

Deep immunophenotyping  
whole blood and synovial fluid  
immune cell populations in  
Rheumatoid Arthritis by mass  
cytometry

Tejal Bhatt

Submitted in accordance with the requirements for the  
degree of

Doctor of Philosophy

The University of Leeds

Faculty of Medicine and Health

And

UCB Pharma, Slough, UK

June 2021

## Statement of integrity

The candidate confirms that the work is her own and that appropriate credit has been given where reference has been made to the work of others.

The coding scripts designed for R Studio were developed by Dr Emma Sutton, UCB Pharma, Slough and remain the intellectual property of UCB Pharma, Slough. With kind permission due to the collaborative nature of this work, these scripts have been applied to analyse the data generated in this work.

This copy has been supplied on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.

2021 The University of Leeds and Tejal Bhatt

## Abstract

Diagnosis and clinical management of Rheumatoid Arthritis (RA) have improved significantly over the last three decades however, not all patients respond successfully to treatment and currently a cure for RA remains elusive. Chronic inflammation is a key feature of RA and if left untreated or improperly managed, can lead to irreversible joint damage. Understanding the immune dysregulation which occurs in RA is central to improving existing approaches for managing RA. Extensive research directed towards immunophenotyping and functional analysis of immune cell subsets in RA have contributed to our current understanding of RA but the impact of these findings have been restricted by existing technology. In this study, deep immunophenotyping was performed using mass cytometry which is a novel multiparameter, high-dimensional single cell technology, to comprehensively interrogate immune cell subsets present in peripheral blood and synovial fluid from patients with RA using a 37 protein marker panel.

This study set out to investigate two main aims, the first of which was to evaluate whether mass cytometry was a suitable technology that could be adopted for large patient cohort immunophenotyping studies which prior to the commencement of this study, had not been reported. Significant effort was invested for protocol validation and optimisation and in addition, with support from UCB Pharma, an automated bioinformatics pipeline was developed to analyse the data without the limitations of traditional gating approaches. The second aim was to assess whether mass cytometry could detect immune cell populations which associated with disease stage or immune signatures which are specific to the local joint microenvironment in RA.

10 palladium isotope tagged barcoded batches were prepared and an internal batch control was included to compare consistency of staining. Data pre-processing steps were applied to obtain a single cell population and discovery hypothesis driven analysis was performed through R Studio using a published pipeline called Diffcyt which assessed immune cell populations across different conditions and between samples.

Diffcyt analysis revealed that decreased percentage changes in innate cell populations are evident early on in RA compared to healthy donors. Differential expression analysis revealed that both innate and T cell subsets in RA peripheral blood have an activated phenotype characterised by CD27, CD38, CD28 and HLA-DR suggesting the beginning of a hyper chronic inflammatory environment.

In addition, analysis of immune cell populations in synovial fluid further corroborated reports of pathologically expanded memory CD4 T cell populations present in RA synovial fluid compared to peripheral blood. Furthermore, a specific CD8 NK cell immunophenotype was detected in RA synovial fluid suggesting a potential role in crosstalk between innate and adaptive immunity.

This research has demonstrated that mass cytometry can be used to comprehensively interrogate the immune landscape in large patient cohorts and using the methodology described here, successfully identifies cell populations that support findings previously reported by other researchers giving confidence in the data obtained by mass cytometry. It is hoped that the methodology for analysing high-dimensional cytometry data will provide a template for future analysis of either this dataset or new datasets and that the cell populations identified here will inform further investigation in RA.

## Acknowledgements

I am indebted to my supervisors Professor Maya Buch, Dr Aamir Aslam, Dr Ash Maroof and Dr Sean Mason for the opportunity to undertake this project. I thank all supervisors for their invaluable guidance and support. In addition, I thank Dr Elena Jones and Dr Ed Vital for their invaluable insight and discussions regarding this project and for their support and enthusiasm.

I am deeply grateful to Dr John Fitton who consented and obtained patient samples from the RADAR BMC cohort and the unenviable task of providing me the clinical details for many of the patients, without which, this research could not have progressed. Also, my thanks to Rachael Peake who provided the clinical information for the early RA cohort and Dr Agata Burska who consented and obtained the healthy donor samples. I would like to extend my thanks to the clinicians, clinical trials teams and the laboratory staff (Diane, Katie, Sharon, and Thelma) at Chapel Allerton Hospital, for their support with patient cohort consent and collection. Naturally, I am deeply indebted to all the patients and healthy donors who generously consented to being included in this project.

I thank Catherine Simpson, Emma Sutton and Suzanne Cole at UCB Pharma for supporting me with mass cytometry, and I further thank Emma Sutton for her guidance with the bioinformatics analysis and development of the R scripts and for generously allowing me to use these for the analysis of this work. In addition, I would like to thank the many individuals at UCB who made me feel welcome and supported me in my research endeavours.

I would like to thank Dr Michelle Campbell for her encouragement and guidance and for being there on the other end of the phone for me.

I am grateful to have been a part of the Leeds Institute of Rheumatology and Musculoskeletal Medicine who provided my student stipend and to UCB Pharma for providing the funding for the consumables used in this project.

I would like to thank Dr Frederique Ponchel for her support and understanding during the course of this research and the School of Medicine for providing me with an extension to complete this work for submission.

Finally, I thank all who have encouraged me along the way and wished me well during the course of this research. I am forever indebted to all those in my past who have unconditionally supported me throughout my academic journey and provided the important foundations for me to build upon. This is not the end but merely the beginning, and I look forward to exploring the ever evolving world of Immunology.

## Table of Contents

<b>Statement of integrity .....</b>	<b>2</b>
<b>Abstract .....</b>	<b>3</b>
<b>Acknowledgements .....</b>	<b>4</b>
<b>Table of Contents.....</b>	<b>5</b>
<b>Table of Flowcharts.....</b>	<b>10</b>
<b>Table of Figures.....</b>	<b>11</b>
<b>List of Tables .....</b>	<b>14</b>
<b>Abbreviations .....</b>	<b>15</b>
<b>Chapter 1 Introduction .....</b>	<b>18</b>
1.1 Rheumatoid Arthritis .....	18
1.2 Epidemiology.....	19
1.3 Clinical manifestation and diagnosis of RA .....	19
1.4 Risk factors contribute to the development of RA .....	20
1.4.1 Genetic predisposition to RA.....	21
1.4.2 Epigenetics .....	21
1.4.3 Smoking.....	22
1.4.4 Gender .....	22
1.4.5 The microbiome.....	23
1.4.6 Lifestyle .....	23
1.5 Immune dysregulation in active RA.....	23
1.5.1 Cytokines in RA.....	24
1.5.2 B cells .....	25
1.5.3 T cells .....	25
1.5.4 Monocytes and macrophages .....	26
1.5.5 Innate and non-immune cell populations in RA pathology.....	27
1.6 Treatment in RA .....	28
1.6.1 Remission in RA .....	28
1.7 Conceptual frameworks aim to define stages of disease progression within RA	29
1.7.1 Pre-clinical and early RA phase.....	29
1.7.2 Refractory RA.....	30
1.8 The multi-omics era and revolution of single cell technologies.....	31
1.8.1 Challenges of compensation in flow cytometry .....	31
1.8.2 Mass cytometry in discovery analysis.....	32
1.8.3 Synovial pathotypes aim to enhance RA classification and treatment approach	34
1.9 Study hypotheses and aims.....	35
1.9.1 Technical hypotheses.....	35
1.9.2 Hypothesis to interrogate the immune landscape in RA peripheral blood and synovial fluid.....	36
<b>Chapter 2 Materials &amp; Methods .....</b>	<b>38</b>
2.1 Patient cohort.....	38

2.1.1	Ethical approval for patient biological samples.....	38
2.1.2	Patient recruitment criteria .....	39
2.1.3	Collection and biobanking of whole blood samples for mass cytometry .	42
2.1.4	Collection and preparation of synovial fluid samples for mass cytometry	43
2.2	Materials .....	43
2.3	Reagents.....	44
2.3.1	Immunophenotyping panel for mass cytometry .....	45
2.3.2	Immunophenotyping panel for flow cytometry .....	48
2.3.3	Reagent kits.....	50
2.4	Methods.....	51
2.4.1	Justification for antibodies selected for mass cytometry .....	51
2.4.2	Flow cytometry .....	52
2.4.2.1	Preparation of peripheral blood mononuclear cells from whole blood .....	52
2.5	CD138 plasmablast cell differentiation.....	52
2.5.1	Purified B cell isolation and labelling .....	52
2.5.1.1	Plasmablast differentiation assay protocol .....	53
2.6	Mass cytometry laboratory methods .....	53
2.6.1	Lanthanide assignment for mass cytometry antibodies .....	53
2.6.2	Antibody conjugation.....	53
2.6.2.1	Antibody concentration .....	53
2.6.2.2	Preloading the polymer with lanthanide .....	53
2.6.2.3	Purifying the lanthanide-loaded polymer .....	53
2.6.2.4	Buffer exchange and partial reduction of antibody.....	53
2.6.2.5	Purifying the partially reduced antibody.....	54
2.6.2.6	Conjugate the antibody with lanthanide-loaded polymer.....	54
2.6.2.7	Washing the metal-conjugated antibody.....	54
2.6.2.8	Antibody conjugation confirmation.....	54
2.6.2.9	Antibody titration .....	54
2.6.3	Compensation for mass cytometry.....	55
2.6.4	Barcoding patient samples for mass cytometry .....	55
2.6.4.1	Preparation of internal batch control for barcoding .....	55
2.6.4.2	Antibody staining.....	56
2.6.4.3	Preparation of samples on the day of mass cytometry .....	56
2.6.5	Data acquisition on mass cytometer .....	57
2.7	Data analysis.....	57
2.7.1	Normalization and concatenation of files .....	57
2.7.2	Debarcoding .....	57
2.7.3	R Studio and statistical analysis.....	57
<b>Chapter 3 Antibody clone validation.....</b>		<b>60</b>
3.1	Introduction .....	60
3.1.1	Optimisation and selection of antibody clones for novel technologies....	60
3.2	Results.....	62

3.2.1	Flow cytometry light scatter properties are redundant with fix freeze processed samples .....	62
3.2.2	CD45+ single cell gating strategy for flow cytometry .....	63
3.2.3	Validation of antibody clones by flow cytometry .....	64
3.2.4	Validation of lanthanide conjugated antibodies for mass cytometry .....	91
3.2.5	Titration and validation of lanthanide conjugated antibodies for mass cytometry	94
3.3	Discussion.....	104
<b>Chapter 4 Methodology for analysing mass cytometry data.....</b>		<b>106</b>
4.1	Introduction .....	106
4.1.1	High throughput data acquired by mass cytometry requires bioinformatics pipelines .....	106
4.2	Results.....	110
4.2.1	Data pre-processing.....	110
4.2.1.1	Debarcoding samples after acquisition on mass cytometer .....	110
4.2.1.2	Identifying and obtaining single cells using Gaussian parameters for mass cytometry .....	113
4.2.1.3	Compensation for mass cytometry data .....	116
4.2.1.4	Clustering algorithms show batch variation.....	119
4.2.1.5	CytoNorm algorithm for normalising antibody staining across sample batches .....	121
4.2.2	Validation of Diffcyt pipeline for mass cytometry data analysis .....	123
4.2.2.1	Preparing the metadata and marker data frame to import into the Diffcyt pipeline.....	123
4.2.2.2	Equal number of events from each sample were analysed by downsampling .....	126
4.2.2.3	Marker expression intensities and expression levels can be assessed across all samples .....	127
4.2.2.4	Multi-dimensional scaling plot measures similarities in marker differential expression across samples and between conditions .....	129
4.2.2.5	FlowSOM clustering and identification of immune cell populations for statistical differential testing.....	133
4.2.2.6	Differential analysis and hypothesis testing.....	138
4.3	Discussion.....	139
4.3.1	Strengths and limitations of the Diffcyt workflow .....	139
4.3.2	Data pre-processing provides confidence for downstream differential analysis	141
<b>Chapter 5 Immunophenotyping whole blood across the stages of Rheumatoid Arthritis</b>		<b>143</b>
5.1	Introduction .....	143
5.1.1	Recent findings of pathologically expanded immune cell subsets in RA by single cell technologies.....	143
5.1.2	Recent findings of pathologically expanded fibroblast subsets in RA by single cell technologies .....	145
5.1.3	Aims .....	146

5.2	Results.....	147
5.2.1	Evidence of immune dysregulation in two pre-clinical RA patients .....	147
5.2.1.1	Cell percentage increase in B cell population at the pre-clinical RA stage .....	148
5.2.2	Analysing MDS plots to determine immune relationship between health, early RA and established RA .....	149
5.2.3	Distinct immune cell populations identify early RA from health .....	154
5.2.3.1	MDS plot analysis that granulocytes influence distinction between early RA and healthy donors .....	154
5.2.3.2	Differential abundance and expression testing reveals specific immune cell populations are dysregulated between early RA and health and span both adaptive and innate immune cell populations .....	156
5.3	Discussion.....	166
5.3.1	Strengths and limitations of study design and analysis through Diffcyt pipeline .....	166
5.3.2	Innate cell populations abundance are significantly decreased in early RA peripheral blood compared to health.....	168
5.3.2.1	MAIT cells are decreased in early RA .....	168
5.3.2.2	Early RA plasmacytoid dendritic cells do not display an activated phenotype .....	169
5.3.2.3	NK and CD8 NK cells in early RA .....	170
5.3.2.4	Neutrophils interact with innate immune cell populations in early RA .....	171
5.3.3	Activation markers – a therapeutic target for RA? .....	172
5.3.4	Significant changes in whole blood immune cell populations were not identified in established RA .....	173
<b>Chapter 6 Immunophenotyping immune cell populations in paired RA synovial fluid and peripheral blood .....</b>		<b>176</b>
6.1	Introduction .....	176
6.1.1	Synovial fluid in joint pathology .....	176
6.1.2	Immune cell populations detected in RA SF.....	176
6.1.2.1	Neutrophils.....	176
6.1.2.2	CD4 <sup>+</sup> T cell subsets .....	177
6.1.2.3	Memory CD8 <sup>+</sup> T cell subsets .....	179
6.1.2.4	NK cells and rare immune cell populations detected in RA SF .....	180
6.1.3	Aims .....	181
6.2	Results.....	182
6.2.1	Comparison of immune cell populations present in paired whole blood and synovial fluid in RA.....	182
6.2.1.1	MDS analysis shows global protein marker staining similarity is specific to biological microenvironment and not influenced by donor.....	182
6.2.1.2	FlowSOM clustering shows the changes occurring at the immune cell level between paired RA peripheral whole blood and synovial fluid ....	183
6.2.1.3	Chronic activation of T cell subsets in RA SF .....	188

6.2.1.4	PD-1 expression is increased in memory CD8 T cells in RA SF compared to matched peripheral blood.....	192
6.2.1.5	CD8 NK cells display a chronically activated functional immunophenotype in RA SF compared to matched whole blood .....	192
6.3	Discussion.....	194
6.3.1	Strengths and limitations of the study.....	194
6.3.2	Immune cell populations differ between donor matched RA whole blood and synovial fluid.....	195
6.3.3	Chronic pathologic activation within the memory CD4 T cell compartment in the RA SF .....	195
6.3.4	CD8+ NK cells – the bridge between innate and adaptive immunity in RA?.....	197
<b>Chapter 7</b>	<b>General Discussion .....</b>	<b>200</b>
7.1	Summary of findings.....	200
7.2	Limitations of the study .....	201
7.2.1	Critical appraisal of mass cytometry.....	201
7.3	Future direction.....	203
7.3.1	Imaging whole synovial biopsy tissue .....	203
7.3.2	Blood immunophenotyping early RA.....	204
7.3.3	Deep immunophenotyping mucosal tissue could provide new insight into pre-clinical RA .....	204
7.3.4	Mutual pathways between RA and Covid-19? .....	205
7.4	Concluding remarks .....	206
<b>Bibliography.....</b>		<b>207</b>
<b>Appendix A .....</b>		<b>244</b>
<b>Appendix B .....</b>		<b>260</b>
<b>Appendix C.....</b>		<b>314</b>
<b>Appendix D .....</b>		<b>321</b>

## Table of Flowcharts

<b>Flow chart 2-1   Inclusion and exclusion criteria for patient cohort .....</b>	<b>41</b>
---	-----------

## Table of Figures

Figure 2-1   Schematic depicting the workflow of barcoding .....	56
Figure 3-1   Comparison of fresh and fixed whole blood from the same donor against light scatter parameters. ....	62
Figure 3-2   Gating strategy to obtain single cell CD45 <sup>+</sup> cell population in peripheral whole blood .....	64
Figure 3-3   Validation of CD3, CD19, CD8a and IgD by flow cytometry. ....	66
Figure 3-4   Validation of CD45RO, CD27, CD14 and CD45RA by flow cytometry. ....	69
Figure 3-5   Validation of $\gamma\delta$ TCR, CD123, CD25 and CD16 by flow cytometry. ....	72
Figure 3-6   Validation of HLA-DR, CD138, TNFR2 and Perforin by flow cytometry. ....	75
Figure 3-7   Validation of FC $\epsilon$ RI $\alpha$ , CTLA4, Va7.2 and CD28 by flow cytometry.....	78
Figure 3-8   Validation of CD86, CD4, CD203c and CD11c by flow cytometry. ....	81
Figure 3-9   Validation of CD56, CD161, PDL1 and CD40 by flow cytometry. ....	84
Figure 3-10   Validation of Granzyme B, PD-1, NKp44 and CD80 by flow cytometry. ....	87
Figure 3-11   Validation of CD127, NKp46, CD38 and CD11b by flow cytometry. ....	91
Figure 3-12   Validation of each lanthanide conjugated antibody for mass cytometry using Universal Comp beads.....	92
Figure 3-13   Manual gating strategy for mass cytometry to obtain a single cell population	95
Figure 3-14   Validation of lanthanide conjugated antibodies for mass cytometry using 1.25 $\mu$ g of antibody.....	96
Figure 3-15   Validation of lanthanide conjugated antibodies for mass cytometry using 1.25 $\mu$ g of antibody.....	98
Figure 3-16   Validation of lanthanide conjugated antibodies for mass cytometry using 1.25 $\mu$ g of antibody.....	99
Figure 3-17   Validation of lanthanide conjugated antibodies for mass cytometry using 1.25 $\mu$ g of antibody.....	101
Figure 3-18   Validation of lanthanide conjugated antibodies for mass cytometry using 1.25 $\mu$ g of antibody.....	102
Figure 3-19   Gating strategy for marker CD138 expression on plasma cells using 1.25 $\mu$ g of antibody.....	103
Figure 4-1   Methodology of analysis applied to data from mass cytometry. ....	109
Figure 4-2   Fluidigm debarcoding software allows samples to be individually identified.	112
Figure 4-3   Automated gating strategy using Gaussian parameters for identifying single cell population.....	115
Figure 4-4   Automated gating for obtaining a single cell population can be applied on a large data set .....	116
Figure 4-5   Compensation matrix used to correct for metal spillover in adjacent channels. ....	118
Figure 4-6   No difference can be observed between uncompensated and compensated channels .....	119
Figure 4-7   Batch variations are observed by two different clustering algorithms .....	120
Figure 4-8   FlowSOM analysis shows that once normalisation through the CytoNorm script has been applied, batches control staining variation becomes consistent. ....	122

Figure 4-9  Bar graph shows 27,000 cells from each sample (3 synovial fluid (SF) and 3 peripheral blood (PB)) included in the Diffcyt analysis after downsampling in R Studio. ....	127
Figure 4-10  Non-redundancy score (nrs) plot shows the antigens defined as 'lineage' markers for each sample included in the analysis. ....	128
Figure 4-11  Smooth densities of marker expression for each sample is calculated for lineage and functional markers .....	129
Figure 4-12  Optimising the Diffcyt pipeline identified batch effects across the neutrophil population when analysing health and early RA .....	132
Figure 4-13  tSNE plot shows the expression level of the CD4 marker. ....	134
Figure 4-14  Heatmap shows 40 clusters identified by FlowSOM and the merging of these clusters into defined cell populations.....	135
Figure 4-15  Heatmap of immune cell populations identified in peripheral whole blood after cluster merging .....	136
Figure 4-16  Immune cell populations, after clustering analysis can be visualised in a FlowSOM tSNE plot .....	137
Figure 5-1  Multi dimensional scaling (MDS) plot observes the staining similarity globally across all samples included for analysis using 27,000 events from each sample. ....	148
Figure 5-2  Increase in percentage of B cells in pre-clinical RA in comparison to health and early RA.....	148
Figure 5-3  MDS plot analysis does not reveal 3 distinct groups of healthy, early RA and established RA.....	150
Figure 5-4  MDS analysis identifies early RA and health cluster into 2 distinct groups ..	151
Figure 5-5  Distinction between health and established RA can be visualised by MDS plot	151
Figure 5-6  No clear distinction between early RA and established RA by MDS analysis	152
Figure 5-7  Multi dimensional scaling (MDS) plot shows early RA (n=23) and refractory (REF) RA (n=23) are indistinguishable based on global staining similarity analysis. ....	153
Figure 5-8  Distinction between non-refractory (NR) (n=11) and refractory (REF) (n=17) is not evident by multi dimensional scaling analysis.....	153
Figure 5-9  Granulocyte population influence distinction between health and early RA	155
Figure 5-10  Absence of granulocyte population results in a loss of distinction between early RA and health .....	156
Figure 5-11  FlowSOM clustering identifies 18 immune cell populations in early RA and health whole blood profiles .....	157
Figure 5-12  Differential abundance heatmap represents significant changes in immune cell populations between early RA (n=23) and health (n=14) .....	158
Figure 5-13  Differential abundance testing through the Diffcyt script shows percentage changes in specific cell populations between early RA (n=23) and health (n=14).	159
Figure 5-14  Differential expression heatmap represents 14 significant changes in functional activation markers across different immune cell populations between early RA (n=23) and health (n=14). ....	161
Figure 5-15  Differential expression testing through the Diffcyt script reveal significant changes in median expression of functional markers across different immune cell populations between early RA and health. ....	163

**Figure 5-16 | Percentage of MAIT cells decrease in established RA (n=23) (depicted in orange) compared to health (n=13)..... 164**

**Figure 5-17 | Differential expression testing comparing established RA with health identifies trends in functional marker expression in B cells, memory CD4 T cells and MAIT cells. .... 165**

**Figure 6-1 | Heterogeneity of memory CD8 T cells outlining their phenotypic characterisation, function and where known, transcription factor profile <sup>370</sup> ..... 180**

**Figure 6-2 | Multi-dimensional scaling (MDS) analysis shows the distinction of three paired RA synovial fluid (SF) and peripheral whole blood (PB). .... 183**

**Figure 6-3 | FlowSOM clustering comparing paired RA peripheral whole blood (PB) and synovial fluid (SF) immune cell populations. .... 185**

**Figure 6-4 | Heatmap showing the lineage markers used to define immune cell subsets 187**

**Figure 6-5 | Box plot representation of median expression of functional protein markers within the memory CD4 T cell population reveal a distinct immune profile (PD-1<sup>+</sup> CTLA4<sup>+</sup> CD27<sup>-</sup> CD28<sup>-</sup> HA-DR<sup>+</sup>) in RA synovial fluid..... 188**

**Figure 6-6 | Gating strategy for identifying T peripheral helper cells by mass cytometry in RA synovial tissue and matched peripheral blood and synovial fluid. .... 190**

**Figure 6-7 | FlowSOM clustering analysis shows the expansion of CD4<sup>+</sup> CD45RO<sup>+</sup> PD-1<sup>+</sup> T cells in RA synovial fluid which immunophenotypically resemble T peripheral helper cells.191**

**Figure 6-8 | Box plot representation of median expression of functional protein markers within the memory CD8 T cell population reveals higher median expression of checkpoint inhibitors CTLA-4 and PD-1 in RA synovial fluid compared to matched peripheral whole blood..... 192**

**Figure 6-9 | Box plot representation of median expression of functional protein markers within the CD8 natural killer (NK) cells population reveals a chronically activated immune phenotype but an incapability of inducing apoptosis through Granzyme B and Perforin ..... 193**

## List of Tables

<b>Table 1-1   Comparison of the 1987 ACR and 2010 EULAR criteria for diagnosing RA <sup>7,18</sup> ..</b>	<b>19</b>
<b>Table 2-1   Number of healthy and patient peripheral blood samples obtained at a single time point.....</b>	<b>42</b>
<b>Table 2-2   Number of established Rheumatoid Arthritis longitudinal peripheral blood samples .....</b>	<b>42</b>
<b>Table 2-3   Number of paired peripheral blood and synovial fluid samples and single synovial fluid samples obtained from inflammatory arthritis .....</b>	<b>42</b>
<b>Table 2-4   General equipment and plastic ware.....</b>	<b>43</b>
<b>Table 2-5   General reagents, company and catalogue number .....</b>	<b>44</b>
<b>Table 2-6   Metal lanthanides abbreviation for mass cytometry .....</b>	<b>45</b>
<b>Table 2-7   Mass cytometry immunophenotyping panel for this study .....</b>	<b>46</b>
<b>Table 2-8   Mass cytometry immunophenotyping panel for this study .....</b>	<b>48</b>
<b>Table 2-9   Mass cytometry antibody conjugation reagents and equipment, Fluidigm catalogue number 201300 .....</b>	<b>50</b>
<b>Table 2-10   Additional reagents required for antibody conjugation not provided with kit50</b>	
<b>Table 2-11   Antibody conjugation with lanthanide metal equipment.....</b>	<b>51</b>
<b>Table 2-12   Fluidigm, Cell-ID™ 20-Plex Pd Barcoding kit .....</b>	<b>51</b>
<b>Table 2-13   Plasmablast differentiation reagents .....</b>	<b>51</b>
<b>Table 4-1   Classification of lineage and functional protein markers for Diffcyt analysis</b>	<b>125</b>
<b>Table 4-2   Immune cell markers used to define cell populations in FlowSOM .....</b>	<b>137</b>
<b>Table 5-1   Significant changes in functional protein marker median expression in immune cell subsets in early RA patients identified by differential expression testing in Diffcyt pipeline .....</b>	<b>160</b>

## Abbreviations

ACPAs – anti citrullinated peptide autoantibodies  
ACR – American College of Rheumatologists  
AMP - Accelerating Medicines Partnership  
APC – allophycocyanin  
APCs – antigen presenting cells  
CD – cluster of differentiation  
CIA – collagen induced arthritis  
CRP – C reactive protein  
CTLA4 – cytotoxic T lymphocyte-associated protein 4  
CyTOF – cytometry time of flight (mass cytometry)  
DAS – disease activity score  
DCs – dendritic cells  
Di – dual counts  
DA - Differential abundance  
DE - Differential expression  
DMARDs – disease modifying anti rheumatic drugs  
dPBS – distilled phosphate buffered saline  
EDTA - Ethylenediaminetetraacetic acid  
ESR – erythrocyte sedimentation rate  
EULAR – European League Against Rheumatism  
FACS – fluorescence activated cell sorting  
FLS - Fibroblast-like synoviocytes  
Fc – fragment crystallizable region  
FITC – fluorescein isothiocyanate  
FLS – fibroblast-like synoviocytes  
(Flow)SoM – self-organizing map (clustering algorithm)  
FSc – forward scatter  
FCS – flow cytometer standard  
GWAS – genome wide association studies

HLA-DR – human leukocyte antigen – DR isotype  
IA – inflammatory arthritis  
IFN $\gamma$  – interferon gamma  
Ig – immunoglobulin  
IL-6 – interleukin 6  
JAK – Janus Kinase  
MAIT - mucosal associated invariant T cells  
MDS – multi-dimensional scaling  
MHC II – major histocompatibility complex II  
miRNAs – microRNAs  
MS – multiple sclerosis  
MDS - multidimensional scaling  
NK cells – Natural Killer cells  
NICE – National Institute for Health and Care Excellence  
OA – Osteoarthritis  
PADI – peptidyl arginine deiminase  
PBS – phosphate buffered saline  
pDCs – plasmacytoid dendritic cells  
PC5.5 – phycoerythrin-cyanine 5.5  
PECy7 – phycoerythrin-cyanine 7  
PE – phycoerythrin  
PerCP – peridinin chlorophyll – A protein  
PFA – paraformaldehyde  
PDL1 - programmed death ligand 1  
RA – Rheumatoid Arthritis  
RADAR (BMC) - Rheumatoid Arthritis Disease Research (Biologics monitoring clinic)  
RF – Rheumatoid Factor  
RNA – ribonucleotide acid  
RT – room temperature  
SSc – side scatter  
SE – shared epitope

Sf – synovial fluid

SFMCs – synovial fluid mononuclear cells

SLE – systemic lupus erythematosus

SoM – self-organising map

TCR – T cell receptor

TLR – toll-like receptor

TNF – tumour necrosis factor

Tph – T peripheral helper cells

Tregs – T Regulatory cells

UK – United Kingdom

VAS – visual analogue scale

# Chapter 1 Introduction

## 1.1 Rheumatoid Arthritis

Rheumatoid Arthritis (RA) is a chronic, debilitating inflammatory disease which primarily targets the small joints in the hands and feet and has a preponderance towards females at a ratio of 3:1<sup>1</sup>. The age of onset for RA is usually between the 40-60 years, however, this does not preclude earlier or later onset outside of this range<sup>1</sup>.

The presence of autoantibodies provides the single most compelling evidence that RA is an autoimmune disease with two classes of autoantibodies identified in individuals with RA; Rheumatoid Factors (RF) and anti-citrullinated peptide autoantibodies (ACPAs) which target the Fc (fragment crystallizable region) portion of IgG (immunoglobulin G) antibodies<sup>1</sup>.

In 1940, Waaler was the first to report the occurrence of RF in RA serum<sup>2</sup>. Waaler obtained serum from patients with RA, and observed that RA serum inhibited haemolysis of sheep red cells and caused the cells to agglutinate<sup>2</sup>. The activating component in the serum was later identified as RF which targets the Fc portion of immunoglobulins<sup>2</sup>. Further work subsequently showed that RF are produced by B cells and form immune complexes typically IgM and IgA isotypes, that bind Fc fragments of IgG. RF are present in 60-80% of patients with RA, however, RF has been identified in other diseases including systemic lupus erythematosus (SLE) and mixed connective tissue disease in addition to chronic infection and old age, suggesting that RF are not specific to RA<sup>3</sup>. Furthermore, RF is present in healthy individuals, and a titre of  $\geq 50$  Units/mL is considered specific to RA<sup>3</sup>.

Although high RF titres are associated with a poor prognosis in RA, their lack of specificity has left researchers in RA looking for other autoantibodies that may be implicated in RA pathology<sup>4</sup>. ACPAs are present in 70-90% of patients and have a 90-95% specificity to RA<sup>5</sup>. ACPAs occur by amino acids undergoing the post translational modification process of citrullination<sup>6</sup>. An arginine residue is replaced by citrulline which is catalysed by the peptidylarginine deiminase (PADI) enzymes<sup>6</sup>. To date, certain citrullinated proteins have been identified as autoantigens that can form ACPAs including structural proteins such as type I and II collagen, fibrin and fibrinogen<sup>6</sup>.

Autoantibodies are important in their diagnostic and prognostic value and are central to the classification criteria for identifying RA in patients<sup>7</sup>. ACPA specificities and titres in longitudinal analysis of patients have been detected up to 10 years before symptoms of RA begin to appear and have shown a significant increase in titre 6 months before the onset of RA suggesting that the pathogenesis directly affects titre levels<sup>8</sup>. Patients who have high titres of autoantibodies have poorer prognostic outcomes in particular structural damage and joint erosions<sup>9</sup>. However, serology can change during the course of disease for some patients suggesting that serology alone cannot be relied upon to predict the course of RA progression<sup>10,11</sup>.

Furthermore, the focus of RA research has returned to the synovium (also known as synovial tissue), the primary site of joint pathology in RA. The synovium is a soft tissue lining which covers the spaces across the diarthrodial joints, tendon sheaths and bursae<sup>12</sup>. The synovium consists of two layers: the upper layer known as the intimal lining and the lower layer known as the subintimal lining. In health, the synovium is a largely acellular structure with a 1-2 cell layer thick intimal lining consisting of two cell subtypes: macrophage-like synoviocytes and fibroblast-like synoviocytes (FLS)<sup>12</sup>. In addition, a small volume of synovial fluid is present to lubricate the joint. In RA, the synovial lining becomes inflamed and an infiltration of immune cells can be detected. Recent work has focused on how the synovium is implicated in RA revealing distinct immune signatures that suggest different pathotypes exist in RA<sup>13,14</sup>.

The aim of this chapter is to provide an overview of RA in terms of epidemiology, diagnosis, treatment and understanding the conceptual frameworks that surround the onset of RA and the 'stages' within RA progression. Lastly, the recent advancements in single cell technologies, particularly in cytometry, have provided a renewed opportunity to interrogate multi-parameter protein expression on immune cell populations. The final section of this chapter will focus on mass cytometry and its applications for immunophenotyping in RA.

## 1.2 Epidemiology

Most of the epidemiological studies in RA have been conducted in Europe and North America with prevalence estimated to be between 0.5%-1% in Caucasian individuals<sup>15</sup>. Whilst it is thought that this percentage remains consistent across the population worldwide, it should be applied to other populations with a degree of caution<sup>16</sup>. In certain populations, the prevalence of RA is much higher such as in the Pima Indians (5.3%) and Chippewa Indians (6.8%) and the incidence rate is lower in China and Japan (0.2%-0.3% respectively) suggesting genetics play a role in the prevalence of RA. Interestingly, some studies have reported that the incidence rate in United States has declined which can be attributed to improved treatment<sup>16,17</sup>.

## 1.3 Clinical manifestation and diagnosis of RA

To date, there is no accurate and consistent diagnostic test or biomarker which can be used to diagnose RA. Instead classification criteria are used by rheumatologists which include clinical manifestations and serology assays (ACPA and/or RF positive) that inform diagnosis. Erythrocyte sedimentation rate (ESR) and C reactive protein (CRP) are calculated as markers of general inflammation. RA is regarded as a heterogeneous disease and patients differ in terms of duration of disease, joints affected, treatment response and prognostic outcomes. Therefore, the classification criteria is to be used in conjunction with the clinician's expertise. Two classification criteria are currently used to guide the diagnosis of RA; the 1987 American College of Rheumatologists (ACR) and the 2010 European League Against Rheumatism (EULAR) which are outlined in Table 1-1<sup>7,18</sup>. RA is currently characterised by soft synovial tissue swelling known as clinical synovitis accompanied by morning stiffness which lasts longer than 15 minutes<sup>18</sup>. Joints are usually affected in a symmetrical manner, but it is also possible to observe inflammation in an asymmetrical pattern. Although RA is a heterogeneous disease, there are distinct features that are specific to RA. This includes the joints that are involved which are the metacarpophalangeal and proximal interphalangeal joints but not the distal interphalangeal joints which is commonly associated with osteoarthritis<sup>18</sup>. Moreover, RA is a disease that can be distinguished from other arthritic pathologies due to the irreversible damage caused to the bone and the loss of function as a result<sup>18</sup>.

**Table 1-1 | Comparison of the 1987 ACR and 2010 EULAR criteria for diagnosing RA<sup>7,18</sup>.**

1987 ACR Criteria		2010 EULAR Criteria		
Criteria	Description	Score	Description	Score
Morning stiffness	In and around joints for at least 1 hour	1	Clinical synovitis/swelling in at least 1 joint not explained by another disease	Not given
Joint involvement	Physician observed soft tissue swelling or fluid in 3 of 14 possible joints	1	1 large joint 2-10 large joints 1-3 small joints (with/out large joints) 4-10 small joints (with/out large joints) > 10 joints (at least 1 small)	0 1 2 3 5
Arthritis of hand joints	At least 1 swollen hand or wrist area	1	Not measured	Not measured
Symmetric arthritis	Simultaneous bilateral movement	1	Not measured	Not measured
Rheumatoid nodules	Subcutaneous nodules over bony protrusions, extensor surfaces, or in juxta articular regions observed by physician	1	Not measured	Not measured
Serology	Positive RF serum test	1	Negative RF and negative ACPA Low positive RF or ACPA High positive RF or ACPA	0 2 3
Radiographic changes	Erosions or unequivocal bony decalcification in or adjacent to the involved joints but not alike osteoarthritis	1	Not measured	Not measured
Acute phase reactants	CRP and ESR	Not measured	Normal CRP and ESR Abnormal CRP or ESR	0 1
Duration of symptoms	First 4 criteria must be present for at least 6 weeks	Not measured	< 6 weeks ≥ 6 weeks	0 1
Criteria score required	≥4/7		≥6/10	

Whilst the synovial joint is the primary site of pathology, RA is regarded as a systemic disease with cardiovascular disease being the most cited cause of death in patients<sup>19</sup>. Understanding which patients are susceptible to developing cardiovascular disease or whether cardiovascular disease precedes onset of RA is not fully understood although current research suggests that genetic and serological factors play a role<sup>20</sup>. Additional comorbidities associated with RA include type II diabetes mellitus, obesity and respiratory disease<sup>21</sup>. Many studies have used the collagen induced arthritis (CIA) mouse model to understand RA but it is often argued that this model is not fully representative of RA due to mice not developing systemic complications as identified in patients with RA<sup>22</sup>.

Mental health particularly depression and fatigue has been reported by patients through health assessment questionnaires<sup>23</sup>. The cause of fatigue and depression can largely be attributed to the chronic pain that is associated with RA<sup>23</sup>. Furthermore, for some patients the loss of independence or work due to joint damage or associated comorbidities can affect mental well-being<sup>23</sup>.

The classification criteria are important for guiding the clinician when diagnosing RA. However, research and knowledge about RA have increased significantly since the ACR 1987 and EULAR 2010 guidelines were compiled<sup>7</sup>. The classification criteria is driven by seropositive serology although a proportion of patients are seronegative which is currently not included in the classification criteria<sup>24</sup>. It has been observed that seronegative patients fare less well in their prognosis compared to seropositive patients<sup>24</sup>. The classification criteria to some extent is tentative in order to prevent a misdiagnosis, given the treatment for RA aims to suppress immunological pathways. As more research emerges, it is likely that the classification criteria will be updated in order to improve disease management and thus patient quality of life.

#### 1.4 Risk factors contribute to the development of RA

### 1.4.1 Genetic predisposition to RA

It is not fully understood what causes RA although many risk factors have been identified over the years that have been shown to increase susceptibility of developing RA. However, evidence shows that RA has a significant genetic component which is estimated to be 60% inheritable<sup>25</sup>. Human Leukocyte Antigen – DR isotype (*HLADR*) is the most strongly associated gene with predisposition to RA first observed by Stastny in 1976<sup>26</sup>. *HLADR* is a polymorphic gene; the third hypervariable region DRB1 is known to be important for T cell antigen recognition and the alleles HLA-DRB\*01 and HLA-DRB\*04 are strongly associated with the risk of developing RA<sup>27</sup>. The shared epitope (SE) hypothesis refers to five amino acids that form a functional unit in the DRB1 region. However, this hypothesis has evolved since to suggest that whilst the three amino acids at positions 72-74, RAA (arginine, alanine, alanine) remain fixed, the amino acids at positions 70 and 71 can vary thus modulating the risk of developing RA depending on the sequence<sup>27,28</sup>. DR4 is however polymorphic in the general population, and present in other autoimmune diseases; it is therefore clear that these polymorphisms alone are not enough to break peripheral tolerance and initiate the development of RA.

In addition to the *HLA-DR* gene, genome wide association studies (GWAS) with fine mapping have identified more than a 100 loci as having an association with RA, although a much weaker correlation in comparison to *HLA-DR*<sup>29</sup>. Peptidyl arginine deiminase 4 (*PADI4*) is a member of the family of enzymes that catalyse the post-translational modification reaction of arginine to citrulline which results in ACPAs. In a Japanese cohort, single nucleotide polymorphisms formed a haplotype and further functional studies determined that this haplotype affected stability of the mRNA causing increased production of citrullinated peptides<sup>29</sup>. However, in 2004 Barton et al. reported that *PADI4* could not be identified in the United Kingdom using a Caucasian cohort however, in 2005, researchers in Japan reconfirmed the presence of *PADI4* in a different Japanese cohort suggesting that it is at least associated in Japanese RA patients<sup>30,31</sup>. *PTPN22*, which encodes a hematopoietic-specific tyrosine phosphatase, is a risk allele present in approximately 17% of Caucasian individuals in the general population rising to 28% in Caucasian RA patients<sup>32</sup>. It was originally identified in type I diabetes<sup>33</sup> but has since been associated with juvenile idiopathic arthritis, psoriatic arthritis, psoriasis and multiple sclerosis thus implicating it as a strong candidate for autoimmune disease association<sup>34</sup>. The genes *PRL* and *NFIA* associate with seronegative RA and the genes *AFF3*, *CD28* and *TNFAIP3* are attributed to seropositive RA suggesting genes can drive differences in serology<sup>35,36</sup>.

### 1.4.2 Epigenetics

Genetic cues alone cannot explain the aetiology of RA and environmental triggers are thought to combine with a genetic predisposition and lead to the onset of RA. Epigenetics is the process by which heritable phenotypic changes can occur without affecting the DNA sequence. There are many types of epigenetic mechanisms, DNA methylation being the most common, typically observed using RA peripheral blood comparing to healthy or disease controls. Numerous cell types are known to have aberrant methylation patterns including B cells, T cells and monocytes. Various genes have been identified as either hypermethylated such as *CTLA-4* or hypomethylated such as *FoxP3* in T regulatory cells from RA patients<sup>37,38</sup>.

microRNAs (miRNAs) have been identified to influence gene regulation by either repressing translation or degrading mRNA. miRNA-146a and miRNA-155 are the best characterised in terms of

RA with elevated levels of both having been observed in peripheral blood and synovial tissue<sup>39,40</sup>. It remains unclear whether aberrant miRNA expression is a result of RA or whether dysregulation leads to the onset of RA pathogenesis, however, this along with epigenetic modulation could be harnessed as both a biomarker and a therapeutic intervention.

### 1.4.3 Smoking

Susceptibility to RA cannot be entirely explained by genetic factors and it is thought that 40% of risk factors can be attributed to environmental influence of which smoking is one of the leading causes. Individuals who have a long history of smoking 20 pack cigarettes a day are at a greater risk of developing RA compared to those who do not smoke<sup>41</sup>. Duration, in terms of number of years an individual had smoked had a higher association with risk of RA than number of cigarettes smoked<sup>42</sup>. In addition, individuals who smoke are shown to have a 20-fold increase in the shared epitope and are typically RF positive compared to non-smokers. Moreover, smoking is thought to increase the risk of ACPAs in these patients who have the shared epitope<sup>43</sup>. Smoking is thought to be associated with poor increases pulmonary infections and disrupt the microbiota composition.

Smoking has long been associated with early mortality in the general population and smoking is associated with a poorer prognostic outcome in RA. Patients who cease smoking show remarkable improvement in their lipid profile and cardiovascular health and lower disease activity<sup>44</sup>.

Intriguingly it has been observed that the non-tobacco component of cigarettes has been associated with the risk of RA and not the tobacco component suggesting that other particulates within cigarettes are responsible. Interestingly, studies have looked at particulate exposure in work-related environments with findings from a study in Sweden showed that men who had been exposed to silica were at a twofold increased risk of developing RA compared to men who had not been exposed<sup>45</sup>. Furthermore, a threefold increased risk to RA was observed in men whose occupation involved stone crushing or rock drilling<sup>45</sup>. However, smoking could not be discounted as a confounding factor in this study as some of the individuals were current smokers or had smoked in the past. In a separate study conducted in Malaysia, a gene-environment interaction between HLA-DRB1 SE and textile dust exposure increased risk of RA in ACPA-positive and ACPA-negative women who had never smoked<sup>46</sup>. The authors of this study go on to speculate that textile dust could have a general immune activating role compared to both silica and smoking due to presence in ACPA negative patients as well<sup>46</sup>.

It has been noted that data regarding the role of smoking in RA can be conflicted suggesting that whilst smoking is thought to be implicated in RA, it cannot be concluded that smoking is solely responsible for increasing the risk of RA<sup>47</sup>. Further work elucidating the mechanistic process between particulate exposure and the increased risk of RA would be essential to confirm clinical observations. For now, it is clear however, that smoking combined with genetic and other environmental factors can enhance the risk of RA.

### 1.4.4 Gender

Females are at a threefold increased risk of developing RA. This in part is driven by hormones, in particular oestrogen which is thought to have the ability to immunomodulate the immune system although this has been disputed by other researchers within the field<sup>48</sup>. Women who have not given birth are at a greater risk of developing RA compared to women who have given birth although this is not definitely proven<sup>49</sup>.

### 1.4.5 The microbiome

The gut and oral microbiome have been implicated in RA pathogenesis and appear different when compared to non-RA individuals<sup>50</sup>. It is thought that the beginning stages of RA are proinflammatory and that microbiota may influence this environment by interacting with immune cells and further promoting immune dysregulation<sup>51</sup>. Evidence identifying IgA immune complexes in the circulation and synovial fluid of patients promote the hypothesis that RA may be of mucosal origin or at the very least mucosal sites contribute to RA pathology<sup>52</sup>. To date, two main species of bacteria have been associated with microbiota dysbiosis in patients with RA. The first of these is *Prevotella copri* which is increased in new onset RA patients compared to patients with established RA<sup>53</sup>. In addition, the oral microbiome has been associated with RA pathogenesis. Patients with RA who had periodontal disease revealed oral microbiome dysbiosis which when treated alleviated RA symptoms<sup>53</sup>. *Porphyromonas gingivalis* has provided a mechanistic link between the oral microbiome and citrullination of peptides ultimately leading to autoantibody production and synovial inflammation. Further evidence suggests that *Porphyromonas gingivalis* can also manipulate the TLR2 pathway compromising its ability to clear bacteria.

### 1.4.6 Lifestyle

Some lifestyle factors have shown to increase the risk of developing RA. Obesity has shown a modest association with predisposition to RA with an odds ratio of 1.45 in individuals with a body mass index over 30kg/m<sup>2</sup><sup>54</sup>. Lower socioeconomic status and manual labour jobs increase an individual's predisposition to RA. Conversely, moderate consumption of alcohol has been associated with lowering the risk of developing RA<sup>55</sup>. Furthermore, moderate consumption of alcohol is associated with lowering disease severity in RA<sup>55</sup>.

As can be observed, there are many risk factors that have been associated with increasing susceptibility to RA reflecting the heterogeneity of the disease. It is unlikely that an individual with RA would be predisposed to all the risk factors neither is it likely that one risk factor is solely responsible for RA. The lack of mechanistic studies showing how these factors cause disease make it difficult to prevent RA. However, understanding the aetiopathogenesis of RA and the associated risk factors remains an active research interest.

## 1.5 Immune dysregulation in active RA

The transition from a preclinical state to clinical synovitis comprises of numerous immunological pathways that become dysregulated<sup>56</sup>. It is apparent that the presence of autoantibodies alone do not result in RA and therefore the concept of the elusive 'second hit' has been much sought for to identify the mechanism by which the pathology manifests<sup>56</sup>. It has been hypothesised that the second-hit comprises of a combination of vascular, neuro-regulatory, microtrauma or transient-infection-dependent pathways<sup>56</sup>. This change in the vasculature allows for immune complexes to bind to Fc receptors within the synovium and an immune cascade to ensue including synovial innate cells to release vasoactive mediators and increase antibody activity within the joint<sup>56</sup>. Additionally, inflammatory events within the joint can be exacerbated by complement activation, toll-like receptor engagement and osteoclastogenesis<sup>56</sup>. This coupled with the production of proinflammatory chemokines and cytokines result in clinical synovitis. Clinical synovitis during the pre-RA phase is initially not obvious however, as symptoms of RA clinically manifest, it is apparent that normal synovial lining layer thickness is altered from 2-3 cells thickness to 10-12 cells depth

mainly consisting of macrophages and fibroblast-like synoviocytes<sup>57</sup>. However, the complexity of the immune dysregulation that occurs in RA is becoming better characterised with increasing access to biological material including synovial tissue biopsies allowing for a more robust interrogation of the immune infiltrate<sup>58</sup>. The following sections aim to summarise the contribution of cellular and cytokine biology which contribute to RA pathogenesis and is well characterised in patients who have active RA.

### 1.5.1 Cytokines in RA

It is evident that cytokines actively contribute to RA pathogenesis and play an important role in inflammation, articular destruction and comorbidities associated with RA<sup>59</sup>. Of these, TNF inhibition has been regarded as one of the biggest treatment successes in RA. The importance of TNF in RA was demonstrated in 1989, by elegant in vitro cell culture experiments in which disaggregated RA synovial mononuclear cell cultures, along with IL-1, IL-6, GM-CSF and IL-8 were spontaneously and chronically produced over a 5 day culture period<sup>60</sup>. Intriguingly, when TNF production was inhibited in vitro, it was discovered that spontaneous production of IL-1 protein and *IL-1B* mRNA expression also reduced<sup>60</sup>. The importance of TNF inhibition has been further identified to reduce the expression of GM-CSF, IL-6 and IL-8<sup>61,62</sup>. TNF is regarded as one of the most successful treatment blockades in RA and is also used to treat other chronic inflammatory arthritis conditions including juvenile arthritis, psoriatic arthritis and ankylosing spondyloarthritis<sup>63</sup>.

The discovery and effectiveness of TNF blockade has led to additional cytokines within the TNF superfamily being identified including B lymphocyte stimulator (BLyS) and a proliferation-inducing ligand (APRIL)<sup>64</sup>. BLyS promotes the survival and differentiation of B cells and is produced by numerous cell types including stromal cells, B cells, activated T cells, stimulated neutrophils, monocytes, macrophages and plasmacytoid dendritic cells<sup>65</sup>. Whilst BLyS can act alone its bioavailability is limited however, together with APRIL, they can maintain the activation of B cells and thus promote autoimmune disease<sup>65</sup>. Increased levels of BLyS have been observed in serum, synovial fluid and saliva in RA and have been associated with severity of disease which improves upon inhibition with Belimumab, a drug which inhibits BLyS<sup>66</sup>.

The RANKL (Receptor activator of nuclear factor- $\kappa$ B ligand)/RANK/osteoprotegrin pathway that is connected between bone resorption and bone formation<sup>67</sup>. RANKL is a member of the TNF superfamily and combined with M-CSF can form osteoclasts. Expression of RANKL is regulated by TNF and modulated by osteoprotegrin<sup>67</sup>. In a clinical trial study conducted in Japan, it was observed that denosumab which inhibits RANKL, was capable of suppressing bone erosion and bone loss in patients with RA with or without concomitant osteoporosis<sup>68</sup>.

In addition to TNF, inhibition of IL-6 is regarded as a successful therapeutic target and tocilizumab is widely used in the clinic to treat patients with RA, in particular those patients who have not responded well to TNF treatment<sup>69</sup>. In the early 1990s, in vitro cell culture experiments revealed that IL-6 might be involved in osteoporosis, cartilage destruction and synovial inflammation associated with RA and animal models showed that IL-6 inhibition was effective in preventing RA<sup>70,71</sup>.

IL-6 activates cells via a signalling mechanism that requires two receptor components, IL-6R $\alpha$  and glycoprotein 130 which together form a hexameric complex which induces signal transduction<sup>72</sup>. IL-6 can signal through two different pathways, classical signalling, or trans-signalling. Classical signalling occurs by IL-6 binding to its membrane-bound receptor IL-6R $\alpha$  whereas trans-signalling

binds to circulating soluble IL-6R $\alpha$  which together form a complex with glycoprotein 130<sup>73</sup>. IL-6R $\alpha$  is expressed on only a few cell types including hepatocytes, monocyte, macrophages neutrophils and some T cell subsets<sup>72</sup>. Activation of the receptor complex occurs by phosphorylation of JAK 1 and 2 and tyrosine kinase 2 which can be therapeutically targeted however, in RA the receptor is targeted as the concentration is less variable across patients thus simplifying dose regimes<sup>69</sup>.

A multitude of predominantly proinflammatory cytokines have been discovered that are capable of joint destruction. These include IL-1 (and members of IL-1 superfamily including IL-1 $\alpha$ , IL-1 $\beta$ , IL-18 and IL-33), IL-7, IL-12, IL-17, IL-18, IL-21, IL-23 and IL-32<sup>74</sup>. There are anti-inflammatory cytokines that have been associated with RA pathology including IL-4, IL-10, IL-13 and IL-35 but their role in RA is less prominent in comparison to proinflammatory cytokines<sup>74</sup>.

### **1.5.2 B cells**

Adaptive and innate immune cells are both implicated in RA pathogenesis. The most studied immune cell populations are B and T cells due to their association with autoantibody production.

B cells are important in RA due to their ability to produce physiologically important proteins including RF, ACPAs and proinflammatory cytokines (TNF, IL-6, IL-12, IL-23 and IL-1) and their ability to activate T cells through co-stimulatory molecules<sup>75</sup>. Under normal conditions, autoreactive B cells are eliminated by immune checkpoint mechanisms which are stationed throughout the B cell development stages predominantly during the early immature to immature B cells in the bone marrow (central tolerance), or before B cells become mature naïve B cells in the periphery (peripheral tolerance)<sup>75,76</sup>. Despite these immune checkpoints, it is estimated that 20% of the antibodies produced by mature naïve B cells entering the periphery are autoreactive and shown to be even higher in RA patients<sup>77,78</sup>. It was also observed that some clones of mature naïve B cells displayed reactivity to citrullinated peptides and immunoglobulins in RA suggesting aberrant processes within the tolerance mechanisms thought to be attributed to the genetic predisposing gene PTPN22 which may exert its influence on the B cell activation threshold<sup>76,79</sup>. Autoreactive B cells are also capable of pathologically contributing to immune dysfunction, inflammation and bone damage<sup>80</sup>.

B cells also contribute to the formation of ectopic lymphoid-like structures (ELS) which are observed to develop in the synovial membrane in approximately 25% of patients with RA<sup>81</sup>. ELS show a high degree of cellular organization which resemble structures similar to those observed in lymph nodes and are composed of distinct T and B cell areas which show a diffused or aggregated composition and a network of follicular dendritic cells<sup>82</sup>. Evidence suggests that ELS are capable of inducing B cell somatic hypermutation and class switch recombination in the local synovial environment by the presence of the enzyme activation-induced cytidine deaminase and ACPA positive plasma cells<sup>81</sup>. This evidence further suggests that B cells are able to deviate from the norm and bypass trafficking to the lymph nodes to differentiate into antibody secreting cells thus expediting the production of autoantibodies locally. The proximity of B and T cells in ELS suggest that they can communicate with one another closely meaning that these B cells can present autoantigen to T cells, leading to T cell activation, proliferation and proinflammatory cytokine production indicating that these structures can self-perpetuate autoimmunity<sup>81</sup>.

### **1.5.3 T cells**

Evidence that T cells are central to RA pathogenesis first emerged in the 1980s where increased numbers of activated T cells were observed in the peripheral blood, synovial membrane and synovial fluid<sup>83,84</sup>. It is evident that CD4+ T cells play a key role in RA pathogenesis highlighted by the fact that the RA genetic risk HLA-DR loci preferentially map to enhancers and promoters which are active in CD4+ T cell subsets. There are numerous subsets of CD4+ T cell subsets distinguished by their differential capacity to produce cytokines<sup>85</sup>. At the simplest level, CD4+ T cells can be divided into two subsets known as T helper 1 (Th1) and T helper 2 (Th2). Th1 cells are important in the defence against intracellular pathogens and secrete IFN $\gamma$ , TNF and IL-2 whereas Th2 cells mediate against parasitic infections such as helminths and secrete IL-4 and IL-5<sup>85</sup>. Although this is a simplistic classification, it is apparent that Th1-like CD4+ T cells are more abundant in RA than Th2 CD4+ T cells as evidenced by the detection of a significant increase in IFN $\gamma$  cells in RA synovial fluid that was not observed in peripheral blood<sup>86</sup>. Furthermore, the same study also noted that IL-4 cells did not change between the synovial fluid and peripheral blood which further supported the skewing towards a Th1 phenotype within RA<sup>86</sup>. Effector functions of Th1 CD4+ T cells in RA are thought to be involved in macrophage activation and B cell isotype switching.

The Th1 and Th2 paradigm has evolved with additional subsets of CD4+ T cells have been characterised. IL-17 is a proinflammatory cytokine which was identified as a product of activated T cells leading to the coining of Th17 cells<sup>87</sup>. Th17 were shown to be distinct from the Th1/2 paradigm as identified by the lineage-specific transcription factors ROR $\gamma$ t. CD4+ T cells secreting IL-17 were detected in the RA synovium in 1999 although at the time, the term Th17 cells had not been coined so was termed as being produced by 'some' Th1 cells<sup>88</sup>. IL-17 from Th17 cells can stimulate epithelial, endothelial and fibroblast cells to produce proinflammatory cytokines including IL-6, IL-8 and GM-CSF and also play a role in recruiting neutrophils as evidenced in RA synovial fluid<sup>89</sup>.

Regulatory T cells (Tregs) are specialised CD4+ T cells which are capable of modulating the immune response to environmental pathogens and suppressing inappropriate immune responses to self-antigens. Tregs were first characterised by Sakaguchi et al. as CD4+ CD25+ and further distinguished by the transcription FoxP3<sup>90</sup>. Evidence of Treg involvement in RA initially came from studies in mice where depletion of CD4+CD25+ T cells showed a more severe arthritis<sup>91</sup> and conversely when CD4+CD25+ T cells were adoptively transferred into the joint of mice, this decelerated disease progression<sup>92</sup>. Tregs numbers in RA peripheral blood is inconsistent with some reports stating that Tregs are elevated and other reports suggesting that they are diminished although in RA synovial fluid there is a general consensus that Tregs are elevated<sup>93</sup>. The role of Tregs in RA is still under investigation and whether increasing their numbers in the joint would alleviate inflammation and promote self-tolerance has yet to be definitively shown in clinical trials.

In recent years, single cell immunophenotyping of T cells have revealed novel subsets of CD4 T cells which pathologically contribute to RA will be discussed later on.

#### **1.5.4 Monocytes and macrophages**

At present, it is widely accepted that there are three subsets of human monocytes present in peripheral circulation known as classical (CD14+CD16-), intermediate (CD14+CD16+) and non-classical (CD14-CD16+)<sup>94</sup>. Within RA, it is the intermediate subset which is observed to be at a higher frequency when compared to health and correlated with reduced response to methotrexate<sup>95,96</sup>.

It was previously thought that all circulating monocytes gave rise to tissue macrophages<sup>97</sup>. In light of elegant fate-mapping lineage studies performed in mouse embryos, the relationship between

monocytes and macrophages is more complex than initially thought<sup>98</sup>. These studies have shown that tissue-resident macrophages are seeded from embryonic precursors giving rise to tissue specific macrophages where in adulthood macrophages can undergo self-renewal although it is uncertain whether this happens in the synovium<sup>98</sup>. Macrophages are phagocytic cells which are capable of antigen presentation, T cell activation and cytokine production. Broadly, human macrophages can be divided into two subsets known as classically activated (M1) and alternatively activated (M2) macrophages although it is known that macrophages are dynamic and capable of responding to environmental stimuli meaning that they can fluctuate between M1 and M2 states<sup>99</sup>. M1 macrophage differentiation is driven by proinflammatory cytokines GM-CSF, TNF and IFN $\gamma$  and are present in the synovial sublining layer which positively correlate with disease activity<sup>100</sup>. Activated M1 macrophages are prolific producers of effector cytokines including TNF, IL-1 and IL-6 contributing to the chronic inflammatory environment within the joint and the ability to recruit more cells into the synovium<sup>101</sup>. M2 macrophage differentiation is driven by IL-4 and IL-13 and is thought to have an anti-inflammatory function characterised by its production of IL-10 and TGF- $\beta$ . There is no clear evidence to suggest that there is a presence of M1 or M2 macrophage phenotype in the joint however, abundant levels of TNF and IL-1 are found in the joint whereas IL-10 is relatively diminished in the joint<sup>102</sup>.

### **1.5.5 Innate and non-immune cell populations in RA pathology**

In addition to the above immune cell subsets which are implicated in RA, considerable evidence is emerging that innate immune cell populations are implicated in RA pathogenesis including dendritic cells (DCs), neutrophils and natural killer (NK cells)<sup>103</sup>. The concept of innate immunity is evolving and it is becoming increasingly evident that the binary distinction of adaptive and innate immune cell populations is a reductionist approach<sup>104</sup>. Innate immune cell populations are longer-lived than previously thought<sup>105</sup>, capable of memory<sup>106</sup> and also capable of immunoregulatory functions as well as proinflammatory functions<sup>107</sup>. In RA, innate cell populations are known to contribute to the excess inflammation and are capable of producing proinflammatory cytokines however, they have not been as extensively studied as their adaptive immunity counterparts, although interest in their pathologic contribute is increasing. In the subsequent chapters, the role of innate cell populations in RA will be elaborated upon in the context of their contribution to RA pathology.

FLS are not immune cells however, interact closely with other immune cell populations. In RA, FLS and macrophage-like synoviocytes are the two cell subsets that are present in the synovial intimal lining<sup>108</sup>. The synovial intimal lining sits on a delicate network of connective tissue which contains fibroblasts along with blood vessels and macrophages<sup>108</sup>. It is evident that within the rheumatoid joint, the number of FLS increases considerably contributing to the formation of the pannus, which is an abnormal layer of fibrovascular tissue within the synovial lining<sup>108</sup>. RA FLS are able to expand in the joint in response to the proinflammatory cytokines produced by immune cells and infrequent apoptosis<sup>108</sup>. FLS are capable of undergoing epigenetic reprogramming particularly by DNA methylation and it is thought that this process contributes to their pathological function in RA<sup>108</sup>. Fibroblasts will be reviewed in more detail further on in sections 1.8.2 and 1.8.3 of this chapter and in subsequent chapters their phenotypic subsets will be further embellished upon.

Finally osteoclasts are cells which play a key role in maintaining skeletal homeostasis by supporting steady-state bone remodelling in the bone marrow<sup>109</sup>. Osteoclasts are multinucleated cells derived from a myeloid lineage and are capable of differentiation in response to environmental cues and stimulation from M-CSF, RANKL and osteoprotegerin<sup>109</sup>. In RA pathology, osteoclasts are capable of

causing bone erosion through a combination of specific enzymes and an ATPase proton pump that enables degradation of bone matrix and solubilize calcium<sup>110</sup>. During RA pathology, osteoclasts are found in the pannus of the synovial intimal lining and link joint inflammation and structural damage by resorbing mineralised tissue adjacent to the joint resulting in tissue destruction<sup>110</sup>.

## 1.6 Treatment in RA

At present, methotrexate remains the cornerstone of RA treatment<sup>111</sup>. Methotrexate is most effective when prescribed at the earliest opportunity of disease however, compliance to methotrexate can diminish over time as patients experience adverse side effects<sup>112</sup>.

Both treatment options and treatment strategy in RA have evolved considerably over the last 30 years. Initially disease modifying anti rheumatic drugs (DMARDs) were considered to halt the progression of RA<sup>113</sup>. Examples of DMARDs include sulfasalazine, hydroxychloroquine, azathioprine and cyclosporin<sup>113</sup>. However, it became apparent that whilst DMARDs reduced symptoms and disease activity for some patients, it did not reverse joint damage<sup>113</sup>.

The advent of biologics has revolutionised treatment options in RA. Biologics are designed to target specific arms of the immune system. Cellular components of the immune system are targeted and the most effective have been treatment directed towards B cells and T cells. These treatments are known as rituximab (B cells) and abatacept (T cells)<sup>114</sup>. Cytokine therapy include TNF (tumour necrosis factor) and IL-6 (interleukin – 6) inhibitors. 5 classes of TNF inhibitors exist: etanercept, infliximab, adalimumab, certolizumab pegol and golimumab<sup>115</sup>. These different classes of TNF inhibitors target different epitopes of the TNF protein. IL-6 is another cytokine that has been targeted for treatment in RA and tocilizumab (IL-6 inhibitor) has been shown to be effective for many patients<sup>69</sup>. Small molecule inhibitors are a recent addition to biologic therapy. Baricitinib and tofacitinib target the Janus kinase (JAK) pathway but target different JAK molecules; baricitinib blocks JAK1 and JAK2 and tofacitinib blocks JAK1, JAK2 and JAK3<sup>114</sup>.

### 1.6.1 Remission in RA

Treat-to-target is the current treatment strategy which looks at selecting treatment which will result in low disease activity thus slowing disease progression and lead to remission<sup>116</sup>. If disease activity is not lowered, change to treatment should be made that will lead to low disease activity .

Remission is currently the gold standard for assessing treatment success as present treatment does not cure RA. Remission is defined as an 'absence of articular and extra-articular inflammation and immunologic activity related to RA'<sup>117</sup>. Attaining remission is currently expensive and not many patients can achieve this status without lifelong treatment. In 2010, a report by the National Rheumatoid Arthritis Society calculated the cost of treatment in RA to be £689 million in the United Kingdom (UK)<sup>118</sup>. Remission continues to be an active area of research in the management of RA. Encouragingly, more patients achieve remission status with the availability of targeted biologics but drug-free remission is significantly harder to achieve<sup>119</sup>.

For many patients, the prospect of lifelong treatment is undesirable and becomes an unsustainable approach when considering the cost of treatment, clinic time dedicated to reviewing patients and potential toxicity to the patient from continuous treatment. These reasons alone support the need for understanding drug-free remission which has been observed in 10-15% of patients<sup>120</sup>. Unfortunately, 50% of those patients who achieve drug-free remission relapse and begin to flare after the cessation of treatment<sup>120</sup>. Identifying a biomarker which would distinguish patients who do flare from those that do not flare after treatment withdrawal would prove to be invaluable and further refine the remission criteria. In 2019, Baker et al. set out to identify biomarkers which

distinguished patients likely to flare from those who did not after cessation of treatment <sup>121</sup>. Patients were deemed to be in remission by assessing their DAS28 and CRP which if less than 2.4 allowed them to qualify for remission <sup>121</sup>. In the CD4+ T cell population, two gene transcripts which were statistically significant identified patients who were likely to flare and one gene transcript was statistically significant for patients who did not flare. The cytokine IL-27 was associated with increased risk of flare after DMARD cessation <sup>121</sup>. Clinical factors also defined patients who were at a reduced risk of flare if they fulfilled the ACR/EULAR remission criteria at baseline, had a longer interval since their last change in DMARD therapy and had a longer disease duration <sup>121</sup>. Although the study consisted of a small cohort, and these results would need to be confirmed in larger cohorts, it strengthens the argument for inclusion of drug-free remission in remission guidelines <sup>122</sup>.

## **1.7 Conceptual frameworks aim to define stages of disease progression within RA**

### **1.7.1 Pre-clinical and early RA phase**

Within RA research, biologics have offered more precise treatment to ensure that not only the symptoms of RA are managed but the underlying chronic inflammation is targeted. In RA, the concept of early RA has received active research interest. Early RA aims to capture RA preferably before the condition manifests in the joint. Early RA was first put forward as a hypothesis at the beginning of the 1990s based on the existence of a time frame referred to as the 'window of opportunity' <sup>123</sup>. The window of opportunity is the concept where treatment is optimal resulting in long term benefits that closely resemble a 'cure' <sup>123,124</sup>. Before the concept of early RA was put forward, patients were historically treated by using the least effective and toxic drugs first and when joint erosion became visible, it was considered justifiable to administer DMARDs <sup>124</sup>. At the time, RA was thought to be a mild disease with slow progression but data from MRI now informs us that joint erosion can occur within weeks of symptom onset <sup>125</sup>. Duration of symptoms is regarded as the best response to therapy and evidence demonstrated that patients presenting with less than 1 year disease duration showed 53% patients responded to DMARD treatment which diminished as the duration of symptoms increased <sup>126</sup>.

Furthermore, conceptual frameworks exist to capture the progression of RA as it is no longer considered to be a uniform disease but more akin to a 'syndrome' due to its heterogeneous nature in both diagnosis and prognosis <sup>127,128</sup>. Currently there are six phases (A-F) defined in RA progression. Phases A and B refer to genetic and environmental risk factors for RA respectively, phase C refers to the systemic autoimmunity associated with RA, phase D refers to symptoms without clinical arthritis and phases E and F refer to undifferentiated arthritis and RA respectively <sup>129</sup>. These phases assist in identifying patients who are at the highest clinical risk of developing RA however, not all patients experience each phase. Treatment is observed to be most effective when given at the earliest opportunity and known as the 'window of opportunity' <sup>124</sup>. Considerable research effort has been directed towards the concept of early RA and what constitutes early RA as traditionally this was viewed as initial phase after arthritis became clinically detectable <sup>129</sup>. Recent studies have offered to challenge this presumption by suggesting that disease mechanisms become altered in the preclinical peripheral and by the time clinical arthritis is diagnosed, this is not the beginning of disease but rather the culmination of a plethora of well-established pathological events <sup>129</sup>. Ultrasound imaging in preclinical RA has shown that patients can display levels of sub-clinical inflammation in the at-risk individual <sup>130</sup>. Identifying biomarkers that could detect early RA and thus distinguish patients from

undifferentiated RA has been actively researched with several markers being reported including a type I interferon gene signature which distinguished DMARD-naïve patients who would go onto develop persistent inflammatory arthritis<sup>131-134</sup>. In patients with early inflammatory arthritis the gene *ST6GALNAC1*, an enzyme involved in glycosylation, could be a potential biomarker for undifferentiated arthritis compared to healthy controls<sup>131</sup>. In addition, *SIGLEC1* was also considered to be a useful marker of disease activity in undifferentiated arthritis<sup>131</sup>.

The lack of a robust clinical biomarker to identify early RA makes it difficult to precisely identify a time frame for early RA. However, researchers have made efforts to identify immune profiles which would differentiate early RA from other arthritides. Research published in 2005 looked at cytokine profiles of T cell and stromal cell origin in synovial fluid taken from 36 patients with early inflammatory arthritis and non-inflammatory synovial fluid were used as controls<sup>135</sup>. Within the first 12 weeks of symptoms onset, patients who subsequently went onto developing RA had a distinct and consistent synovial fluid cytokine profile<sup>135</sup>. The cytokines that were associated with early RA were IL-1, IL-2, IL-4, IL-13, IL-15 and IL-17 which were found to be expressed by T cells, stromal cells and macrophages<sup>135</sup>. This cytokine profile was specific to early RA synovial fluid and not observed in any other arthritides. Another study looked at CD4+ T cell subsets across the inflammatory arthritis continuum to see whether changes within the subsets could predict progression from one stage to the next<sup>136</sup>. Naïve CD4+ T cells predicted disease flare and disease progression along the whole inflammatory arthritis continuum suggesting that naïve CD4+ T cells can be used to distinguish individuals who are at high-risk of developing RA<sup>136</sup>. Lastly, a review aiming to define cytokine profiles across the stages of RA suggests that during the pre-clinical phase of RA, IL-17/IL-23, IL-8, IL-4, IL-5 and type I interferon cytokines are more prominent and may explain why not all patients respond well to TNF and IL-6 inhibitors as these cytokines become more prominent during early RA<sup>137</sup>.

## 1.7.2 Refractory RA

The prospect of numerous treatment options and in particular biologics has undeniably improved the clinical management of RA. Disappointingly, not all patients with RA respond successfully to treatment and it is common for patients to cycle through multiple rounds of different biologics, with the expectation that they can achieve remission<sup>138</sup>. These patients are at the opposite end of the spectrum to those with pre-clinical and early RA, and are termed as 'refractory RA' which at its broadest definition 'indicates the inefficacy of multiple agents in conjunction with unabating and systemic inflammation'<sup>139</sup>. A more refined definition currently suggests following methotrexate inefficacy, failure of two mechanistically different inhibitors however, at this stage remains speculative and open to further refinement in the presence of empirical data<sup>139,140</sup>. What precisely constitutes refractory RA or more specifically how to identify patients who are likely to develop refractory RA is not well understood, although it is estimated that approximately 6-17% of patients who are refractory<sup>138,139,141-143</sup>.

Whilst the definition for defining refractory RA remains dynamic, it is a concept that is gaining traction in the clinical setting to identify better treatment strategies as it currently remains an unmet clinical need<sup>143</sup>. For example, small molecule inhibitors such as JAK inhibitors have shown to be effective in a subset of patients who were refractory to previous multiple biologics<sup>141</sup>. Another small inhibitor, Evobrutinib which inhibits Bruton's tyrosine kinase, crucial for B cell development, failed to improve response rate in patients with refractory RA demonstrating that not all small molecule

inhibitors are successful in treating refractory RA as a result of its heterogeneity between patients<sup>144</sup>.

As little is known about refractory RA, simple questions such as whether it is driven by predominantly innate or adaptive immune cells is yet to be elucidated or whether it is driven by non-immunological pathways such as stromal cells<sup>138</sup>. It is argued however, that if refractory RA was driven by stromal cell pathology, a more consistent presentation would be observed whereas refractory RA is heterogeneous<sup>138</sup>. Finally, treatments that have historically not worked well in RA such as IL-1 blockade and IL-23-IL-17 inhibitors may be successful in treating refractory RA<sup>138</sup>.

## **1.8 The multi-omics era and revolution of single cell technologies**

The revolution in single cell technologies has provided an unprecedented platform for comprehensive profiling at the cellular and molecular level of translational research. This revolution has been informally coined the 'multiomics era' where biological tissue is subjected to in-depth interrogation of the genomic, epigenomic, metabolomic, transcriptomic and proteomic tissue landscape<sup>145</sup>. Multi-omic technology has been refined over time resulting in robust, high-resolution technologies which are capable of incorporating multiple parameters which can scale up the number of analytes to be analysed in a single cell<sup>146</sup>.

One single cell technology which has withstood the test of time and become a routinely used technique in cell biology and immunology research is flow cytometry which measures protein expression on single cells in solution. Antibodies which specifically recognise extracellular or intracellular proteins which are measured by a flow cytometer which utilizes lasers as light sources to produce both scattered and fluorescent light signals that are read by detectors such photodiodes or photomultiplier tubes<sup>147</sup>. These fluorescent signals are converted into electronic signals that measure and electronically record the fluorescence signals for the duration of acquisition which can subsequently be written into a standardised format (.fcs) data file<sup>147</sup>. Flow cytometry is primarily used for immunophenotyping immune and non-immune cells but can also be used for molecular assays measuring various cell functions including apoptosis, phagocytosis, phosphorylation, cell cycle analysis and this possibility continues to increase. Cell sorting is another main feature of flow cytometry which requires a specific cell sorting flow cytometer to selectively isolate a specific population of cells using a characterised immunophenotype. These cells are subsequently used in downstream in vitro experiments to determine the functional role of the specific cell population. All of these features of flow cytometry have made the technique versatile for numerous research questions concerning both pathology and health.

### **1.8.1 Challenges of compensation in flow cytometry**

Flow cytometry is widely regarded as a pioneering multiparameter single cell technology, however, a major limitation is spillover from one fluorescent channel into another channel. This fluorescence spillover can be corrected by applying a calculation which adjusts for spillover known as 'compensation'<sup>148</sup>. If compensation is not adjusted for, this can lead to inaccurate interpretation of the data known as a 'false positive' if a fluorescence signal is emitted due to spillover signal from another channel or a 'false negative' when a fluorescence signal is not recorded resulting in the interpretation that the protein is not present on the cell. Compensation can be calculated manually however, this approach is infrequently used as it is considered inaccurate. Automated compensation software to correct fluorescent spillover is present on most flow cytometers or incorporated in cytometry analysis softwares such as FlowJo where a compensation grid is calculated and the

fluorescent channels for each marker plotted against one another to visualise the effect of fluorescence on the population of cells to be analysed.

Designing a flow cytometry experiment requires significant time especially in the event of a new panel design which incorporates proteins which are not routinely studied by the researcher. Careful consideration should be applied from the beginning including the understanding of the cytometer the cells will be acquired on including the number of fluorescent channels that are present, the fluorescent dyes available for conjugating the antibodies and the number of proteins that are to be included in the panel design. Furthermore, based on whether proteins are extra/intracellular will require an understanding of whether cells need to be permeabilised and/or fixed will impact on compensation and requires optimisation prior to sample acquisition. Compensation can fluctuate between each sample and for each acquisition run meaning that controls need to be included each time. Controls for compensation include a negative and positive control for the protein marker and fluorescence minus one controls which include all the fluorescence colours that comprise the panel excluding one fluorophore to observe the staining signal when it is absent<sup>148,149</sup>. Incorporating these experimental controls becomes logistically challenging when the biological sample is limited. Moreover, successful compensation of a panel consisting of  $\geq 20$  protein markers can be challenging even for an experienced cytometrist. Thus deep immunophenotyping by flow cytometry is challenging and despite the multi-parameter capability, it is practically challenging to create protein marker panels in excess of 20 markers without encountering technical challenges. This limitation has restricted the scope of current flow cytometry immunophenotyping data in RA in terms of specific cell populations being interrogated which overlooks the heterogeneity of RA or limited in terms of the number of protein markers that can be included in one panel<sup>150</sup>. The advent of mass cytometry has provided an opportunity for deep immunophenotyping which overcomes the significant challenges presented by compensation in flow cytometry<sup>151,152</sup>.

### **1.8.2 Mass cytometry in discovery analysis**

Mass cytometry has provided an opportunity for both multi-parameter high-dimensional immunophenotyping and retaining high resolution making it a desirable technology to deep profile every single cell from a representative population of cells. Principally mass cytometry is similar to flow cytometry allowing for immunophenotyping single cells in suspension<sup>150</sup>. However, the important difference is that flow cytometry is a technology that measures protein expression by fluorescence and mass cytometry dispenses with fluorescence and instead antibodies are labelled with elemental (metal isotopes from the lanthanide series also known as rare earth metals) reporter ions via covalent conjugation with chelating polymers<sup>153</sup>. Samples are subsequently stained with these lanthanide-labelled antibodies and are introduced as single cell suspensions where they are nebulised into droplets which pass through an inductively coupled mass spectrometer which vaporizes, atomises, and ionizes the sample which means that the cells are no longer viable and after acquisition on a mass cytometer cannot be retrieved for further downstream analysis. The ion cloud created passes through quadrupoles which deflect biologically abundant low-mass ion species such as carbon, nitrogen and oxygen therefore selectively allowing high-mass ion lanthanide reporter masses to be recorded by time-of-flight mass spectrometry and measure the abundance of all reporter masses included in the panel<sup>153</sup>. These reporter masses can be visualised during acquisition to ensure that channels where a signal is expected can be detected. This recording of isotopic masses allows for bound antibody to be quantified which can be used in subsequent analysis to inform the expression of the markers of interest. At present, 37 protein markers can be interrogated in one panel by mass cytometry and it is anticipated this can be increased to 50

markers with intentions to explore new lanthanides and ultimately create panel in excess of 100 markers.

The use of metal isotopes to measure each antibody significantly reduces the challenges introduced by fluorescent spillover in flow cytometry however, careful consideration should be given to panel design to minimise the contamination of isotopic oxidation which can result in adjacent channels recording an unspecific signal. By strategically placing markers, such that proteins which have a stable expression e.g. lineage markers can be conjugated with isotopes occurring for the same metal as contamination into adjacent channels will be easier to correct. This is a straight-forward solution and considerably less problematic to solve compared to fluorescence compensation.

The amount of publications reporting data from mass cytometry has increasingly grown as accessibility to the technology widens. It should be mentioned that although the number of publications have increased, these are often from research groups working in collaboration with other groups/departments to generate and analyse the high-dimensional data. It is important to acknowledge the differences and hence the suitability of mass cytometry in experiments compared to flow cytometry. These include a maximum of 500 events per second can be acquired by mass cytometry compared to several thousand events per second in flow cytometry<sup>154</sup>. The running of a mass cytometry experiment is considerably more expensive than flow cytometry with Fluidigm currently being the only vendor supplying reagents and instrumentation<sup>155</sup>. Lastly, mass cytometry is a commitment as it requires specific facilities adapted to suit the purpose of the technology including gas/liquid argon which delivers the sample to the plasma, calibration of the mass cytometer daily takes between 1.5-2 hours prior to acquisition and dedicated personnel to maintain and operate the machinery<sup>156</sup>.

To date, much of the immunophenotyping analysis in RA has been performed by flow cytometry. Opportunity for in-depth immune profiling with improved precision single-cell technologies has allowed for old questions to be revisited with better insight. Mass cytometry has provided the opportunity for researchers to either probe specific cell populations more intently or to look at a heterogeneous population of cells within a given biological sample. Currently, RA research has fully embraced mass cytometry as an opportunity to answer some of the unmet clinical needs which exist with particular aim towards understanding if there are specific immunophenotypes or immune signatures that correlates with treatment response or disease stage. These questions are broad and are further tailored to the cohort and samples collected.

The use of mass cytometry to understand RA has received considerable attention and publications over the last 5 years have steadily increased which have revealed novel subpopulations of immune cells which are expanded in RA compared to health and inflammatory arthritis which is not RA. The expanded immune cell populations identified are further detailed in chapters 5 and 6 however, a brief overview will be provided here to put the hypotheses and aims of this study into context. Pathological expansion of peripheral helper T cells (characterised as CD4<sup>+</sup>PD-1<sup>hi</sup>CXCR5<sup>+</sup>) discovered by the Accelerating Medicines Partnership (AMP) group where they identified a population of T cells functionally characterised as B helper T cells which are not functionally exhausted despite high PD-1 expression but instead displayed features of chronic activation<sup>157-159</sup>. Another subset of CD4 T cells characterised as CD27<sup>-</sup> and HLA-DR<sup>+</sup> have also been reported to be aberrantly expanded in RA peripheral blood by mixed effects associated single cell modelling and also had features of chronic activation<sup>160</sup>. Analysis of 51 samples of synovial tissue from patients with RA or osteoarthritis detected specific cell populations that were associated with driving joint inflammation in RA including an expansion of THY1<sup>+</sup> HLA-DRA<sup>hi</sup> sublining fibroblasts which demonstrated that stromal

cell populations and subsets can be detected by mass cytometry <sup>161</sup>. 17 unique cell populations were additionally identified by transcriptomics which included *IL-1B+* pro-inflammatory monocytes, autoimmune-associated B cells and CD8 T cell subsets with increased cytolytic function <sup>161</sup>. These cell populations will be discussed in more detail in terms of their phenotype and functional characteristics however, it does demonstrate the value of mass cytometry in revealing subpopulations that contribute to RA pathogenesis.

Mass cytometry remains however, a technology in its infancy and is not routinely used unlike flow cytometry. This could be viewed as a disadvantage but it is already apparent from the conception of this study, that accessibility to mass cytometry is improving and an advantage of this is that protocols can be optimised and refined to enable best practice across all research groups and improve replicability of data. The findings above however, are preliminary in terms of their impact. It is not clear for example to what extent peripheral helper T cells are pathologically implicated in RA and how existing treatments affect this population if at all. Large patient cohorts are not currently acquired by mass cytometry and this can be attributed to both the logistical practicality and the complex data analysis required following acquisition. In particular, mass cytometry data analysis has advanced considerably in the last 5 years and existing approaches such as manual gating and two dimensional clustering visualisation plots <sup>162,163</sup> have been surpassed by approaches which can both cluster and statistically quantify cell populations across defined conditions <sup>164–166</sup>. It is here that this study will aim to bridge the gap between manual and automated gating by using the novel bioinformatics approaches for data analysis to interrogate the immune cell populations in RA peripheral blood and synovial fluid. It also aims to be one of the first studies to acquire a patient cohort of over 150 samples by mass cytometry using a clinically translatable protocol. It is hoped that the work presented here will instil confidence for future researchers to pursue large immunophenotyping studies with an automated bioinformatics analysis pipeline which can expedite findings from bench to clinic.

### **1.8.3 Synovial pathotypes aim to enhance RA classification and treatment approach**

The above studies have extended the discussion about the pathological involvement of cell populations in RA. The single cell technology revolution has enabled us to identify specific subsets within large cell populations such as macrophages, neutrophils, fibroblasts and T and B cells and in addition, specific phenotypes are associated with functional status of the cells. The AMP group and collaborators have made extensive progress in identifying pathologically expanded populations in RA however, in addition to these findings, the stratification of synovial tissue pathotypes has also developed significantly <sup>167</sup>. The three pathotypes that have been consistently identified are lymphoid, myeloid and pauci-immune. Pauci-immune refers to very minimal/absent infiltrating immune cells but presence of a fibroblast phenotype <sup>167,168</sup>.

Based on this stratification, further evidence has shown that these findings correspond with stages of RA. In a cohort of 144 treatment naïve early RA patients, it was demonstrated that three pathotypes exist prior to treatment intervention or further clinical progression which include: lympho-myeloid which is dominated by the presence of B cells as well as myeloid cells, diffuse-myeloid which is characterised by the presence of myeloid lineage cells but poor B cell presence and pauci-immune which demonstrates a lack of immune cells but rich in stromal cells <sup>169–171</sup>. Lymphoid-rich patients also showed high expression of myeloid genes when analysed by Nanostring technology, leading to the term 'lympho-myeloid' <sup>170</sup>. The advantage of this stratification is that

multiple clinical parameter analyses can be performed to determine whether there is a correlation between immune pathotype and clinical outcome. Analysis showed that elevated expression of myeloid and lymphoid genes strongly correlated with worsening outcomes including osteoclastic radiographic joint progression, disease activity, acute phase reactant and autoantibody positivity whereas pauci-immune pathology showed less severe disease activity and radiographic progression<sup>170</sup>. Conversely, in a cohort of 37 RA patients, synovial biopsies were obtained from patients pre-and post TNF blockade treatment using certolizumab pegol where it was observed that those with a lymphoid-myeloid pathotype at baseline, achieved better clinical response compared to those with a pauci-immune pathotype suggesting that these pathotypes are a continuous rather than a discrete distribution<sup>172,173</sup>.

Dennis et al., in 2014 set out to observe whether synovial pathotypes could be identified by specific biomarkers which would also correlate with treatment response<sup>172</sup>. Using a combination of synovial tissue histopathology, transcriptional profiling and serology they found increased levels of soluble ICAM1 in the serum which associated with a synovial myeloid cell pathotype and predicted good response to adalimumab, whereas the chemokine CXCL13 associated with a synovial lymphoid pathotype and predicted good response to tocilizumab<sup>172</sup>.

Therefore, it is evident that single cell technologies have been richly utilised within the RA field and it is anticipated that over time, this will only increase. Both transcriptomic and cytometric approaches have revealed specific immune cell subsets which have a pathologic role in RA. The complex nature of the data means that appropriate analysis approaches need to be applied to achieve the most meaningful analysis from high throughput technology. Although the field of RA has returned to the synovium, blood is easier to obtain and can be more extensively explored with mass cytometry. Furthermore, the immunophenotyping panel in this study is deliberately broad, to capture numerous immune cell populations to understand not only the subsets within the populations but also assess the relationship between different cell populations.

## **1.9 Study hypotheses and aims**

### **1.9.1 Technical hypotheses**

- 1. Mass cytometry is a novel single cell technology yielding reliable, reproducible data for 37 protein markers in a large patient cohort immunophenotyping study.**

Aim 1: To validate antibody clones to be used in mass cytometry by flow cytometry to ensure the staining quality is comparable.

Aim 2: To validate metal lanthanide conjugated by staining beads to ensure that the metal conjugation has covalently attached and that a signal can be read in the appropriate channel on a mass cytometer. 37 protein markers will be titrated to ensure optimal staining by mass cytometry.

Aim 3: To ensure that reliable data is obtained by mass cytometry, patient samples will be processed using the same fixation method immediately upon collection and will simultaneously be barcoded and stained by the 37-antibody marker panel to ensure staining consistency. Inclusion of a batch control will ensure any staining inconsistencies across the batches can be detected.

- 2. Novel automated bioinformatics pipelines produce comparable data to manual gating and expedite multi-parametric, high-dimensional single cell protein analysis.**

Aim 1: To apply an automated bioinformatics script for data clean-up to obtain a single cell population of cells across all samples included in analysis.

Aim 2: Identify a robust mass cytometry data analysis pipeline which can statistically quantify the cell populations between conditions and can be used for discovery analysis. Clustering algorithms will be validated for consistent cell populations being identified across conditions for comparative analysis.

Aim 3: This study will aim to identify the two CD4+ T cell populations which have recently been reported to be pathologically expanded in RA by mass cytometry. Ensuring replicability of data by mass cytometry will increase confidence and lend more weight to investigating specific cell populations further.

## **1.9.2 Hypothesis to interrogate the immune landscape in RA peripheral blood and synovial fluid**

### **3. Immune dysregulation occurs in early RA and specific immune cell populations in peripheral blood can distinguish this phenomenon from health.**

Aim: To compare immune cell populations in peripheral blood from patients clinically defined as 'early' RA to healthy donors to determine if there are specific changes that distinguish early RA from health.

### **4. Innate immune cell populations in peripheral whole blood predominantly drive immune dysregulation in early RA in comparison to health.**

Aim 1: To establish whether innate immune cell populations become dysregulated and drive the inflammation observed in early RA or whether this is still largely dominated by adaptive immune cell subsets.

Aim 2: To determine whether neutrophil populations are altered in early RA compared to health.

### **5. Biological microenvironment shapes the immunophenotype of cell populations revealing specific functional adaptation or response to stimuli within the microenvironment.**

Aim: To compare cells in the joint microenvironment, synovial fluid obtained from patients with RA will be collected in conjunction with matched peripheral blood to compare the influence of microenvironment on immune cell populations including whether T cells are chronically activated in the synovial fluid microenvironment.

The structure of this dissertation will focus on systematically demonstrating the validation and optimisation of using mass cytometry. This will include a step wise description of the methodology in chapter 2, where the main aim is to provide a replicable protocol for future users. This includes cohort selection and criteria, designing a protein panel for mass cytometry and preparing and acquiring data on the mass cytometer, Chapter 3 focusses on the validation of the antibodies used for mass cytometry validated using flow cytometry (an established technique) and mass cytometry to ensure that downstream analyses and conclusions drawn are robust. Chapter 4 will discuss the bioinformatics pipeline used to analyse the data acquired from mass cytometry and thus demonstrate a novel approach to this research. Chapter 5 will look at the immune profiles in whole blood across the different stages of RA using the bioinformatics pipeline described in chapter 4. Chapter 6 aims to explore the immune profiles between synovial fluid and whole blood obtained

from 3 patients with RA to compare the biological microenvironment. Furthermore, synovial fluid offers the opportunity to observe immune cell populations in RA at its primary site of pathology.

## Chapter 2 Materials & Methods

### 2.1 Patient cohort

The intention of this project originally set out to comprehensively immunophenotype synovial biopsy tissue obtained from patients with RA. Currently the field of mass cytometry has been exploited in this way by other research groups. However, it became logistically challenging to obtain synovial biopsy tissue at Chapel Allerton Hospital. Furthermore, if tissue had been possible to obtain, this would have required extensive validation in terms of processing the tissue to obtain optimum cellular yield and validation in processing tissue for mass cytometry. Whilst other groups have successfully developed protocols for synovial biopsy digestion for single cell suspension, and subsequently used mass cytometry to immunophenotype these cells, better techniques exist. The Helios Mass cytometer requires cells to be in suspension and thus disaggregating cells from a tissue biopsy disrupts the architecture and questions determining origin of cells (resident vs. infiltrating) remain unanswered.

The Hyperion imaging mass cytometer would be a better tool for tissue and whilst utilised in cancer biomarker screening from tumour biopsies, imaging mass cytometry remains to be utilised in RA synovial tissue research. Unfortunately, at the time of this research, the Hyperion imaging platform was not available at UCB. Indeed, as the field of tissue research expands, technology continues to improve and provide more options. Examples of this include the CODEX (Co-Detection by inDEXing) technology provided by Akoya Biosciences, which is a multiplexing imaging technology that was originally developed in Garry Nolan's research group at Stanford University. The technology uses antibodies conjugated to oligonucleotides which allows for deep immunophenotyping of more than 40 markers in a single tissue sample<sup>174</sup>. RNA sequencing and in situ hybridisation have also been extensively reported by groups although these techniques do not focus on proteins. It remains clear however, that the questions to be answered must be considered in accordance with the available resources and logistics. Disaggregation of tissue is undesirable, and the field is moving from basic histology to an automated, precision platform.

Therefore, blood samples were collected from patients with RA. Blood samples are easier to scale up over a short time period and less invasive than biopsy samples. It is also possible to obtain healthy control samples to control for findings which can be challenging when working with tissue. Blood samples are still pertinent to RA research as they represent the systemic environment of the disease. Where possible synovial fluid was collected from large joint aspirations as a way of understanding the joint microenvironment. However, presently in clinics at Chapel Allerton Hospital, joints are infrequently aspirated in clinic due to better treatments available resulting in reduced swelling and inflammation.

#### 2.1.1 Ethical approval for patient biological samples

Ethical approval was obtained prior to the commencement of patient recruitment and biological sample collection for this project. Ethical approval was obtained from Leeds Teaching Hospitals Trust Research and Development for the study RADAR (Rheumatoid Arthritis Disease Research), sponsor identification number RR09/9134. Patients and healthy donors were also consented to RADAR BMC (biologics monitoring clinic) which is a sub-study of RADAR. All participants in this study read the patient information sheet and written informed consent was given prior to sample collection. Samples were collected from Chapel Allerton Hospital, Leeds and processed and stored at the

Wellcome Trust Brenner Building at the St. James' Hospital site, Leeds. Ethics allowed for the transportation of samples between Chapel Allerton Hospital and St. James' Hospital and covered processing and storage of biological material at St. James' Hospital. The ethics also encompassed transportation, processing, storage and data collection of samples at UCB Pharma, Slough for research and development purpose. Samples were stored at UCB for longer than two weeks and were stored according to the Human Tissue Act regulation. Training for Human Tissue Act regulation were undertaken at Leeds and UCB according to in-house guidelines.

### **2.1.2 Patient recruitment criteria**

Patients were recruited to this study from two different Rheumatology clinics at Chapel Allerton Hospital, Leeds and a summary of the cohort is shown in Flow chart 2-1.

A power calculation was not performed before patient recruitment. The justification for this approach is that prior to embarking on this work, publications using mass cytometry reported results using a maximum of ten patient samples or less. This can be directly attributed to the novelty of the technology in which validation and protocol development takes considerable time and therefore a patient cohort in excess of 10 patients would be considered a risk. Therefore, this work both acknowledged the limitations of current findings in the literature and exceeded the number of samples processed and analysed by mass cytometry which included patients that would represent the broad spectrum of RA including treatment, treatment timepoints and disease stages. It was not possible to calculate a power calculation a priori to the study due to the uncertainty of how many patients could be recruited within the 9 months, which was the recruitment time frame for this study.

Inclusion criteria for patients recruited for this study were recruited from the RADAR BMC and had been diagnosed with RA for over 12 months, were over the age of 18 years old and had received either DMARDs or biologic treatment and none of these patients were treatment naïve. These patients are referred to as having established RA. Table 2-1 summarises this patient cohort. Single time point peripheral blood samples were obtained from 55 patients. Single time point means that only one sample was collected from the patient at one time point with no follow up sample. This single time point was not identical for each patient. For some patients, it was only possible to obtain a sample at a 'baseline' time point. Baseline time point means that a patient is not currently receiving treatment. This does not mean that the patient has never received treatment; it means that the patient is currently undergoing a 'wash out' phase. The 'wash out' is a window of time between the previous treatment and the new treatment which the patient will commence which was defined as 4 weeks. This 'wash out' phase is important allowing for the previous drug and associated compounds to be removed from the biological system before commencing new treatment. The baseline time point should ideally be obtained for each patient to understand the immune system of that individual which is thought to represent the immune system as close to its 'natural' state although this is not strictly true as patients have been on prior treatment; the long-lasting effects of which are not fully understood. Where possible single time points were taken at baseline, but it was not logistically possible to obtain this time point for all patients. Therefore, for some of the patients, the sample was collected at either three months or six months after they have received the new treatment. The RADAR BMC clinic oversees patients who are switching from one biologic to another biologic. In addition, there are patients who are receiving biologics for the first time as well as patients who would be currently defined as having refractory RA. In addition to the patient cohort, 23 peripheral blood samples from healthy individuals were recruited for this study at

a single time point. These healthy individuals did not have a diagnosis of RA as defined by the current classification system and were not on treatment for another condition that overlapped with those used in this RA cohort.

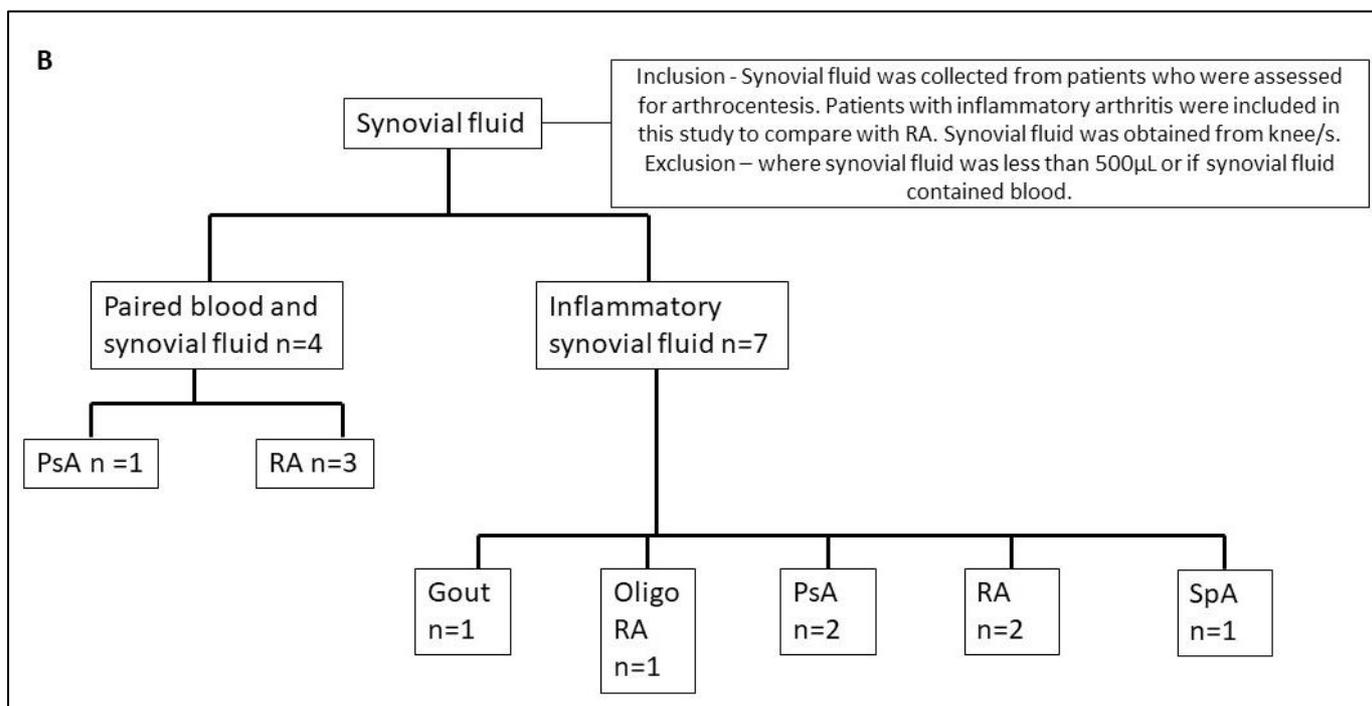
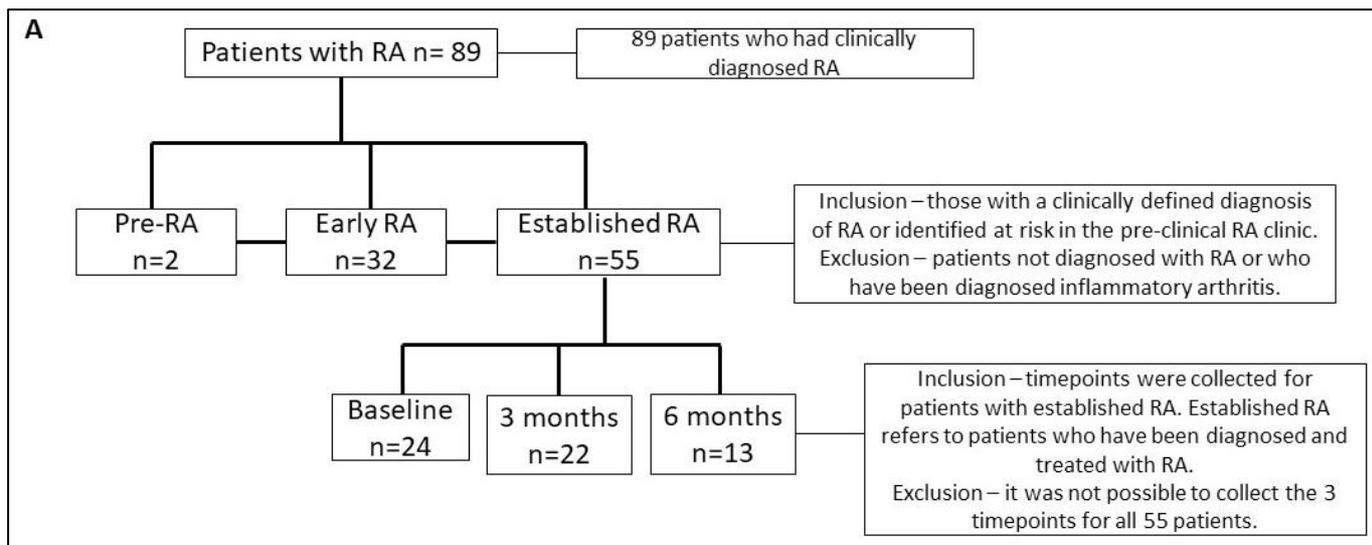
It was possible to recruit an additional cohort of patients from RADAR BMC where a baseline time point and a follow up time point at either 3 or 6 months and where possible both follow up time points were recruited (Table 2-2).

Thirty-two patients with early RA were recruited from the early RA clinic and had been newly diagnosed (less than 12 months) and not received treatment.

Two preclinical RA individuals were recruited from the preclinical RA screening clinic. These individuals did not have diagnosed RA or have symptoms of RA but they have been screened based on risk factors including ACPA/ RF titre.

As synovial biopsies were not logistically possible to obtain, synovial fluid was the closest option to observing the immune system in the joint microenvironment. Due to improved treatment and better treatment outcomes, swelling in larger joints, particularly in RA are not routinely observed in the Chapel Allerton Hospital arthrocentesis clinic. This is an excellent outcome for both clinicians and patients and a testimony to the high level of care and treatment provided by the Rheumatologists at Leeds. Although occurrence was infrequent, where possible, synovial fluid was collected from patients with inflammatory arthritis which would form a modest sized cross-sectional cohort. In total, 11 synovial fluid samples were obtained from the clinic and of these, 4 paired samples with peripheral whole blood were collected to compare synovial fluid mononuclear cells (SFMCs) and peripheral blood microenvironments from the same patient. Table 2-3 shows the pathology from which the synovial fluid (and paired samples) were obtained from.

The experimental design is represented by this wide-ranging cohort, spanning disease duration, time points, treatment and in addition the clinical parameters associated with each patient. This cohort was considered to best represent and answer the hypotheses within this project acknowledging the heterogeneity of RA.



**Flow chart 2-1| Inclusion and exclusion criteria for patient cohort**

**A** Inclusion and exclusion criteria for peripheral blood collected from patients with RA and **B** shows inclusion and exclusion criteria for both paired blood and synovial fluid cohort or unpaired synovial fluid from inflammatory arthritis.

**Table 2-1| Number of healthy and patient peripheral blood samples obtained at a single time point**

Cohort	Number (n)
Healthy	n=23
Pre-Rheumatoid Arthritis	n=2
Early Rheumatoid Arthritis	n=32
Established Rheumatoid Arthritis	n=55

**Table 2-2| Number of established Rheumatoid Arthritis longitudinal peripheral blood samples**

Cohort	Time point	Number (n)
Established RA longitudinal cohort	Baseline	n=24
	3 months	n=22
	6 months	n=13

**Table 2-3| Number of paired peripheral blood and synovial fluid samples and single synovial fluid samples obtained from inflammatory arthritis**

Cohort	Disease	Number (n)
Paired peripheral blood and synovial fluid	Psoriatic Arthritis	n=1
	RA	n=3
Inflammatory arthritis (only synovial fluid)	Gout	n=1
	Oligo RA	n=1
	Psoriatic Arthritis	n=2
	RA	n=2
	Spondyloarthritis	n=1

### 2.1.3 Collection and biobanking of whole blood samples for mass cytometry

Samples were collected 2-4 hours after collection and processed using a method developed by UCB Pharma, Slough to maintain consistent processing across all samples. This procedure was performed in a category II tissue culture hood using sterile, aseptic techniques. In a 50mL falcon tube 5mL of 16% paraformaldehyde (PFA) was added to 40mL of PBS supplemented with 1mM EDTA (Ethylenediaminetetraacetic acid) thus creating a 1.6% PFA solution. 5mL of blood was then added to the 1.6% PFA solution and mixed thoroughly by inversion and left to stand for 5 minutes at room temperature. Samples were then centrifuged at 400g for 7 minutes at room temperature (RT) and the supernatant discarded to leave a solid red blood pellet. This was resuspended in PBS/1mM EDTA and thoroughly mixed and centrifuged at 500g for 7 minutes at RT. The supernatant was removed by using a 25mL stripette and the final small volume removed by a 1mL pipette tip to leave a semi solid blood pellet that was resuspended in 10% DMSO (dimethyl sulfoxide)/90% PBS/1mM EDTA. 1mL of blood was aliquoted into cryovials for cryogenic preservation at -80°C in a CoolCell© freezing box. After cells were frozen for 2 hours they were transferred into a storage box and kept at -80°C.

#### 2.1.4 Collection and preparation of synovial fluid samples for mass cytometry

Where possible, matched synovial fluid with peripheral blood or just synovial fluid was collected from patients undergoing arthrocentesis. The fluid was centrifuged at 400g for 5 minutes and the synovial fluid was filtered through a 70µM filter to remove non-cellular material. The pelleted synovial fluid cells were fixed in 1.6% PFA as described in section 2.1.3.

## 2.2 Materials

**Table 2-4| General equipment and plastic ware**

Equipment	Company
Flow Cytometer Mass cytometrylex S	Beckman Coulter
Mass Cytometer Mass cytometry 2	Fluidigm
Laminar Flow Class II Tissue culture hood	Nuaire
Swing bucket centrifuge	Eppendorf
37°C bead bath	SHEL LAB
37°C/5% CO <sub>2</sub> cell culture incubator	-
Automated cell counter	Sysmex
Haemocytometer	Thermo Fisher Scientific
0.5-10µL, 10-20µL, 20-200µL, 200-1000µL pipettes	Eppendorf
Finnpipette™ F2 Multi-channel pipette 30µL - 300µL	Thermo Fisher Scientific
Serological pipette S1 Pipet fillers	Thermo Fisher Scientific
Nanodrop	Thermo Fisher Scientific
CoolCell freezer box	Biocision
Scotsman Laboratory Ice machine	Progen Scientific
Human Tissue Act approved -80°C freezer storage	-
±4°C Refrigerator	-

-20°C freezer	-
---------------	---

Plastic ware	Company
5mL, 10mL and 25mL stripettes	Fisher Scientific
5mL polystyrene FACS tubes	Fisher Scientific
5mL polypropylene FACS tubes	Fisher Scientific
5mL polystyrene filter cap FACS tubes	Fisher Scientific
15mL falcon tube	Fisher Scientific
50mL falcon tube	Fisher Scientific
12 well cell culture plates	Corning, Fisher Scientific
10, 20, 200, 1000µL pipette tips	Rainin
Pasteur pipettes	VWR
1.5mL Superlock Microcentrifuge tube	Starlab
1.5mL Eppendorf	Fisher Scientific
0.5mL Eppendorf	Fisher Scientific
1.5mL Cryogenic tubes	Corning
Freezer boxes	Camlab
70µM cell strainer	Fisher Scientific
Nunc® MicroWell™ 96-Well Plates, Polypropylene, High Volume (used in this work for Mass cytometry barcoding)	Thermo Fisher Scientific

## 2.3 Reagents

**Table 2-5 | General reagents, company and catalogue number**

Reagent	Company	Catalogue number
Maxpar Cell acquisition solution buffer	Fluidigm	201240
Cell-ID™ Intercalator-Ir—125 µM	Fluidigm	201192A
Deionized water	-	-
EQ™ Four Element Calibration Beads	Fluidigm	201078
Gibco Phosphate buffered Saline Solution, pH 7.4	ThermoFisher Scientific	10010-023
DMSO	Sigma	D2650-100ML

EDTA	ThermoFisher Scientific	15575020
Fetal bovine serum (FBS)	ThermoFisher Scientific	10100147
L-glutamine 200mM	ThermoFisher Scientific	25030081
Lymphoprep	Stemcell Technologies	07861
autoMACS Running Buffer – MACS Separation buffer	MACS Miltenyi Biotec	130-091-221
Maxpar Cell Staining buffer	Fluidigm	201068
Maxpar Water	Fluidigm	201069
Methanol	Merck	179337-1L
OneComp ebeads™ Compensation beads	ThermoFisher Scientific	01-1111-41
Pierce™ 16% Formaldehyde (w/v), Methanol-free	ThermoFisher Scientific	28908
Penicillin-Streptomycin	Merck	P4333-100ML
Fixation/Permeabilisation Solution kit	BD Biosciences	554714
RPMI 1640 media	ThermoFisher Scientific	11875093
Trypan blue	Sigma Aldrich	93595
Viability fixable stain 780	BD Horizon	565388

### 2.3.1 Immunophenotyping panel for mass cytometry

**Table 2-6| Metal lanthanides abbreviation for mass cytometry**

Periodic table abbreviation for lanthanide metal isotope	Lanthanide metal isotope
170Er	Erbium-170
145Nd	Neodymium-145
146Nd	Neodymium-146
209Bi	Bismuth-209
166Er	Erbium-166
160Gd	Gadolinium-160
148Nd	Neodymium-148
142Nd	Neodymium-142
149Sm	Samarium-149
167Er	Erbium-167
158Gd	Gadolinium-158

172Yb	Ytterbium-172
162Dy	Dysprosium-162
89Y	Yttrium-89
169Tm	Thulium-169
171Yb	Ytterbium-171
173Yb	Ytterbium-173
176Yb	Ytterbium-176
168Er	Erbium-168
151Eu	Europium-151
163Dy	Dysprosium-163
147Sm	Samarium-147
175Lu	Lutetium-175
161Dy	Dysprosium-161
156Gd	Gadolinium-156
165Ho	Holmium-165
153Eu	Europium-153
174Yb	Ytterbium-174
154Sm	Samarium-154
155Gd	Gadolinium-155
144Nd	Neodymium-144
152Sm	Samarium-152
159Tb	Terbium-159
150Nd	Neodymium-150
141Pr	Praseodymium-141
143Nd	Neodymium-143
164Dy	Dysprosium-164

**Table 2-7 | Mass cytometry immunophenotyping panel for this study**

Antibody	Clone	Host, Isotype	Metal conjugation	Company	Catalogue number	Dilution
CD3	UCHT1	Mouse IgG1, κ	170Er	Biolegend	300443	1:100
CD4	RPA-T4	Mouse IgG1, κ	145Nd	Biolegend	300541	1:100
CD8	RPA-T8	Mouse IgG1, κ	146Nd	Biolegend	301053	1:100
CD11b	ICRF44	IgG1	209Bi	Fluidigm	3209003B	0.5:100

CD11c	3.9	Mouse IgG1, κ	166Er	Biolegend	301639	1:100
CD14	61D3	Mouse IgG1, κ	160Gd	Thermo Fisher Scientific	14-0149-80	1:100
CD16	3G8	Mouse IgG1, κ	148Nd	Biolegend	302051	1:100
CD19	H1B19	Mouse IgG1, κ	142Nd	Biolegend	302247	1:100
CD25	M-A251	Mouse IgG1, κ	149Sm	Biolegend	356102	2:100
CD27	O323	Mouse IgG1, κ	167Er	Biolegend	302839	1:100
CD28	CD28.2	Mouse IgG1, κ	158Gd	Biolegend	302902	1:100
CD38	HIT2	Mouse IgG1, κ	172Yb	Biolegend	303535	1:100
CD40	HB14	Mouse IgG1, κ	162Dy	Biolegend	313002	1:100
CD45	HI30	Mouse IgG1, κ	89Y	Fluidigm	304045	0.5:100
CD45RA	HI100	Mouse IgG2b, κ	169Tm	Biolegend	304143	1:100
CD45RO	UCHL1	Mouse IgG2a, κ	171Yb	Biolegend	304239	1:100
CD56	CMSSB	Mouse IgG1, κ	173Yb	Thermo Fisher	14-0567-82	1:100
CD80	MEM-233	Mouse, IgG1	176Yb	Biorad	MCA2071	1:100
CD86	FUN-1	Mouse BALB/c IgG <sub>1</sub> , κ	168Er	BD Biosciences	555655	1:100
CD123	6H6	Mouse IgG1, κ	151Eu	Biolegend	306027	1:100
CD127	A019D5	Mouse IgG1, κ	163Dy	Biolegend	351337	1:100
CD138	MI15	Mouse BALB/c IgG <sub>1</sub> , κ	147Sm	BD	551902	1:100
CD152 (CTLA-4)	L3D10	Mouse IgG1, κ	175Lu	Biolegend	349902	1:100
CD161	DX12	Mouse, C3H/Bi IgG <sub>1</sub> , κ	161Dy	BD	556079	1:100
CD203c	NP4D6	Mouse IgG1, κ	156Gd	Biolegend	324602	1:100
FCεRIα	AER-37 (CRA-1)	Mouse IgG2b, κ	165Ho	Biolegend	334602	1:100

Granzyme B	GB11	Mouse, IgG1	153Eu	Biorad	MCA2120	1:100
HLA-DR	L243	Mouse IgG2a, κ	174Yb	Biolegend	307651	1:100
IgD	IA6-2	Mouse IgG2a, κ	154Sm	Biolegend	348202	1:100
NKp44	P44-8	Mouse IgG1, κ	155Gd	Biolegend	325102	1:100
NKp46	9-E2	Mouse BALB/c IgG <sub>1</sub> , κ	144Nd	BD Biosciences	557911	1:100
PD-1	EH12.2H7	Mouse IgG1, κ	152Sm	Biolegend	329941	1:100
PDL1	MIH1	Mouse IgG1, κ	159Tb	Thermo Fisher	14-5983-82	1:100
Perforin	dG9	Mouse IgG2b, κ	150Nd	Biolegend	308102	1:100
TNFR2	3G7A02	Rat IgG2a, κ	141Pr	Biolegend	358402	1:100
Va7.2	3C10	Mouse IgG1, κ	143Nd	Biolegend	351702	1:100
γδTCR	B1	Mouse IgG1, κ	164Dy	Biolegend	331202	1:100

### 2.3.2 Immunophenotyping panel for flow cytometry

**Table 2-8| Mass cytometry immunophenotyping panel for this study**

Antibody	Clone	Host, Isotype	Fluorescence conjugation	Company	Catalogue number	Dilution
CD3	UCHT1	Mouse BALB/c IgG <sub>1</sub> , κ	Brilliant Violet 421	BD Horizon	562426	1:100
CD4	RPA-T4	Mouse IgG <sub>1</sub> , κ	APC/Cy7	Biolegend	300517	1:100
CD8	RPA-T8	Mouse IgG <sub>1</sub> , κ	Alexa Fluor 700	BD Pharmingen	561453	1:100
CD11b	D12	Mouse BALB/c IgG <sub>2a</sub> , κ	PE	BD Biosciences	347557	1:100
CD11c	3.9	Mouse IgG <sub>1</sub> , κ	FITC	Biolegend	301603	1:100
CD14	M5E2	Mouse IgG <sub>2a</sub> , κ	Alexa Fluor 700	BD Pharmingen	557923	1:100
CD16	3G8	Mouse IgG <sub>1</sub> , κ	Pacific blue	Biolegend	980106	1:100

CD19	HIB19	Mouse IgG1, κ	Brilliant Violet 421	BD Horizon	562440	1:100
CD25	M-A251	Mouse BALB/c IgG <sub>1</sub> , κ	PE	BD Pharmingen	555432	1:100
CD27	O323	Mouse IgG1, κ	PE	Biolegend	302807	1:100
CD28	CD28.2	Mouse IgG1, κ	PE	Biolegend	302901	1:100
CD38	HIT2	Mouse IgG1, κ	PE	Biolegend	303505	1:100
CD40	HB11	Mouse IgG1, κ	PE	Biolegend	313005	1:100
CD45	HI30	Mouse IgG1, κ	PECy7	BD Pharmingen	557748	1:100
CD45RA	HI100	Mouse IgG2b, κ	PC5.5	Biolegend	304110	1:100
CD45RO	UCHL1	Mouse IgG2a, κ	PE	Biolegend	304205	1:100
CD56	CMSSB	Mouse IgG1, κ	APC	Thermo Fisher Scientific	17-0567-42	1:100
CD80	MEM-233	Mouse, IgG1	APC	Thermo Fisher Scientific	A15707	1:100
CD86	FUN-1	Mouse BALB/c IgG <sub>1</sub> , κ	Brilliant Violet 510	BD Horizon	563461	1:100
CD123	6H6	Mouse IgG1, κ	FITC	Biolegend	306013	1:100
CD127	R34.34	Mouse, IgG1	PE	Novus Biologicals	DDX0700P-100	1:100
CD138	MI15	Mouse IgG1, κ	PE	Biolegend	356503	1:100
CD152 (CTLA-4)	L3D10	Mouse IgG1, κ	PE	Biolegend	349905	1:100
CD161	DX12	Mouse, C3H/Bi IgG <sub>1</sub> , κ	PE	BD Pharmingen	556081	1:100
CD203c	NP4D6	Mouse IgG1, κ	Alexa 647	Biolegend	324625	1:100
FCεR1α	AER-37 (CRA-1)	Mouse IgG2b, κ	FITC	Biolegend	334607	1:100
Granzyme B	GB11	Mouse, IgG1	PE	ThermoFisher Scientific	GRB04	1:100
HLA-DR	Immu-357	Mouse, IgG1	FITC	Beckman Coulter	IM1638U	1:100

IgD	IA6-2	Mouse IgG <sub>2a</sub> , κ	PE	Biolegend	348203	1:100
NKp44	P44-8	Mouse IgG1, κ	PE	Biolegend	325107	1:100
NKp46	9E2	Mouse BALB/c IgG <sub>1</sub> , κ	PC5.5	BD Biosciences	557991	1:100
PD-1	EH12.2H7	Mouse IgG1, κ	FITC	Biolegend	329903	1:100
PDL1	MIH1	Mouse BALB/c IgG <sub>1</sub> , κ	PE	BD Pharmingen	557924	1:100
Perforin	dG9	Mouse IgG2b, κ	PerCP/Cyanine 5.5	Biolegend	308113	1:100
TNFR2	3G7A02	Rat IgG2a, κ	PE	Biolegend	358403	1:100
Va7.2	3C10	Mouse IgG1, κ	PE	Biolegend	351705	1:100
γδTCR	B1	Mouse IgG1, κ	FITC	Biolegend	331207	1:100

### 2.3.3 Reagent kits

**Table 2-9| Mass cytometry antibody conjugation reagents and equipment, Fluidigm catalogue number 201300**

Product supplied with kit	Volume (for 40 reactions)
R- Buffer	60mL
C-Buffer	55mL
L-Buffer	14mL
W-Buffer	80mL
Maxpar Polymer®	40 tubes
Lanthanide solution	200µL

**Table 2-10| Additional reagents required for antibody conjugation not provided with kit**

Reagents	Company	Catalogue number
Antibody stabiliser	Candor Bioscience GmbH	131125
Cell staining buffer	Fluidigm	201068
100µg of each antibody	As specified	As specified
0.5M TCEP: Pierce™ Bond-Breaker® TCEP solution	Thermo Fisher Scientific	77720

**Table 2-11 | Antibody conjugation with lanthanide metal equipment**

Equipment required for mass cytometry antibody conjugation	Company	Catalogue number (if applicable)
Thermo shaker PHMP-4	Grant-Bio	n/a
Microcentrifuge	Centrifuge 5415D Eppendorf	n/a
Amicon Ultra 0.5mL 50K	Millipore	UFC505096
Amicon Ultra 0.5mL 3K	Millipore	UFC500396

**Table 2-12 | Fluidigm, Cell-ID™ 20-Plex Pd Barcoding kit**

Product supplied with kit (catalogue number 201060)	Volume provided (for 60 samples)
3 sets of 20 barcodes in PCR tubes	10µL each
Maxpar® Fix I Buffer (5x)	15mL
Maxpar Cell Staining Buffer	500mL
Maxpar 10x Barcode Perm Buffer	50mL
Maxpar PBS	500mL

**Table 2-13 | Plasmablast differentiation reagents**

Reagent	Company	Catalogue number
Leucosep tubes	Greiner Bio One	227288
Total B cell isolation kit II, human	Miltenyi Biotec	130-091-151
MACS LS columns	Miltenyi Biotec	130-042-401
QuadroMACS Separator	Miltenyi Biotec	130-090-976
IL-21	Peptotech	200-21
Mega CD40L soluble human recombinant	Enzo Life Sciences	ALX-522-110-C010

## 2.4 Methods

### 2.4.1 Justification for antibodies selected for mass cytometry

The antibodies that were chosen for mass cytometry were selected for comprehensively immunophenotyping abundant and rarer populations of cells in blood samples. The decision was made to not stimulate the blood samples obtained from patients or healthy donors, but to keep the samples in their original biological context to observe the cellular populations. In addition, logistical

factors meant that a cytokine/chemokine panel could not be devised because mass cytometry requires samples to be processed fresh and not frozen. Whilst many antibodies have been conjugated successfully for mass cytometry, it is still limited in comparison to the possibilities available to flow cytometry. Although all antibodies were validated before applying to samples, some antibodies were already approved by the Fluidigm Immune Monitoring kit which provided confirmation for a novel technology.

## **2.4.2 Flow cytometry**

### **2.4.2.1 Preparation of peripheral blood mononuclear cells from whole blood**

Lithium heparinised coagulated whole blood was collected from healthy volunteers and processed for PBMCs by density gradient centrifugation. Using sterile, aseptic techniques in a category 1 tissue culture hood, blood was diluted at a 1:1 volume with PBS and slowly layered at a 45° angle onto the density gradient medium Lymphoprep™ and subsequently centrifuged for 20 minutes at room temperature with acceleration brake set at 5 and deceleration brake set at 0. After centrifugation, 3 layers can be observed with the top layer consisting of plasma, a middle layer consisting of the buffy coat containing the PBMCs and a bottom layer consisting of red blood cells and granulocytes. The buffy coat was removed by Pasteur pipette and placed into a new tube and washed with MACS buffer at 200g for 10 minutes to remove any contaminating platelets. Cells were then washed again in MACS buffer and resuspended in RPMI-1640 media supplemented with 10% FBS and 1% penicillin streptomycin. Cell count was obtained by using a haemocytometer and viability was assessed by trypan blue. Total cell number was calculated using the following equation below:

Total cell number = Cell count x 10<sup>4</sup> x volume of media x trypan blue dilution

## **2.5 CD138 plasmablast cell differentiation**

### **2.5.1 Purified B cell isolation and labelling**

This protocol was kindly provided by UCB Pharma, Slough. 20mL lithium heparinised coagulated whole blood was obtained from 1 healthy donor. Blood was diluted 1:1 using sterile PBS. Diluted blood was layered onto Leucosep tubes in a vertical manner aiming for the centre of the tube and centrifuged at 800g for 15 minutes with acceleration set at 3 and deceleration set at 3. The interface layer of PBMCs was collected in a fresh 50mL falcon tube and the volume made up to 50mL by PBS. Samples were centrifuged at 300g for 8 minutes with the acceleration and deceleration set to the maximum speed. Supernatant was discarded and the cell pellet was resuspended in 6mL PBS. 10µL of cells were taken and counted using a haemocytometer by diluting 1:1 with trypan blue. Cell suspension was centrifuged at 300g for 5 minutes. The supernatant was discarded and the pellet resuspended in RPMI-1640 media supplemented 10% FBS, 1% L-glutamine and 1% Penicillin Streptomycin and centrifuged at 300g for 5 minutes. The cell pellet was resuspended in 40µL of MACS buffer per 10<sup>7</sup> cells and 10µL biotin cocktail per 10<sup>7</sup> cells and incubated for 5 minutes in the refrigerator. After incubation, an additional 30µL MACS buffer per 10<sup>7</sup> cells and 20ul anti-biotin MicroBeads per 10<sup>7</sup> cells were added to the cell suspensions and incubated for 10 minutes at 4°C. LS columns were placed in a QuadroMACS separator and washed through with 3mL of MACS buffer. 15mL falcon tubes were set up beneath the columns ready for sample collection and 1mL of sample

was applied to each column and flow through collected. Columns were additionally washed with 3mL MACS buffer and the flow through collected.

### **2.5.1.1 Plasmablast differentiation assay protocol**

For differentiation of B cells into plasmablasts, 96 well u-bottom plates were plated with 200,000 B cells and stimulated with 10ng/mL CD40L and 50ng/mL IL-21 and incubated for 5 days at 37°C, 5% CO<sub>2</sub> in a cell culture incubator. When incubation was completed, supernatant was removed and cell pellets were resuspended in PBS and placed in a 15mL falcon tube. The differentiated plasmablasts were stained for with the markers CD19, CD27, CD38 and CD138 to confirm the presence of CD138.

## **2.6 Mass cytometry laboratory methods**

### **2.6.1 Lanthanide assignment for mass cytometry antibodies**

Lanthanides were assigned to antibodies by prior knowledge of the immunology of the expression of the marker. Metals that have multiple isotopes were assigned to antibodies that were not close or related in terms of expression. This is because the metal can become oxidised therefore causing the metal to contaminate the subsequent isotope channel.

### **2.6.2 Antibody conjugation**

Whilst most steps remain consistent with manufacturer's protocol, some modifications were made to optimise antibody yield as determined by UCB Pharma. The lanthanide and antibody steps were performed simultaneously.

#### **2.6.2.1 Antibody concentration**

Concentration of the purified carrier-free antibody concentration was determined by Nanodrop after blanking against the buffer they are suspended in. Filter tips were used throughout this protocol to prevent cross-contamination.

#### **2.6.2.2 Preloading the polymer with lanthanide**

The polymer reagent was spun down for 10 seconds in a microcentrifuge to ensure that the reagent was at the bottom of the tube. It was resuspended in 95µL of L-buffer (supplied with the Fluidigm Maxpar antibody labelling kit) and 5µL of lanthanide metal solution was added making the concentration 2.5mM in a total volume of 100µL. This solution was mixed thoroughly by pipetting and incubated at 37°C in a heat block for 35 minutes.

#### **2.6.2.3 Purifying the lanthanide-loaded polymer**

After incubation 200µL of L-buffer was added to a 3kDa filter and 100µL of the metal loaded polymer mixture was added to this. This was centrifuged for 25 minutes at room temperature, flow through discarded and repeated by adding 400µL of C-buffer to the filter and centrifuge at 12,000g for 30 minutes at room temperature.

#### **2.6.2.4 Buffer exchange and partial reduction of antibody**

Next, 100µg of stock antibody was added in up to 400µL of R-buffer to a 50kDa filter and centrifuged for 10 minutes at room temperature. The column flow through was discarded and 100µL of 4mM TCEP-R buffer was added to each antibody and mixed by pipetting. This was incubated at 37°C in a heat block for 30 minutes and paramount not to exceed this time as over reduction of the antibody would damage the epitope.

#### **2.6.2.5 Purifying the partially reduced antibody**

The 50kDa filter containing the partially reduced antibody was retrieved from the heat block and 300µL of C-buffer was to the 50kDa filter to wash the antibody. This was centrifuged at 12,000g for 10 minutes at room temperature, discard the flow through and repeat the wash by adding 400µL of C-buffer to the filter and centrifuge again for 12,000g for 10 minutes at room temperature.

#### **2.6.2.6 Conjugate the antibody with lanthanide-loaded polymer**

The 3kDa filter containing the purified lanthanide-loaded polymer from the centrifuge and the 50kDa filter containing the purified partially reduced antibody from the centrifuge were retrieved and the flow-through for both was discarded. The lanthanide-loaded polymer was resuspended in 100µL of C-buffer and the contents transferred to the corresponding partially reduced antibody in the 50kDa filter. This was mixed gently by pipetting and incubated at 37°C for 90 minutes.

#### **2.6.2.7 Washing the metal-conjugated antibody**

To wash the conjugated antibody, 200µL of W-buffer was added to the 100µL antibody conjugation mixture and centrifuged for 12,000g for 5 minutes and the flow-through discarded. The wash step was repeated for a further three times but with 400µL of W-buffer. After the final wash with W-buffer, 80µL of W-buffer to the 50kDa filter to dilute the 20µL conjugate to a volume of 100µL. The solution was mixed thoroughly by pipetting and the walls of the filter rinsed with the W-buffer. The absorbance of the conjugated antibody was measured at 280nm against a W-buffer blank using the Nanodrop. The volume of antibody stabilization buffer was calculated to elute the antibody at 0.1mg/mL. The filter was inverted in a new collection tube and half of the calculated volume of antibody stabiliser was added to the filter and centrifuged at 1000g for 2 minutes. This step was repeated with the next half of the calculated volume of antibody stabilization buffer and centrifuged at 1000g for 2 minutes. Antibodies were stored at 4°C ready for titration. The process yields 60% of the original concentration of the antibody.

#### **2.6.2.8 Antibody conjugation confirmation**

For each antibody, a separate 1.5mL Eppendorf tube was prepared. In each tube, one drop of OneComp compensation beads was added to the tube and 1µL of antibody added and mixed thoroughly with the beads. The mixture was incubated for 10 minutes at room temperature. Next, 1mL of Maxpar cell staining buffer was added for washing and the tube was centrifuged at 1000g for 3 minutes. The supernatant was discarded and the wash step repeated. The pellet was then resuspended in 100µL CAS and acquired on the mass cytometer.

#### **2.6.2.9 Antibody titration**

All antibodies used for mass cytometry were titrated to ascertain the optimum concentration of antibody required to differentiate between the negative and positive population. To assess this, 1mL

of whole peripheral whole blood from one healthy donor was fixed and frozen as described in section 2.1.3. Antibodies were titrated at the following concentrations: 5µg/mL, 2.5µg/mL, 1.25µg/mL, 0.625µg/mL, 0.3125µg/mL and 0.125µg/mL and collected on the mass cytometer.

### **2.6.3 Compensation for mass cytometry**

As described in the previous section 2.6.2.8, antibodies were added to OneComp compensation beads but this time simultaneously in one tube in their titrated volume. An acquisition of 10, 000 events on the mass cytometer were required to create a compensation matrix which can be applied post acquisition.

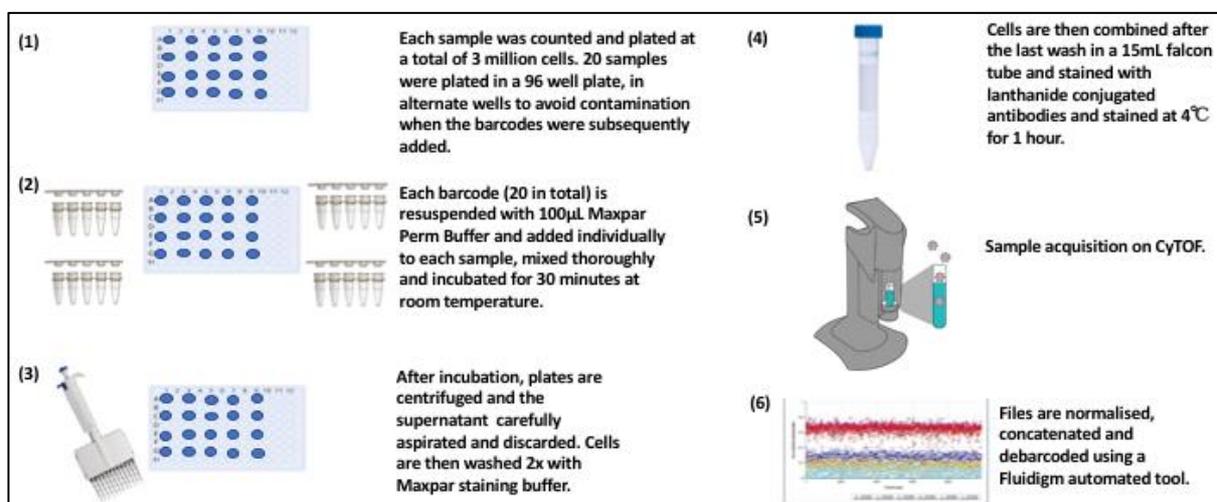
### **2.6.4 Barcoding patient samples for mass cytometry**

All patient and healthy donor samples were processed on the same day to reduce staining variability. Forty samples were retrieved at a time from the -80°C freezer and thawed in a 37°C bead bath. Once the samples were thawed, 1mL of fixed blood or synovial fluid was added to 9mL of 1x permeabilisation buffer in a corresponding labelled 15mL falcon tube and left upright for 15 minutes at room temperature. Samples were centrifuged at 800g for 5 minutes and the supernatant removed by aspiration. Samples were then resuspended in 5mL dPBS/1mM EDTA. Subsequently, 100µL aliquots of each sample was removed and placed into a 0.5mL Eppendorf tube and counted on the automated Sysmex cell counter which had been previously validated in house to count fixed cells. The 15mL falcon tubes containing cells in 5mL dPBS/1mM EDTA were then stored at 4°C overnight. The next day, cells were adjusted to 3 million cells and stored in dPBS/1mM EDTA overnight at 4°C.

The following day, samples were centrifuged at 800g for 5 minutes and flow through removed by aspiration. Samples were washed again in 5mL dPBS/1mM EDTA, centrifuged at 800g for 5 minutes and flow through removed by aspiration. Each sample was resuspended in 1mL barcode perm buffer and centrifuged at 800g for 5 minutes. Flow through was removed by aspiration and each sample was resuspended thoroughly in 100µL of barcode perm buffer and samples were added to every other well of a deep V bottom 96-well plate. Each barcode was resuspended in 100µL of 1x barcode perm buffer and transferred to the corresponding sample. Samples and barcodes were mixed thoroughly by multi-channel pipetting and left to incubate for 30 minutes at room temperature. After 15 minutes of incubation, samples were thoroughly mixed by pipetting. Plates containing the samples and barcodes were centrifuged at 800g for 5 minutes and using a multi-channel pipette, the supernatant was carefully aspirated to avoid disturbing the cell pellet. It was critical to remove the supernatant in its entirety to avoid unbound barcodes contaminating other samples. Cell pellets were washed by adding 300µL of Maxpar cell staining buffer solution to each well and mixed thoroughly by using a multi-channel pipette. This step was repeated following centrifugation. After the second wash step with Maxpar cell staining buffer, each sample was resuspended in 100µL of Maxpar cell staining buffer and combined in a 15mL falcon tube. Sample wells were rinsed an additional two more times to ensure maximum cell recovery. Cells were centrifuged at 800g for 5 minutes and the supernatant discarded by aspiration. Each batch of 40 samples were stored at 4°C in the refrigerator until all samples had been barcoded as antibody staining was to be done simultaneously across all barcoded samples.

#### **2.6.4.1 Preparation of internal batch control for barcoding**

Each barcoding batch contained an internal batch control which was obtained from the same donor. The internal batch control consisted of collecting 40mL of blood. The blood was divided with 10mL of whole blood processed as described in section 2.4.3 which provided the unstimulated control. The remaining 30mL of blood was collected and processed to obtain PBMCs using the method outlined in section 2.5.5.1. This served as a stimulated control for activation markers included in the mass cytometry panel. Isolated PBMCs were stimulated in a 15mL falcon tube with 50ng/mL PMA and 1ug/mL ionomycin for 4 hours at 37°C in a cell culture incubator. Cells were subsequently washed with PBS and centrifuged at 400g for 5 minutes. This was step was repeated and then cells were fixed with 1.6% PFA using the protocol described in 2.4.3. Subsequently, 10mL of 10% DMSO/90% PBS was added to the pellet and aliquoted across 10 tubes to store at -80°C. These internal batch controls were thawed at the same time as the patient samples were prepared as described in section 2.5.3. At the end both the unstimulated cells and stimulated PBMCs were counted and 1.5x10<sup>6</sup> cells of each were combined to create 1 sample containing 3x10<sup>6</sup> cells.



**Figure 2-1 | Schematic depicting the workflow of barcoding**

### 2.6.4.2 Antibody staining

After all samples had been barcoded and combined in their respective 15mL falcon tube, antibody staining commenced. Firstly, 100µL of Maxpar cell staining buffer was required for each sample. The dilution of each antibody was determined by titration. Antibodies were prepared in Maxpar cell staining buffer and left to incubate for 1 hour at +4°C in the refrigerator. Once the incubation was complete, 5mL PBS/1mM EDTA was added, and samples were centrifuged at 800g for 5 minutes. Supernatant was discarded, and the wash step was repeated. After another wash step, 300µL of chilled methanol was added to the cell pellet and the pellet left on ice for 15 minutes. After incubation, the pellet was centrifuged at 800g for 5 minutes and the methanol carefully aspirated and disposed appropriately following local safety procedures. The pellet was washed in 5mL PBS/1mM EDTA and centrifuged at 800g for 5 minutes. If the sample was to be run within 24 hours it was stored overnight in 4% PFA with 1/10,000 DNA Iridium Intercalator added at 4°C in the refrigerator. If the sample was to be run after 24 hours it was stored in 4% PFA in a -80°C freezer.

### 2.6.4.3 Preparation of samples on the day of mass cytometry

On the day of acquisition, the sample was thawed at room temperature and once defrosted, 1/5000 DNA Iridium Intercalator was added and the sample incubated for 1 hour at 4°C in a refrigerator. Following the incubation, 5mL of PBS/1mM EDTA was added to the sample and centrifuged at 800g for 5 minutes and the supernatant discarded. The pellet was washed in 5mL PBS/1mM EDTA and centrifuged at 800g for 5 minutes. The supernatant was discarded and the sample was resuspended in 5mL PBS/1mM EDTA. 1mL of sample was removed and placed into a fresh tube. This 1mL of sample was centrifuged at 800g for 5 minutes and the supernatant discarded. The pellet was resuspended in 1mL of CAS and the remaining 4mL of sample were stored at 4°C in a refrigerator. The 1mL of sample was subsequently acquired on the mass

cytometer and the remaining 4mL of sample was processed in the same way until all sample was used to completion.

### **2.6.5 Data acquisition on mass cytometer**

The mass cytometry machine was turned on daily and calibrated with EQ beads. Set up, calibration and pass checks took between 1.5-2 hours daily. Each day a new nebuliser was attached to the sample line to ensure that the nebuliser was free of any contaminating material to prevent blockages. Before acquisition began with samples, Maxpar water was run through the sample line for 5 minutes, followed by CAS buffer for 5 minutes.

Each barcoded batch was run where possible to completion over 2 consecutive days. It was important to run the samples over consecutive days to reduce any variability in antibody staining. Samples were run at a rate of 300 events per second to reduce the formation of doublets and collection took place over 8 hours.

## **2.7 Data analysis**

### **2.7.1 Normalization and concatenation of files**

After acquisition, .fcs files were normalized to correct for any signal discrepancies. Sample batches were normalized according to the internal reference determined by the EQ beads using the Fluidigm Normalizer built into the Mass cytometry2 v.6.0.626 software, Passport P13H2302\_ver2 (User Guide UG13-02\_150501). Once files were normalized, they were concatenated to create one file for each barcoded batch.

### **2.7.2 Debarcoding**

Each file was debarcoded using The Debarcoder software provided by Fluidigm v6.7.1014. The minimum Barcode Separation value was set to 0.12 as the optimum cut-off for distinguishing single cells from doublets and debris. The Mahalanobis Distance was not altered from default as it was not necessary to perform a secondary filter to remove aggregates.

### **2.7.3 R Studio and statistical analysis**

R studio was downloaded through the R studio open platform <https://rstudio.com/> The version of R Studio used in this study was 'R version 3.6.2 (2019-12-12)' and scripts were run on a UCB provided HP Elitebook Folio 1040 G3 laptop, Intel® Core™ i5-6300U CPU @ 2.40GHz, 2496MHz, 2 Core(s), 4 Logical Processor(s). The scripts used for automated gating and compensation matrix were

developed by Dr Emma Sutton at UCB. The normalisation script was adapted from the CytoNorm script<sup>175</sup> and can be found in full in Appendix III. The complete R Studio scripts used for the analysis in this study are in the supplementary section of this document.

To visualise the cell populations across all the samples the Diffcyt script was run in R studio. Samples were analysed according to the hypothesis as the computing power was limited to run all samples at simultaneously. It was ensured that the files were simultaneously prepared through the automated gating, compensation matrix and normalisation scripts to minimise signal staining discrepancies when analysing through Diffcyt.

The Diffcyt script was developed by Nowicka et al as an approach for analysing data from mass cytometry<sup>166,176</sup>. An automated data analysis approach using the Diffcyt script was selected due to the number of samples to analyse and applying a consistent approach for gating and clustering of cell populations which reduces bias which may increase if done by manual gating. Furthermore, small and rare cell populations can be overlooked if manual gating is initially applied. The Diffcyt script is not a fully automated script therefore encouraging the user to manually check the outcomes of each step to remove computer bias.

The Diffcyt script incorporates many of the features associated with single cell data analysis. Sample files were selected that best answered the hypothesis being tested and placed into a folder. To begin in Diffcyt, sample files were down sampled to the same number of events which would be determined by the file which had the lowest number of events. An excel file which contains information on the sample name, sample category (e.g. healthy, early RA, refractory RA) is uploaded into the script and a separate file containing the phenotyping markers and whether they are lineage and/or functional (denoted by a 1 if applicable or 0 if not). The first set of plots is a staining plot for each sample against all markers to observe if there are any obvious differences in marker expression. At this stage, if there are differences in expression it can be assessed whether these differences can be attributed to batch variation affects or if these differences may be biological. If it is thought that the differences are due to batch variation, staining profiles for samples need to be checked in Flowjo. In this study, staining profiles for all markers were checked after the application of the normalisation algorithm CytoNorm. It was observed that the staining expression for CD11b and CD16 fluctuated across batches which was rectified by CytoNorm.

Once staining expression across all markers and samples is confirmed, a multidimensional scaling (MDS) plot which visualises similarity in individuals of a dataset. Individuals are plotted on an abstract Cartesian axis scatterplot with marker expression being the measure of similarity. The MDS plot performs a global analysis to observe if there are any similarities in the samples and whether a trend can be observed. If nothing is observed by the MDS plot, this does not mean that there is no pattern or trend within the dataset, but it indicates that should there be any trends or differences, these may be more subtle.

The Diffcyt script incorporates the unsupervised clustering algorithm FlowSoM (Flow self-organising map) which can cluster millions of cells within minutes without the need for further sub setting. Clusters obtained from FlowSoM are labelled by the user and these labelled cell populations are visualised on tSNE1 and tSNE2 Cartesian axes. Once the unsupervised clustering step is completed, statistical testing can be performed to obtain a p value. The p value for significance can be set within the script and for this study a p value of  $\leq 0.05$  was significant.

The statistical testing is performed in two parts. Differential testing for cell population abundance and differential expression across the functional markers in each cell population uses the edge R

statistical package to fit models and calculate moderated tests at the cluster level <sup>176</sup>. These moderated tests improve power by sharing information on variability across clusters and return values in the form of raw p-values which is adjusted to account for false discovery rate <sup>176</sup>. For differential abundance, the data is represented as boxplots with percentage of cells denoted on the y axis and condition on the x axis. For differential expression the median expression of markers are measured on the y axis and the functional markers are measured on the x axis.

To confirm the findings from mass cytometry analysis, .fcs files were analysed in FlowJo version 10.7.1 to confirm populations using biaxial gating.

## Chapter 3 Antibody clone validation

### 3.1 Introduction

#### 3.1.1 Optimisation and selection of antibody clones for novel technologies

Mass cytometry is a novel technique when compared to flow cytometry with the clear advantage of including more markers in one panel. This provides the opportunity to either immunophenotype multiple cell populations or to comprehensively immunophenotype specific cell populations of interest. However, the flow cytometry field has further advanced and established protocols exist for immunophenotyping. Despite this, protocols can vary across different research groups due to the particular sample (e.g. patient tissue/cell line), methods used to handle and process the sample, the combination of markers to be analysed, controls included and the specific flow cytometer used for acquisition<sup>177</sup>. Although this list of variables is not exhaustive, it serves to highlight how inconsistencies can affect the antibody staining, which is measured by the cytometer before biological variation can even be considered. If the antibody staining fluctuates across samples, it can pose challenges when analysing data and drawing conclusions.

One of the drawbacks within the flow cytometry field is protocols can differ between groups even if similar biological samples and markers are used. This makes it difficult to compare findings let alone replicate findings although inclusion of validation data and controls used for flow cytometry experiments are part of the requirements to publish in many journals<sup>178</sup>. In flow cytometry, the expansion of panels and thus the inclusion of more markers has been gradual. It is only within the last decade that the increase of more markers (panels consisting of  $\geq 10$  markers) has become possible, creating the need for robust validation. As mass cytometry is regarded as a novel technique, there is encouragement within the field to streamline protocols for end users with concerted efforts to communicate new methods conveyed through publications, webinars and online resources. Mass cytometry is a field that is gaining more traction with the possibility of yielding new insights for many diseases, and has become increasingly accessible for many researchers leading to book chapters and textbooks compiling the latest approaches and protocols within mass cytometry<sup>179,180</sup>.

In this project it was important to validate all antibody clones to ensure that they bound specifically to the epitope of interest. All antibody markers were tested by flow cytometry, in order to evaluate the specificity of the target epitope being bound to the antibody. If antibody markers detected the specific epitope by an established technique such as flow cytometry, this would provide assurance that the antibody markers would be successful in mass cytometry. Testing antibody markers by flow cytometry is cost effective as testing clones initially by mass cytometry would be expensive and an inefficient use of resource.

Sample processing for mass cytometry is similar to flow cytometry although there are some caveats that are specific to mass cytometry. One of the key steps in preparing samples for mass cytometry is applying a fixative to samples. In flow cytometry, extracellular surface markers are often immunophenotyped without the application of a fixative. However, with logistical practicalities to take into consideration, (this study was geographically split between Leeds and Slough) it was not possible to acquire these samples on the mass cytometer on the same day as collection and processing. Therefore, whole blood and synovial fluid was immediately fixed with 1.6% paraformaldehyde, washed with phosphate buffered saline, and aliquoted into cryovials to be stored at  $-80^{\circ}\text{C}$  until it was possible to acquire on the mass cytometer. This fixation process allows for many

cell types to be preserved and minimises cell loss. This method will be referred to as the 'fix freeze method'. As the samples were instantly fixed on the day, a cell viability marker was not included.

The fix freeze method has been published by groups who have established mass cytometry protocols and was further validated in-house at UCB Pharma<sup>181</sup>. I validated the technique to ensure uniform processing across all samples for the purpose of this study. Before the fix freeze technique was applied to the samples, peripheral blood was obtained from one healthy donor to compare antibody staining on fresh samples compared with antibody staining on samples that have been processed by the fix freeze method. In flow cytometry, antibody markers could not be run simultaneously due to the limited channels so antibody markers were validated in smaller panels where the antibody markers included would identify the corresponding cell populations. For some of the antibody clones, several clones were validated to determine which clone had the best specificity. Antibody markers for flow cytometry from reputable companies are often extensively tested in both fixed and unfixed conditions but this should be confirmed by the end user for the specific sample type and processing methods used. The clones shown in the results section will be the final clones that were selected for this study and it is anticipated that findings from this study can be replicated. Furthermore, it is hoped that the protocols can be easily translated from a laboratory setting to a clinical setting diagnostic setting should future findings become clinically relevant.

The 37 antibody markers selected for this study was to broadly immunophenotype immune cells in peripheral blood and synovial fluid from patient with RA. The immunophenotyping panel in this study consisted of a combination of markers that have been well profiled in flow and mass cytometry literature alongside known but less characterised markers. It was only possible to include surface markers for the immunophenotyping panel as the samples were fixed with paraformaldehyde during the processing step. Due to logistical practicalities, it was decided that intracellular cytokines and chemokines would not be included in immunophenotyping panels. Primarily this was due to the cohort mainly comprising of peripheral blood. In order to capture intracellular cytokine and chemokine activity, this would require an additional stimulation step and access to the mass cytometer on the same day which was not possible. Intracellular staining by flow cytometry is a technique that would require a significant amount of validation especially if novel chemokines/cytokines are to be included in panels. Therefore staining for chemokines and cytokines by mass cytometry was beyond the scope of this study despite published reports including RA, including chemokines and cytokines in their mass cytometry panel design<sup>157,182,183</sup>.

Gating strategy and staining profiles within mass cytometry requires experience and familiarity with single cell staining plots. Familiarity with these plots improve over time but initial training and guidance for the researcher are required to achieve a level of confidence when interpreting cytometry data. Numerous papers reporting various markers have demonstrated gating strategies by flow cytometry. These are widely available and can serve as an initial reference point for comparing staining profiles and gating strategy although variables as outlined earlier must be taken into consideration. However, differences in gating and staining profiles exist in mass cytometry compared to flow cytometry. Firstly, mass cytometry does not possess the light scatter properties which are the forward and side scatter parameters. The light scatter properties are independent to antibody markers and can be used to profile lymphocytes, monocytes and granulocytes present in one sample. This is present in flow cytometry and although not exclusively used or relied upon, can be an initial approach for gating strategy which cannot be easily applied to mass cytometry. However, the light scatter properties become less relevant if samples have been fixed as was the case in this study. In mass cytometry, cells undergo a harsher process during acquisition whereby they are vaporized, atomised and ionized before the staining signal is recorded, making the binding specificity of antibody markers crucial in order to identify cell populations. In turn, the antibody staining informs the data analysis downstream which commonly incorporates unsupervised analysis

through bioinformatics and computer algorithms; thus the specificity of antibody staining is important when interpreting results and forming conclusions.

## 3.2 Results

### 3.2.1 Flow cytometry light scatter properties are redundant with fix freeze processed samples

Sample preparation for mass cytometry used the fix freeze method on whole blood and synovial fluid cells which comprises of fixing the samples on day of collection from the clinic with 1.6% paraformaldehyde and storing at -80°C until required for acquisition by thawing the sample for mass cytometry. This processing method changes the cellular profile visually when compared to a freshly processed sample which consists of lysing red blood cells, washing with phosphate buffer saline, staining with antibodies and acquiring events on the flow cytometer. Methods for this process are fully detailed in Chapter 2. The difference in cellular profiles affected by the processing method can be observed during the acquisition process on the flow cytometer and subsequently data analysis in the single cell analysis software Flowjo shown in Figure 3-1. This is particularly noticeable when viewing the cells against the forward and light scatter properties. As mentioned earlier, these light scatter properties are not present in mass cytometry but by flow cytometry demonstrate the affect of processing methods on the cellular profile. This is a widely known phenomenon in the cytometry field and Figure 3-1 illustrates the impact that processing methods have on cells.

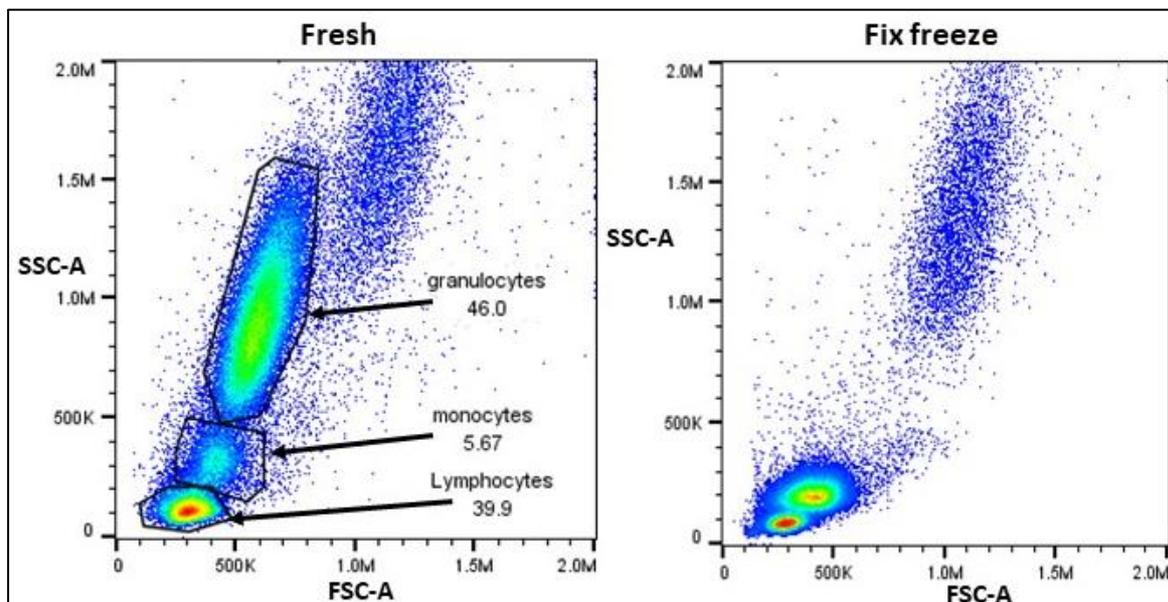


Figure 3-1 | Comparison of fresh and fixed whole blood from the same donor against light scatter parameters.

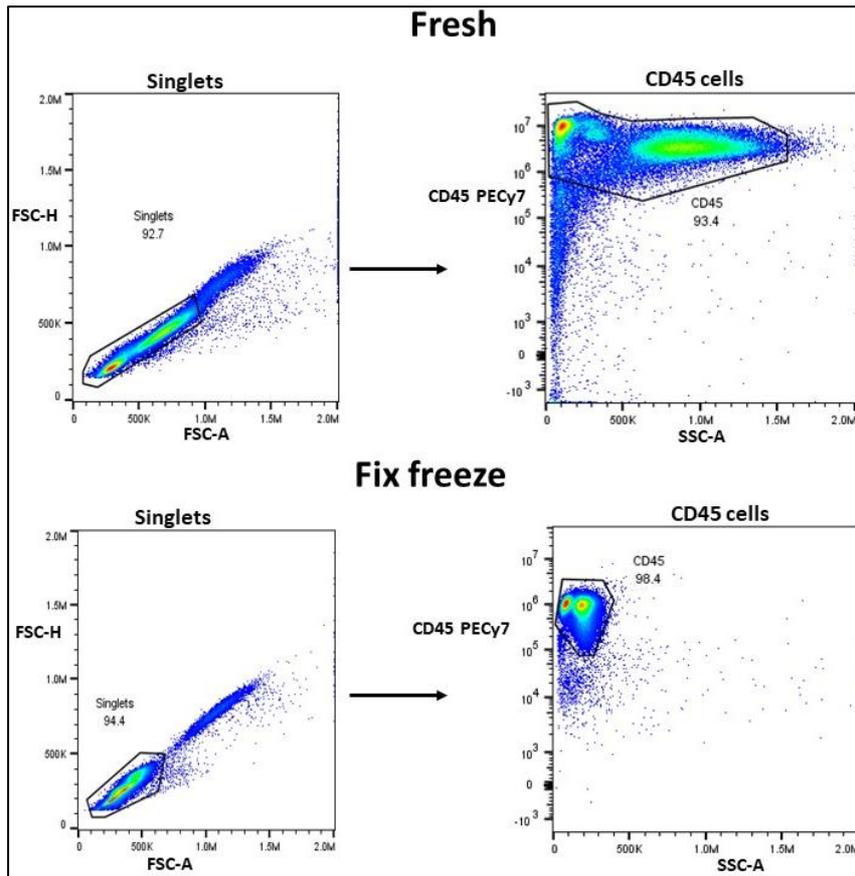
(figure legend on next page)

Whole blood from the same donor was processed by two different methods and acquired on the same day. Freshly processed blood prepared for flow cytometry was run on the Cytoflex S where 100,000 events were collected. Granulocytes, monocytes, and lymphocytes populations can be observed based on the forward and side scatter light properties. In contrast, when the blood sample was processed by 1.6% paraformaldehyde fixation, frozen at -80°C and subsequently thawed (fix freeze process), the light scatter parameters can no longer be relied upon to distinctly identify granulocytes, monocytes and lymphocytes with the cell populations resulting in a condensed profile. In both the profiles, a population with high SSC-A (which sit above the granulocyte population) can be observed. It is uncertain what this population of cells are although speculative analysis suggests that these could be eosinophils but this has not been confirmed by comprehensive immunophenotyping. The two profiles are for illustrative purposes only but underscores the importance of robust immune markers to identify cell populations and to not be reliant upon light parameters. Analysis done using FlowJo software version 10.7.1 **FSC-A** forward side scatter area, **SSC-A** side scatter area

In a freshly processed blood sample acquired on the flow cytometer, using the light scatter parameters, three broad immunological cell populations can be observed which is often a starting point for basic flow cytometry training and serves as an introduction to gating and general visualisation of flow cytometry data. Granulocytes typically consisting of neutrophils, have a larger scatter profile compared to the lymphocyte and monocyte populations. The granulocyte population is typically the largest immune cell population in whole blood shown as 46% followed by lymphocytes (39.9%) and monocytes (5.67%) in Figure 3-1. In health, these cell percentages can subtly vary for each donor, but the pattern is broadly similar, with the granulocyte population comprising the largest immune cell population in blood. However, this distinction of granulocytes, monocytes and lymphocytes are absent with the fix freeze processing method when looking at the cell events against the forward and side scatter parameters. The cellular profile condenses to the lower left corner of the axes demonstrating that the cells do not retain their shape after undergoing fixation and freezing processes. The loss of cell shape is not an issue and does not compromise the integrity or the interpretation of the data but strengthens the need for robust and reliable antibodies to identify cell populations.

### **3.2.2 CD45+ single cell gating strategy for flow cytometry**

To analyse the antibody clones, a two-step initial gating strategy was applied which consisted of removing the doublet cells and debris using the forward scatter area and height parameters and gating on the CD45<sup>+</sup> cell population which is present on all differentiated haematopoietic cells and absent on erythrocytes (Figure 3-2). The gating profile for freshly processed whole blood and fix freeze whole blood is shown in Figure 3-2 to demonstrate the cell profiles from the same healthy donor.



**Figure 3-2 | Gating strategy to obtain single cell CD45<sup>+</sup> cell population in peripheral whole blood**

Gating strategy shown for both fresh and fixed frozen processed whole blood from the same donor to obtain a single cell CD45<sup>+</sup> population. Analysis done using FlowJo software version 10.7.1 **FSC-A** forward side scatter area, **SSC-A** side scatter area, **FSC-H** forward side scatter height, **PECy7** Phycoerythrin-Cyanine 7

### 3.2.3 Validation of antibody clones by flow cytometry

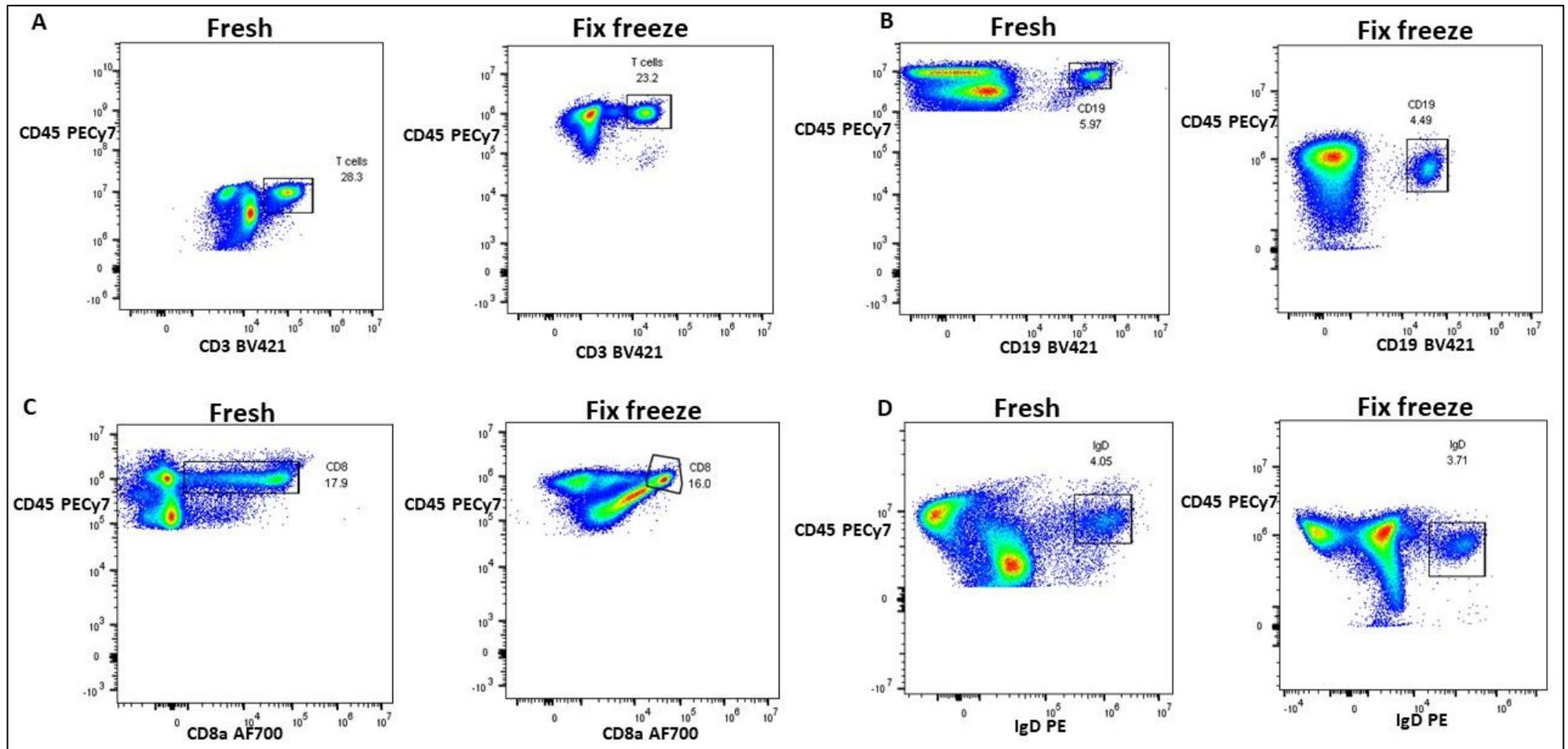
The same healthy donor was used for all antibody marker clone validation experiments to determine how the staining pattern appears in whole blood in health without any activation of cells. Due to the limitation of the flow cytometer as to the number of fluorophore channels that can be analysed in one panel it was decided to run each marker separately with CD45 only. This was the simplest way to assess the specificity of the antibody marker removing the complication of panel designs to accommodate fluorophores and compensation calculation. Each clone used in this study is detailed in Chapter 2. Fresh and fix freeze methods were applied to each antibody to compare percentage of cells staining positive for the marker and these have been adjacently positioned to compare the staining profiles. Freshly processed and fix freeze processed blood was run separately on the flow cytometer. Due to the effects of the fix freeze on the cells, the axes within the FlowJo analysis software were set to a biexponential scale. As anticipated, the process of fixing and freezing samples results in partial cell loss when compared to freshly processed samples. However, the fix freeze method was applied to all samples acquired on the mass cytometer thus removing this variation in cell percentage. Each marker was gated against CD45 to assess the staining profile. After applying the initial clean up gating (shown in Figure 3-2) the markers were assessed against CD45 to view their staining profiles (Figure 3-3). For freshly processed whole blood the percentage of cells for CD3, CD19, CD8a and IgD were 28.3%, 5.97%, 17.9% and 4.05% respectively whereas for fix freeze

processed whole blood the percentage of cells were 23.2%, 4.49%, 16% and 3.71% respectively. Therefore, fix freeze processing leads to a decreased percentage in positive cell staining for samples when compared to fresh processing. However, the markers clearly define a positive population and the cell percentages are comparable with fresh processed sample.

CD3 is an important marker in this study as it is used to identify the T cell population. The design of the antibody panel has many markers that would subset out T cells making it crucial for the CD3 marker to be specific. The CD3 marker separates the T cell population distinctly in a heterogeneous population of cells in both fresh and fix freeze sample processing conditions. Similarly, CD19 is a marker that has been selected to identify the B cell population. It was decided to use CD19 as a generic marker for B cells rather CD20 which can also be used either in conjunction with CD19 or as a substitute as some patients with RA are receiving the treatment Rituximab which targets CD20 on B cells. A large proportion of studies that immunophenotype T and B cells either isolate T or B cells from the blood prior to immunophenotyping to ensure the population of cells being assessed by cytometry techniques are already a 'pure' population with no other immune cell populations included. Alternatively, another approach is to isolate the lymphocyte and/or monocyte cell populations from blood and remove the granulocyte population as these are often seen as an expendable cell population due to their large number. However, the scope of mass cytometry allows for a more encompassing approach which means that the role of B and T cells amongst other cell populations gives us a more realistic 'snapshot' of the cellular interaction within health and disease.

CD8a is a marker can be used to identify various cell populations, but it is commonly used to immunophenotype cytotoxic T cells. In addition, CD8a is also expressed on a population of natural killer (NK) cells and MAIT cells.

IgD is used to identify various subpopulations of B cells. IgD is an immunoglobulin that is expressed on naïve B cells and the loss of surface IgD expression and expression of CD27 is associated with identifying classical switched memory B cells<sup>184</sup>. B cells expressing both IgD and CD27 are characterized as non-switched memory B cells<sup>184</sup>.



**Figure 3-3 | Validation of CD3, CD19, CD8a and IgD by flow cytometry.**

Gating for single cells and CD45+ cell population was performed for each antibody validation experiment in the analysis as shown in Figure 3-2. On the dot plot, each dot represents a single cell and density of a collection of single cells which equate to a population are denoted by blue, yellow, green and red colouration.

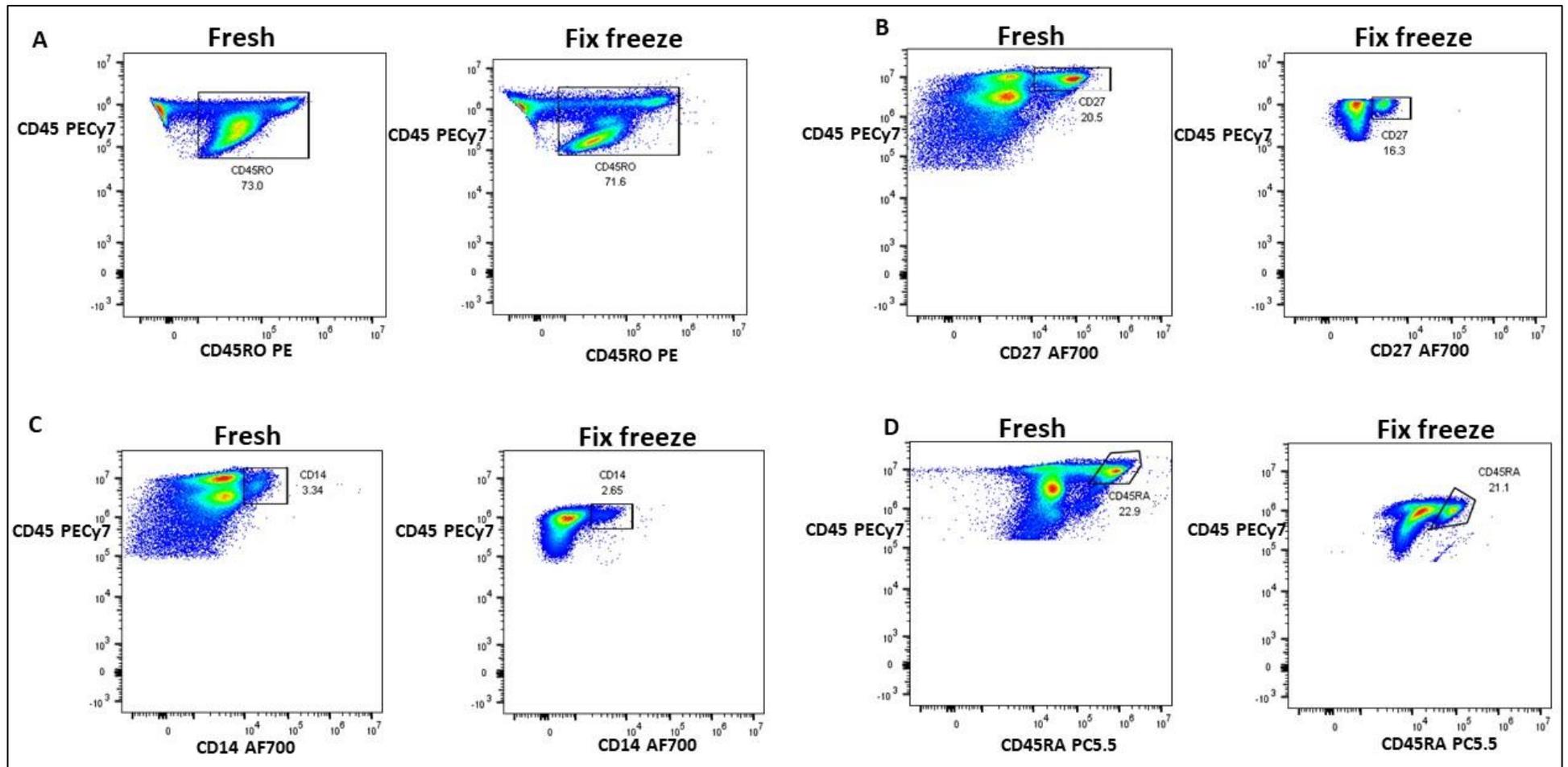
Antibody validation was carried out on peripheral whole blood and 100,000 events were acquired on Cytoflex S. Each marker was plotted against CD45 to ascertain the percentage of cells staining positive in fresh and fixed frozen blood processing. **A-D** Antibody validation for markers CD3, CD19, CD8a and IgD respectively. Analysis done using FlowJo software version 10.7.1  
**PECy7** Phycoerythrin-Cyanine 7, **AF700** Alexa Fluor 700, **BV421** Brilliant Violet 421, **PE** Phycoerythrin

Figure 3-4 shows the validation for CD45RO, CD27, CD14 and CD45RA and for freshly processed whole blood the percentage of cells were 73%, 20.5%, 3.34% and 22.9% respectively whereas for fix freeze processed whole blood the percentage of cells were 71.6%, 16.5%, 2.65% and 21.1% respectively.

CD45RO is a marker that is typically used to identify memory T and B cells. Figure 3-4A shows the granulocyte population also stains positive for CD45RO and this staining pattern is also seen in mass cytometry data (shown in Chapter 2). CD45RA is a marker used in immunophenotyping naïve B and T cells.

CD27 is expressed on predominantly on T, NK and B cells and is a member of the tumour necrosis factor receptor family <sup>185</sup>. CD27 plays a key role in T cell activation by providing a costimulatory signal and increases T cell proliferation and differentiation <sup>186</sup>. Whilst knowledge of CD27 has existed since 1994, it has garnered recent attention as it may serve as an important immune modulation target which offers new treatment options <sup>186</sup>.

CD14 is expressed on macrophages and monocytes and is a coreceptor for the detection of lipopolysaccharide <sup>187</sup>. Monocytes have received considerable attention over the last decade as evidence suggests that they have a role in chronic inflammatory diseases including expansion of the intermediate monocyte subset in RA <sup>96,188-190</sup>. CD14 is regarded as a broad marker of monocytes and along with CD16, three monocyte subsets have been characterised. These subsets are defined as classical monocytes (CD14<sup>+</sup> CD16<sup>-</sup>), intermediate monocytes (CD14<sup>+</sup>CD16<sup>+</sup>) and non-classical monocytes (CD14<sup>-</sup> CD16<sup>-</sup>) although data from mass cytometry immunophenotyping recommends using CCR2, CD36, HLA-DR and CD11c to characterise monocyte subsets accurately especially when discriminating classical and intermediate monocytes <sup>191</sup>. However, characterisation of monocytes is not fully understood in particular with regards to the precise function of the intermediate and non-classical monocyte subsets. Whilst this study was not exclusively investigating monocytes, CD86, CD11c and HLA-DR were included to further define the three monocyte populations.



**Figure 3-4 | Validation of CD45RO, CD27, CD14 and CD45RA by flow cytometry.**

Gating for single cells and CD45+ cell population was performed for each antibody validation experiment in the analysis as shown in Figure 3-2. On the dot plot, each dot represents a single cell and density of a collection of single cells which equate to a population are denoted by blue, yellow, green and red colouration.

Antibody validation was carried out on peripheral whole blood and 100,000 events were acquired on Cytoflex S. Each marker was plotted against CD45 to ascertain the percentage of cells staining positive in fresh and fixed frozen blood processing. A-D Antibody validation for markers CD45RO, CD27, CD14 and CD45RA respectively. Analysis done using FlowJo software version 10.7.1 **PECy7 Phycoerythrin-Cyanine 7, AF700 Alexa Fluor 700, PC5.5 phycoerythrin-cyanine 5.5 PE Phycoerythrin**

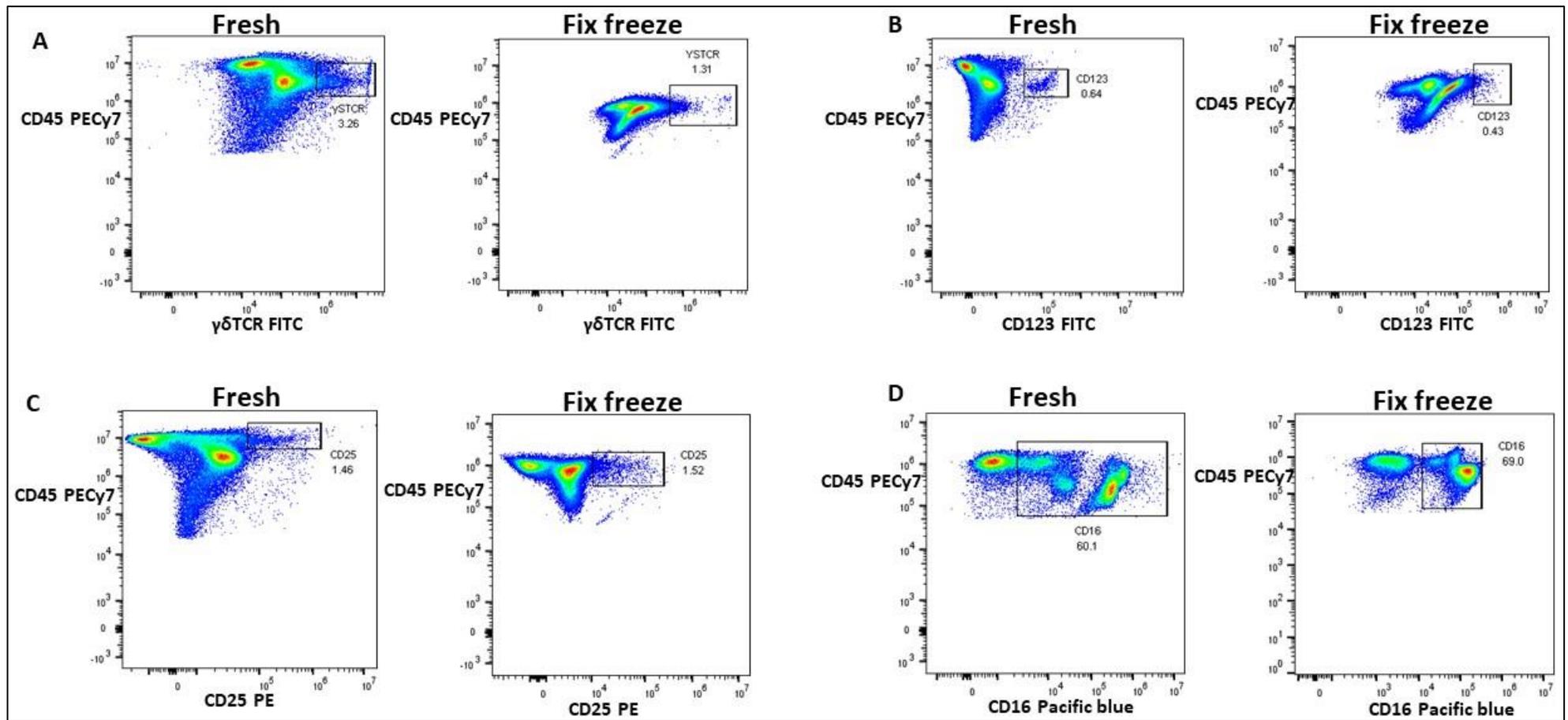
Figure 3-5 shows the validation for  $\gamma\delta$  T cell receptor (TCR), CD123, CD25 and CD16 and for freshly processed whole blood the percentage of cells were 3.26%, 0.64%, 1.46% and 60.1% respectively whereas for fix freeze processed whole blood the percentage of cells were 1.31%, 0.43%, 1.52% and 69% respectively.

$\gamma\delta$  T cells are an unconventional population of T cells and are a small population of T cells found in the peripheral blood. Conventional populations of T cells namely CD4 and CD8 T cell express  $\alpha\beta$  receptors however,  $\gamma\delta$  T cells are composed of  $\gamma\delta$  chains<sup>192</sup>. Whilst the selection process of  $\gamma\delta$  T cells is less well understood, they display unique characteristics.  $\gamma\delta$  T cells can target molecules without the need for MHC molecules and are capable of killing infected cells and microbes. In addition they are able to phagocytose large particulates including bacteria and are capable of acting as professional antigen presenting cells<sup>193</sup>.

CD123 is a marker for dendritic cells. Dendritic cells can be further subsetted into myeloid dendritic cells with additional markers CD11c and CD11b (both included in this study) and CD33. Plasmacytoid dendritic cells can be defined by CD123 and CD303 and CD304<sup>194</sup>. Furthermore, dendritic cells express HLA-DR and lack canonical markers of other cell subsets such as CD3 and CD19.

CD25 is a marker typically used to define T regulatory (Tregs) cells along with CD4 and transcription factor FOXP3<sup>195,196</sup>. Recently T regulatory cells have also been defined as a subpopulation identified as CD25<sup>high</sup> and FOXP3<sup>high</sup> and CD127<sup>lo</sup> predicting treatment response in aplastic anemia<sup>197</sup>.

CD16 also known as Fc $\gamma$ RIII is essential for antibody-dependent cellular cytotoxicity and is expressed on neutrophils, natural killer cells, macrophages (not found in peripheral blood) and monocytes<sup>198</sup>. As mentioned earlier, CD16 is used in collaboration with CD14 to subset monocyte populations. CD16 is a useful marker for subsetting neutrophils along with CD11b.



**Figure 3-5 | Validation of  $\gamma\delta$ TCR, CD123, CD25 and CD16 by flow cytometry.**

Gating for single cells and CD45+ cell population was performed for each antibody validation experiment in the analysis as shown in Figure 3-2. On the dot plot, each dot represents a single cell and density of a collection of single cells which equate to a population are denoted by blue, yellow, green and red colouration.

Antibody validation was carried out on peripheral whole blood and 100,000 events were acquired on Cytoflex S. Each marker was plotted against CD45 to ascertain the percentage of cells staining positive in fresh and fixed frozen blood processing. A-D Antibody validation for markers  $\gamma\delta$ TCR, CD123, CD25 and CD16 respectively. Analysis done using FlowJo software version 10.7.1

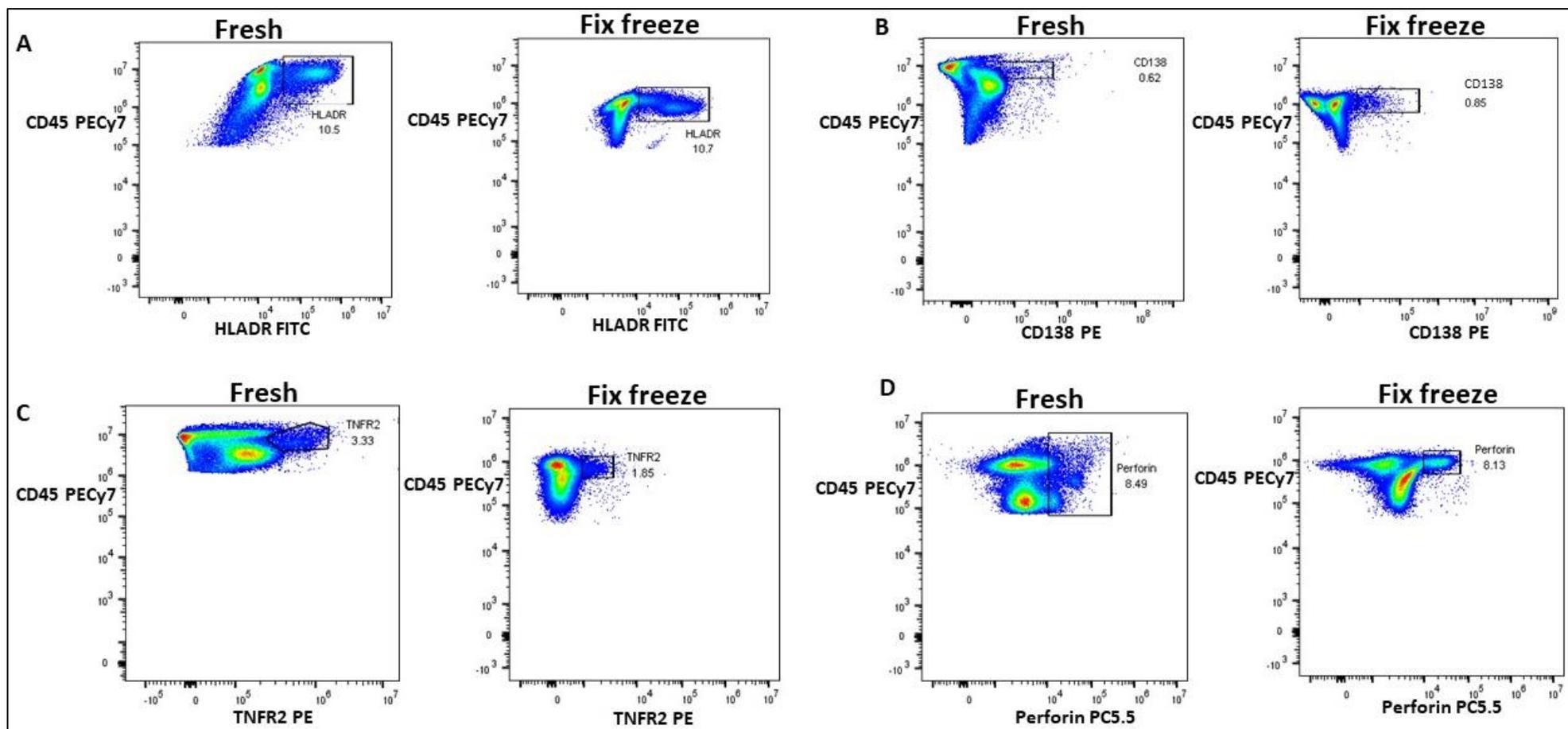
**PECy7 Phycoerythrin-Cyanine 7, FITC fluorescein isothiocyanate, PE Phycoerythrin**

Figure 3-6 shows the validation for HLA-DR, CD138, TNFR2 and Perforin and for freshly processed whole blood the percentage of cells were 10.5%, 0.62%, 3.33% and 8.49% respectively whereas for fix freeze processed whole blood the percentage of cells were 10.7%, 0.85%, 1.85% and 8.13% respectively.

HLA-DR (also known as MHCII) is mostly expressed on antigen presenting cells including B cells, dendritic cells, NK cells and monocytes and has also been observed to be expressed on activated effector T cells <sup>199,200</sup>. HLA-DR can also be used as a measure of activation on cells. As mentioned in Chapter 1, *HLA-DR* gene is central to the aetiopathogenesis of RA and has been incorporated in the immunophenotyping panel to investigate its expression at the protein level.

CD138 is expressed on plasma cells which are present as a small percentage in whole blood. Tumour necrosis factor 2 (TNFR2) is expressed on immune cells but is restricted to myeloid cells, Tregs, glial cells and endothelial cells <sup>201,202</sup>. Only the membrane bound TNF induces TNFR2 activation and not soluble TNF. TNFR2 is thought to be responsible for the survival and maturation of immune cells <sup>202</sup>.

Perforin is a pore forming cytolytic glycoprotein which is essential for immune cells to kill infected cells. Perforin is found in the granules of NK cells and cytotoxic CD8<sup>+</sup> T cell lymphocytes. Perforin forms a pore in target cell membranes which then polymerizes and forms a channel in Ca<sup>2+</sup> dependent manner <sup>203,204</sup>. Perforin works with granzyme and together they are the main pathway used by cytotoxic lymphocytes to eliminate virus-infected or transformed cells <sup>204</sup>.



**Figure 3-6 | Validation of HLA-DR, CD138, TNFR2 and Perforin by flow cytometry.**

Gating for single cells and CD45+ cell population was performed for each antibody validation experiment in the analysis as shown in Figure 3-2. On the dot plot, each dot represents a single cell and density of a collection of single cells which equate to a population are denoted by blue, yellow, green and red colouration.

Antibody validation was carried out on peripheral whole blood and 100,000 events were acquired on Cytoflex S. Each marker was plotted against CD45 to ascertain the percentage of cells staining positive in fresh and fixed frozen blood processing. A-D Antibody validation for markers HLA-DR, CD138, TNFR2 and Perforin respectively. Analysis done using FlowJo software version 10.7.1

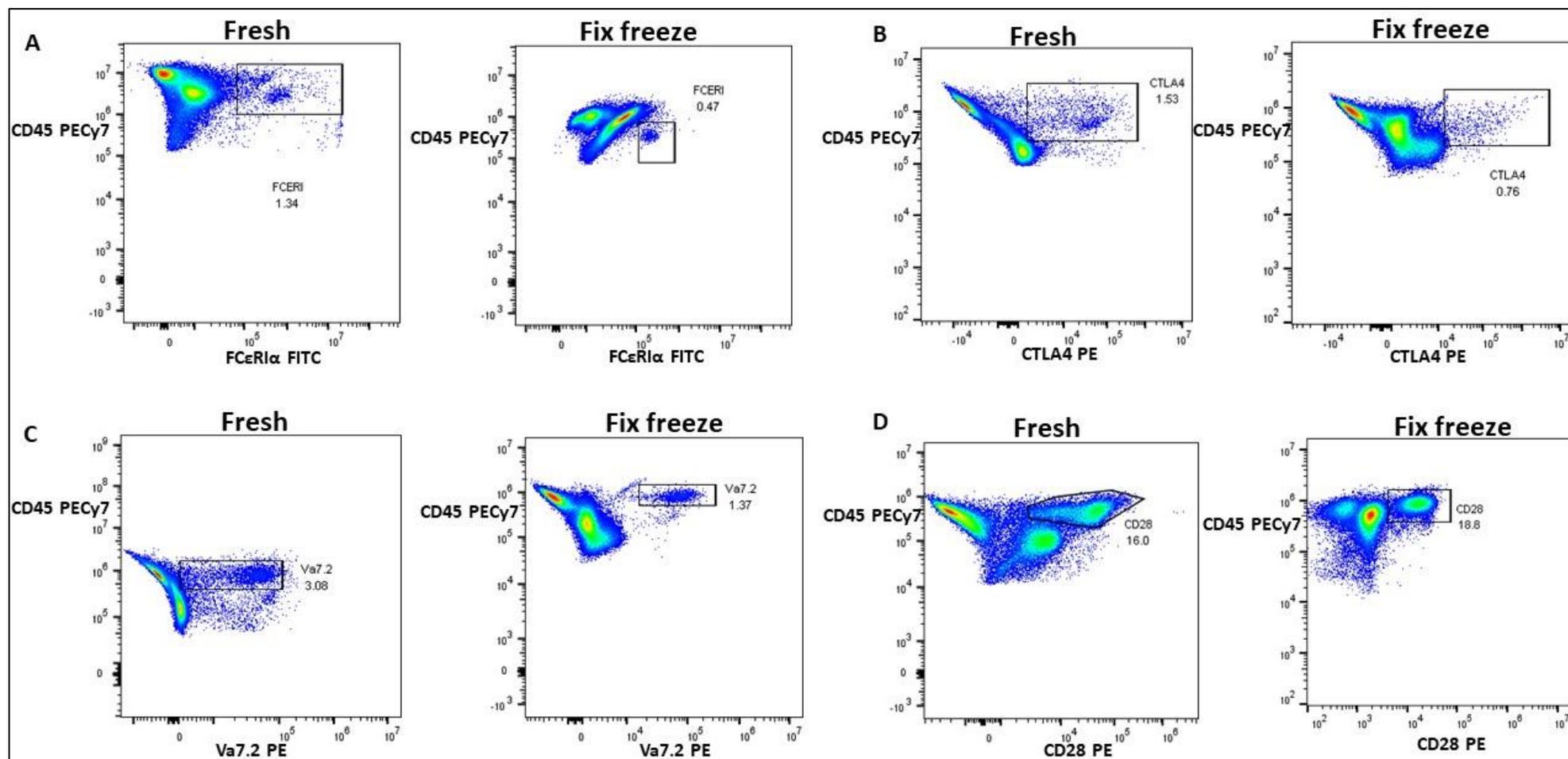
**PECy7** Phycoerythrin-Cyanine 7, **FITC** fluorescein isothiocyanate, **PC5.5** phycoerythrin-cyanine, **PE** Phycoerythrin

Figure 3-7 shows the validation for FCεRIα, CTLA4, Va7.2 and CD28 and for freshly processed whole blood the percentage of cells were 1.34%, 1.53%, 3.08% and 16% respectively whereas for fix freeze processed whole blood the percentage of cells were 0.47%, 0.76%, 1.37% and 18.8% respectively.

FCεRIα is an antibody binding site for the high affinity IgE receptor. IgE is associated with allergy and parasitic infections. FCεRIα is expressed on mast cells (present in tissue and absent in peripheral blood) and basophils.

T cell activation is a process consisting of two signals. The first signal consists of an antigen presenting cell presenting an epitope through MHC I or MHC II to the TCR on the T cell. To result in a fully activated T cell, CD28 binds to CD80/CD86 costimulatory receptors. However, CTLA4 (cytotoxic T lymphocyte-associated protein 4) can bind with higher affinity than CD28 and bind to CD80/CD86 and inhibit T cell activation<sup>205,206</sup>. T cell activation signalling through CD28 is better characterised than T cell inhibition through CTLA4. Studies have shown that CTLA4 and CD28 both bind with a higher affinity to CD80 than CD86 although very often CD80/CD86 are written together as the exact mechanism which determines how binding works has not been fully elucidated. CTLA4 is regarded as one of the immune checkpoints of the immune system that can be targeted by therapy and has shown to be useful in treating many cancers<sup>207</sup>. Abatacept which is an inhibitor of CTLA4 is commonly prescribed in the clinic to treat RA<sup>208</sup>.

Va7.2 is an α chain segment on the TCR which joins with the Jα33 segment which is known as an invariant TCR. These T cells are known as mucosal associated invariant T (MAIT) cells and possess both innate and adaptive immune cell characteristics. MAIT cells are restricted by the non-polymorphic MHC class I related (MR1) molecule which is expressed by many cell types<sup>209</sup>. MAIT cells are capable of combating bacterial disease by secreting IFN-γ and TNF observed in lung epithelial cells taken from healthy individuals which were infected in vitro with *Mycobacterium tuberculosis*<sup>210</sup>. MAIT cells are found in peripheral blood in healthy individuals as well as in tissues and are usually characterised with Va7.2 and CD161<sup>211</sup>.



**Figure 3-7 | Validation of FCεRIα, CTLA4, Va7.2 and CD28 by flow cytometry.**

Gating for single cells and CD45+ cell population was performed for each antibody validation experiment in the analysis as shown in Figure 3-2. On the dot plot, each dot represents a single cell and density of a collection of single cells which equate to a population are denoted by blue, yellow, green and red colouration.

Antibody validation was carried out on peripheral whole blood and 100,000 events were acquired on Cytoflex S. Each marker was plotted against CD45 to ascertain the percentage of cells staining positive in fresh and fixed frozen blood processing. A-D Antibody validation for markers FCεRIα, CTLA4, Va7.2 and CD28 respectively. Analysis done using FlowJo software version 10.7.1 **PECy7** Phycoerythrin-Cyanine 7, **FITC** fluorescein isothiocyanate, **PE** Phycoerythrin

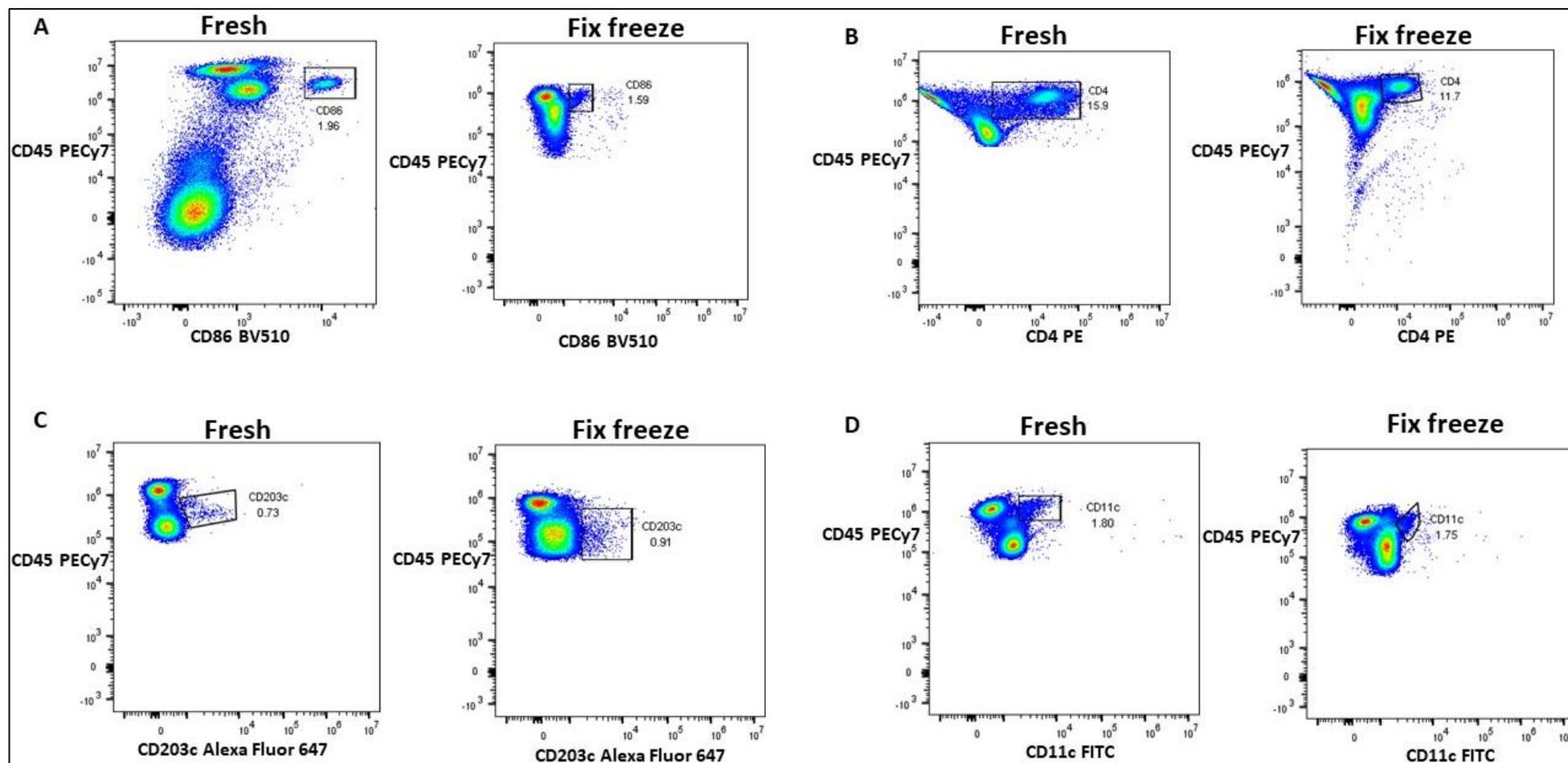
Figure 3-8 shows the validation for CD86, CD4, CD203c and CD11c and for freshly processed whole blood the percentage of cells were 1.96%, 15.9%, 0.94% and 1.80% respectively whereas for fix freeze processed whole blood the percentage of cells were 1.59%, 11.7%, 0.77% and 1.75% respectively.

CD86 is involved as a costimulatory molecule in B and T cell activation. It is present on antigen presenting cells including monocytes, B cells, dendritic cells and macrophages. Its role in T cell activation is one that is extensively studied along with CD80 and whilst their biophysical properties are well elucidated, their functional role in T cell activation remains less understood and an active research area <sup>212</sup>.

CD4<sup>+</sup> T cells are a well characterised and studied immune cell subset within the T cell population and interacts with MHC class II (HLA-DR). CD4<sup>+</sup> T cells are essential to the immune system in its effort to fight disease best demonstrated in individuals infected with human immunodeficiency virus have a depleted number of CD4 T cells and the ability for the host to fight infection is diminished <sup>213</sup>.

CD203c is a transmembrane protein expressed on basophils which are present in the granulocyte population and constitute 0.5-1% of circulating white blood cells <sup>214</sup>. Basophils are associated with allergy and asthma and are functionally important in secreting histamine which induces inflammation <sup>214</sup>.

CD11c is expressed on various cell populations including dendritic cells, monocytes, macrophages, neutrophils and B cells. CD11c is a member of leukointegrin family and is implicated in phagocytosis, cell migration and cytokine production by monocytes and macrophages.



**Figure 3-8 | Validation of CD86, CD4, CD203c and CD11c by flow cytometry.**

Gating for single cells and CD45+ cell population was performed for each antibody validation experiment in the analysis as shown in Figure 3-2. On the dot plot, each dot represents a single cell and density of a collection of single cells which equate to a population are denoted by blue, yellow, green and red colouration.

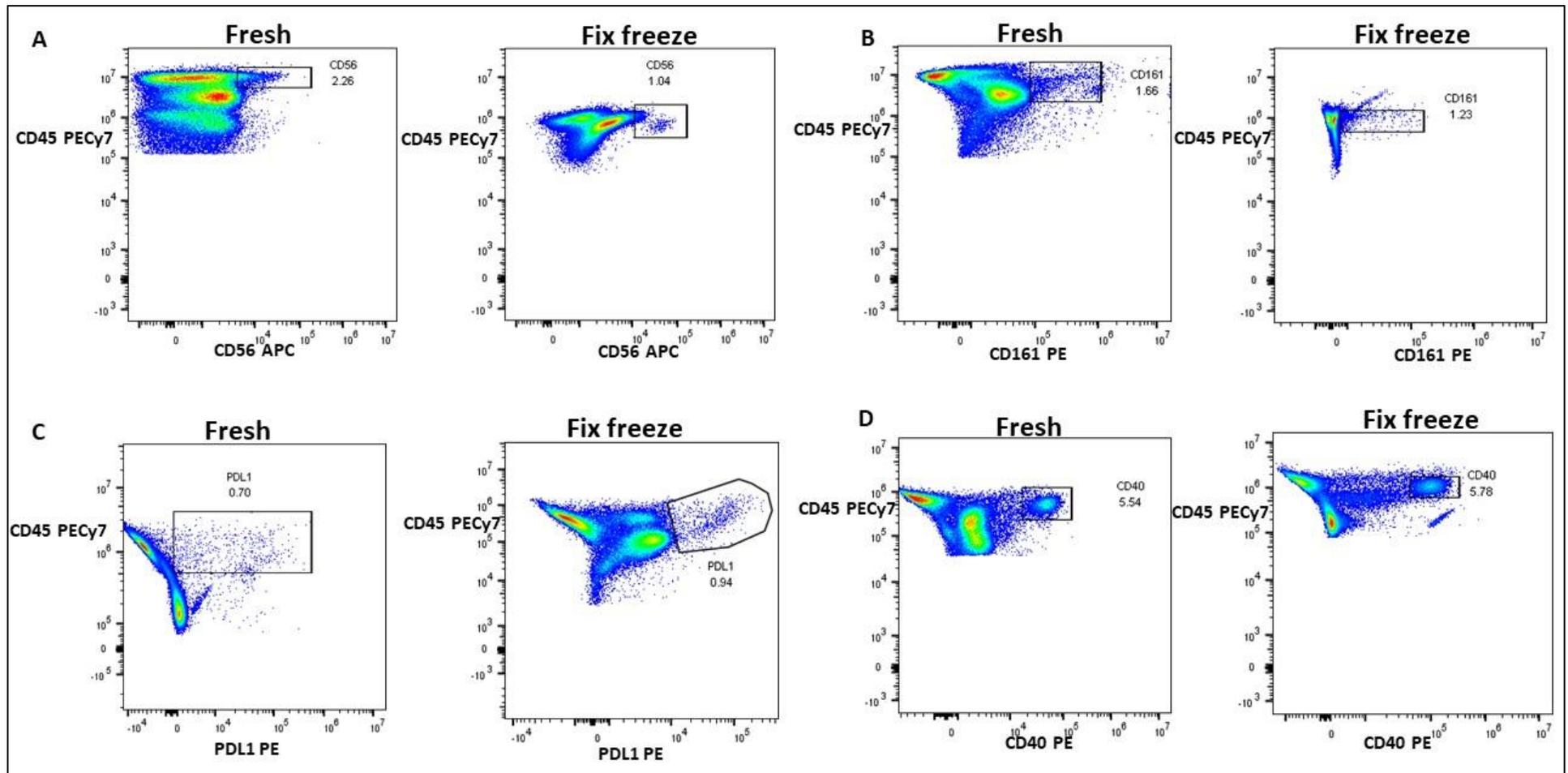
Antibody validation was carried out on peripheral whole blood and 100,000 events were acquired on Cytoflex S. Each marker was plotted against CD45 to ascertain the percentage of cells staining positive in fresh and fixed frozen blood processing. A-D Antibody validation for markers CD86, CD4, CD203c and CD11c respectively. For the validation of CD203c shown in C, the granulocyte population was removed due to high background staining. Analysis done using FlowJo software version 10.7.1 **PECy7** Phycoerythrin-Cyanine7, **FITC** fluorescein isothiocyanate, **BV510** Brilliant violet 510, **PE** Phycoerythrin

Figure 3-9 shows the validation for CD56, CD161, PDL1 and CD40 and for freshly processed whole blood the percentage of cells were 2.26%, 1.66%, 0.70% and 5.54% respectively whereas for fix freeze processed whole blood the percentage of cells were 1.04%, 1.23%, 0.94% and 5.78% respectively.

CD56 is mainly expressed on NK cells with the marker used to distinguish between CD56<sup>bright</sup> and CD56<sup>lo</sup> subsets. In health, NK cells are present at varying levels between 0.61-16.87%<sup>215</sup>. CD161 is a type II transmembrane glycoprotein and is a member of the C type lectin superfamily and is expressed on NK cells and natural killer T cells<sup>216</sup>. In addition CD161 along with va7.2 and CD8 can identify MAIT cells<sup>211</sup>.

PDL1 (programmed death ligand 1) expression has been identified on macrophages, some activated T and B cells, dendritic cells and some epithelial cells particularly under inflammatory conditions such as IFN- $\gamma$  activated monocytes and not abundantly expressed in healthy individuals<sup>217,218</sup>. The PD-1/PDL1 axis has been extensively studied in cancer and tumour immunology as an immune checkpoint blockade target<sup>219</sup>.

CD40 is a member of the tumour necrosis factor receptor superfamily and is well characterised in B cell signalling where CD40 is involved in promoting germinal centre formation, B cell activation and proliferation and is known to play a role in isotype switching<sup>220</sup>. CD40 binds to CD40 ligand which is situated on CD4+ T cells. CD40 however, is not restricted to B cells and is also present on monocytes, dendritic cells, NK cells and granulocytes<sup>220,221</sup>. CD40 plays a different role in these cells to that of B and T cells whereby it promotes cell survival and cytokine production<sup>220</sup>.



**Figure 3-9 | Validation of CD56, CD161, PDL1 and CD40 by flow cytometry.**

Gating for single cells and CD45+ cell population was performed for each antibody validation experiment in the analysis as shown in Figure 3-2. On the dot plot, each dot represents a single cell and density of a collection of single cells which equate to a population are denoted by blue, yellow, green and red colouration.

Antibody validation was carried out on peripheral whole blood and 100,000 events were acquired on Cytoflex S. Each marker was plotted against CD45 to ascertain the percentage of cells staining positive in fresh and fixed frozen blood processing. A-D Antibody validation for markers CD56, CD161, PDL1 and CD40 respectively. Analysis done using FlowJo software version 10.7.1

**PECy7** Phycoerythrin-Cyanine7, **APC** allophycocyanin, **PE** Phycoerythrin

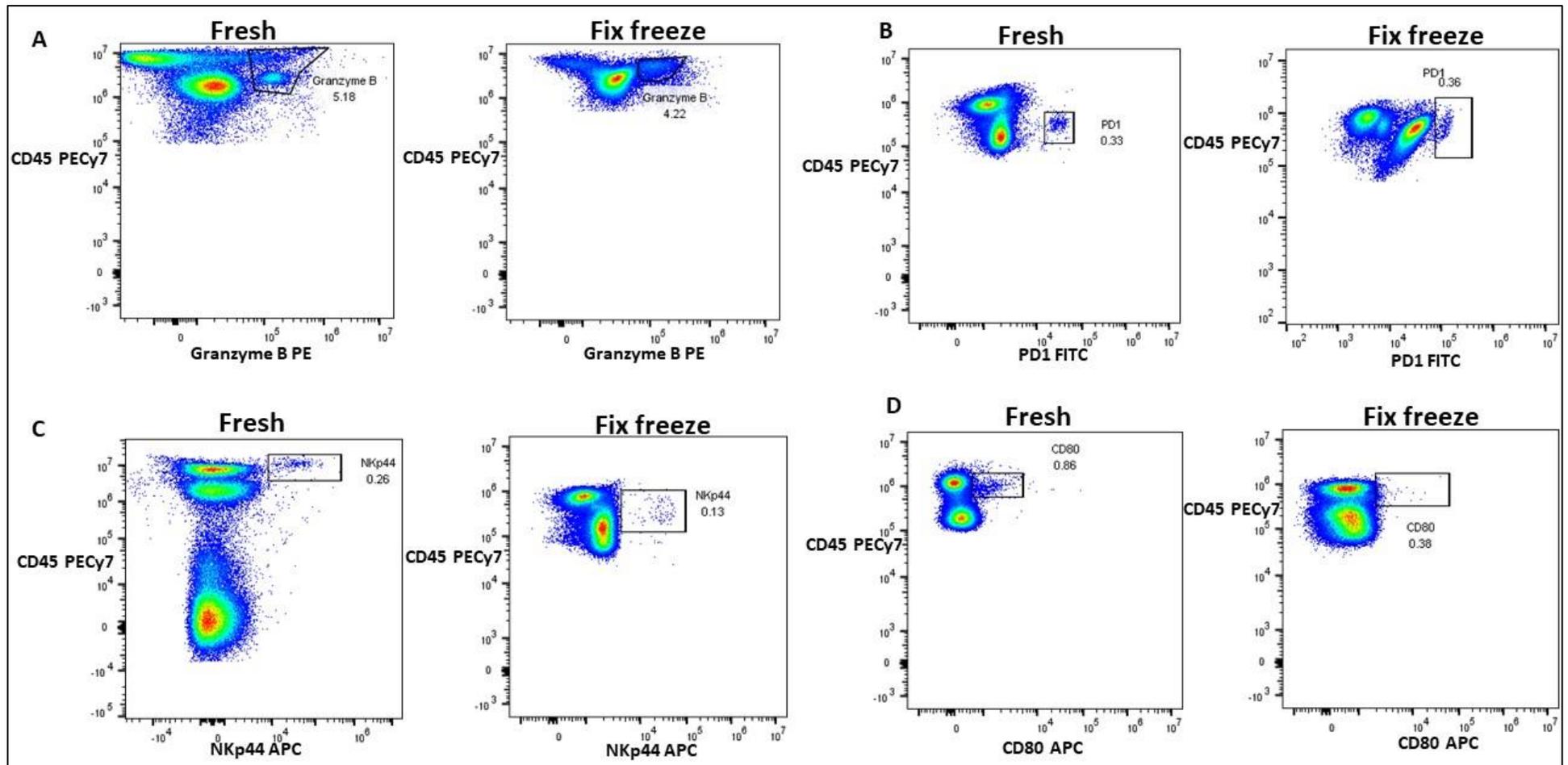
Figure 3-10 shows the validation for granzyme B, PD-1, NKp44 and CD80 and for freshly processed whole blood the percentage of cells were 5.18%, 0.33%, 0.26% and 0.86% respectively whereas for fix freeze processed whole blood the percentage of cells were 4.22%, 0.36%, 0.13% and 0.38% respectively.

Granzyme B is a serine protease that in combination with perforin mediates target cell apoptosis found in the granules of cytotoxic CD8+ T cells and NK cells <sup>222</sup>. Currently 4 other granzymes have been identified which are granzymes A, H, K and M. Granzyme B is perhaps the best studied out of the granzyme family and whilst an extensive amount of literature focuses on its cytotoxic role, evidence has emerged that suggest granzymes may have additional roles <sup>223</sup>. These roles include immune regulation such as cytokine processing and extracellular matrix degradation suggesting a role for granzymes in fibrosis <sup>223</sup>. Whilst granzyme B levels are present as part of a healthy immune system, they have been implicated in autoimmune disease including myositis and SLE and in RA, granzyme B has been shown to positively correlate with disease activity <sup>224</sup>. It has been hypothesised that granzyme B cleaves autoantigens which can lead to the creation of immunogenic protein fragments which are preferentially recognised by autoantibodies present in patients with autoimmune disease <sup>224</sup>.

Programme cell death protein 1 (PD-1) is an inhibitory receptor that is expressed by all T cells during activation <sup>225</sup>. PD-1 is a transmembrane molecule and is a member of the CD28 immunoglobulin family and in addition to being found on conventional T cells, PD-1 can be expressed by T regulatory cells, B cells, NK cells and myeloid cell populations <sup>226</sup>. PD-1 is considered to be an immune checkpoint molecule and PD-1 inhibitors exist and are particularly used as part of cancer treatment <sup>227</sup>. PD-1 often shows high and sustained expression especially during persistent antigen encounter often as a result of chronic infections and cancer and can limit the activation and function of potentially pathogenic self-reactive CD4+ and CD8+ T cells <sup>225</sup>. PD-1 is important for maintaining homeostasis of the immune system but when this balance becomes dysregulated chiefly in response to chronic pathogens and tumours, PD-1 can fail to limit protective immunity <sup>225</sup>. PD-1 has become synonymously linked with 'T cell exhaustion', a concept in which T cells progressively lose their effector functions over time in response to chronic antigenic stimulation and thus sustained chronic inflammation <sup>228</sup>. However, as PD-1 is expressed on a variety of cell populations, it is important to look at the context of this expression and therefore it is not always a marker of exhaustion.

NKp44 is expressed on IL-2 activated NK cells and enhances NK cell mediated cytotoxicity triggering receptor involved in non-MHC restricted cytotoxicity by activated NK cells <sup>229</sup>. NKp44 along with NKp30 and NKp46 are important receptors for NK cell cytotoxicity <sup>230</sup>. NKp30 and NKp46 are constitutively expressed on NK cells whereas NKp44 is expressed only upon activation.

CD80 is found on various surfaces of antigen presenting immune cells including B cells, monocytes and dendritic cells and is involved in the immunological synapse of T and B cells. It has been hypothesised that CD80 binds to CD28 with higher affinity than CD86 <sup>212</sup>. However, the mechanistic role of CD80 in T cell activation is yet to be fully elucidated.



**Figure 3-10| Validation of Granzyme B, PD-1, NKp44 and CD80 by flow cytometry.**

Gating for single cells and CD45+ cell population was performed for each antibody validation experiment in the analysis as shown in Figure 3-2. On the dot plot, each dot represents a single cell and density of a collection of single cells which equate to a population are denoted by blue, yellow, green and red colouration.

Antibody validation was carried out on peripheral whole blood and 100,000 events were acquired on Cytoflex S. Each marker was plotted against CD45 to ascertain the percentage of cells staining positive in fresh and fixed frozen blood processing. A-D Antibody validation for markers Granzyme B, PD-1, NKp44 and CD80 respectively. Analysis done using FlowJo software version 10.7.1

**PECy7** Phycoerythrin-Cyanine7, **FITC** fluorescein isothiocyanate, **PE** Phycoerythrin, **APC** allophycocyanin

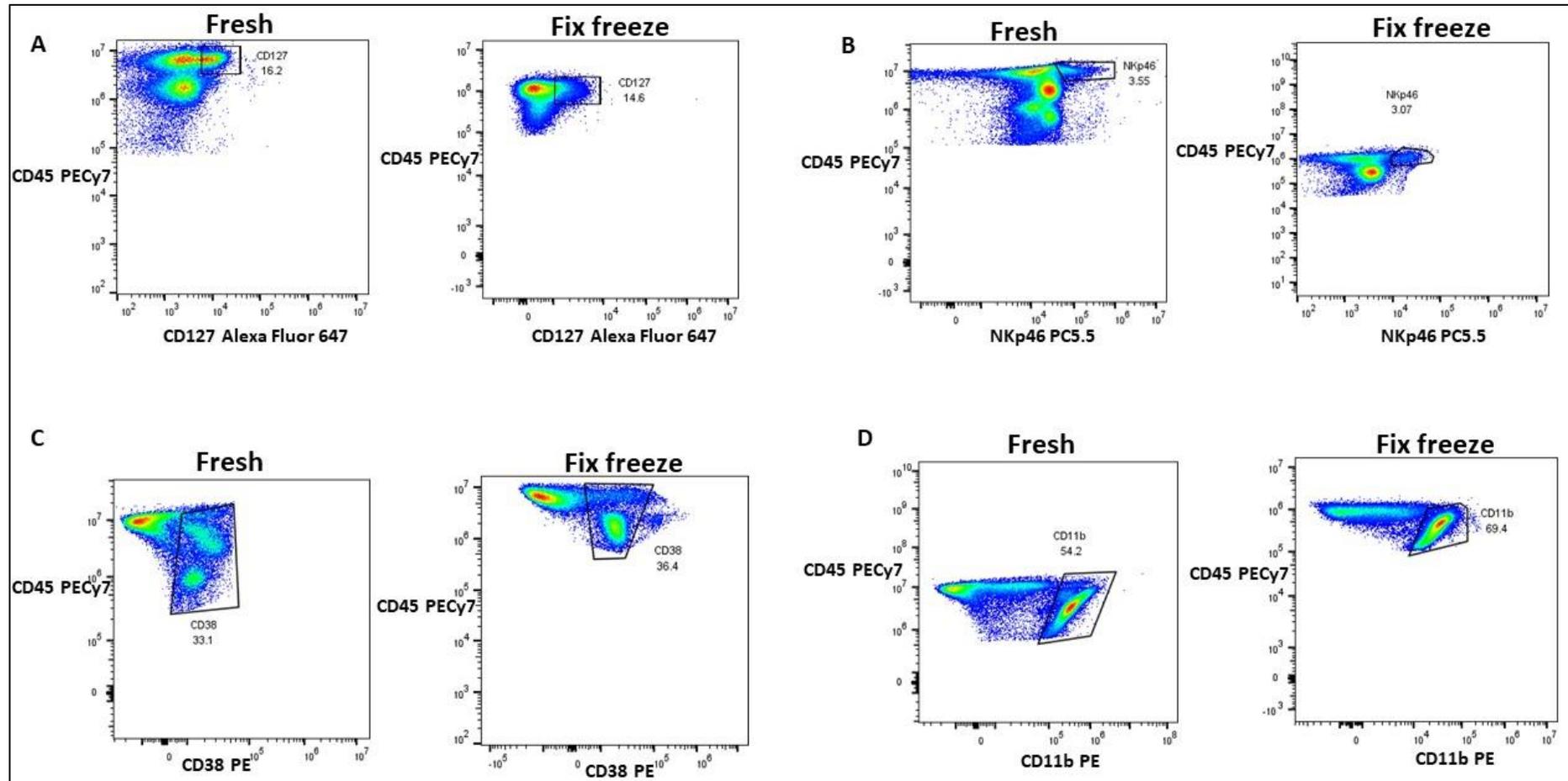
Figure 3-11 shows the validation for CD127, NKp46, CD38 and CD11b and for freshly processed whole blood the percentage of cells were 16.2%, 3.55%, 33.1% and 54.2% respectively whereas for fix freeze processed whole blood the percentage of cells were 14.6%, 3.07%, 36.4% and 69.4% respectively.

CD127 is also known as IL-7R $\alpha$  which is a receptor found on IL-7 cytokine. Upon engagement of this interaction, a signal transduction pathway is initiated through JAKs 1 and 3. IL-7R $\alpha$  signals are important for VDJ recombination and thus present on T and B cell lineages<sup>231</sup>. CD127 is used in immunophenotyping for T regulatory cells with expression being low on this population of T cells.

NKp46 is expressed by all CD56<sup>dim</sup> CD16<sup>+</sup> and CD56<sup>bright</sup> CD16<sup>-</sup> human NK cells irrespective of their activation status<sup>232</sup>. Blocking the expression of NKp46 by specific monoclonal antibodies can result in decreased NK cell cytotoxicity<sup>232</sup>.

CD38 is a multi-functional transmembrane, ecto-enzyme protein which is ubiquitously expressed on many immune cell types including bone marrow progenitors, NK cells, neutrophils, monocytes and activated B and T cells<sup>233</sup>. CD38 is functionally important in the synthesis of cyclic ADP-ribose which is a potent regulator of cytoplasmic Ca<sup>2+</sup> mobilization<sup>234</sup>.

CD11b is a member of the  $\alpha$ 2-integrin family and with CD18 forms a heterodimer to form the complement receptor CR3<sup>235</sup>. Conventionally CD11b is expressed on myeloid lineage cells including neutrophils, monocytes, macrophages, NK cells and granulocytes<sup>236</sup>. CD11b is thought to play a role in the migration of leukocytes from peripheral blood to sites of inflammation and has been further implicated in cell-mediated cytotoxicity, chemotaxis and phagocytosis<sup>236</sup>.



(figure legend on next page)

### **Figure 3-11 | Validation of CD127, NKp46, CD38 and CD11b by flow cytometry.**

Gating for single cells and CD45+ cell population was performed for each antibody validation experiment in the analysis as shown in Figure 3-2. On the dot plot, each dot represents a single cell and density of a collection of single cells which equate to a population are denoted by blue, yellow, green and red colouration.

Antibody validation was carried out on peripheral whole blood and 100,000 events were acquired on Cytoflex S. Each marker was plotted against CD45 to ascertain the percentage of cells staining positive in fresh and fixed frozen blood processing. A-D Antibody validation for markers CD127, NKp46, CD38 and CD11b respectively. Analysis done using FlowJo software version 10.7.1

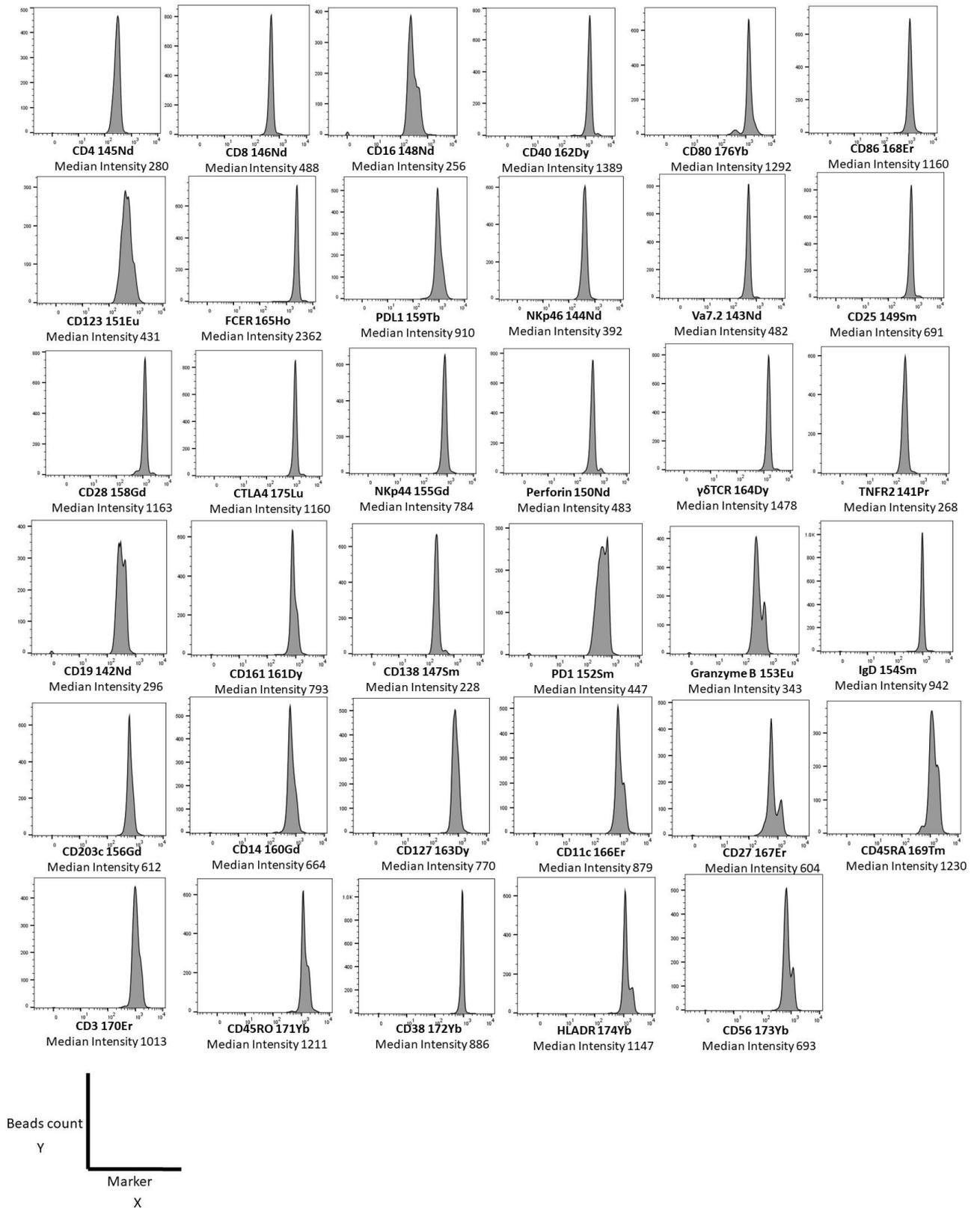
**PECy7** Phycoerythrin-Cyanine 7, **PC5.5** phycoerythrin-cyanine, **PE** Phycoerythrin

### **3.2.4 Validation of lanthanide conjugated antibodies for mass cytometry**

Mass cytometry is a novel technique which means that validating the methods was crucial to ensure antibody staining was optimum. Prior to running any cohort samples through the mass cytometer, each antibody that was conjugated with a metal lanthanide was checked to ensure that the signal emitted from the lanthanide metal and the cell population identified by the antibody marker was detected correctly. Ultracomp™ compensation beads which are commonly used in flow cytometry were used for lanthanide metal detection to ensure that a signal was recorded in the correct metal channel. It is important to emphasise that the Ultracomp™ beads were used as the antibody bound better to these beads to elicit a signal compared to other compensation beads. It is important to assess that the antibody has bound successfully to the metal lanthanide as the chemical reaction that takes place to allow this conjugation process to occur is not fully understood or explicitly explained by Fluidigm most likely due to intellectual property.

However, it is possible that the conjugation process may not work successfully leaving the metal lanthanide and antibody unbound therefore making this validation step important. Figure 3-12 shows the validation checks for each antibody that was conjugated and used in the immunophenotyping panel. To assess conjugation, histograms were plotted in Flowjo analysis software and median intensity was calculated for each metal lanthanide conjugated to an antibody on the mass cytometer. For each metal lanthanide conjugated antibody, a single peak should be detected. CD45 and CD11b were purchased from Fluidigm conjugated to Yttrium and Bismuth respectively. These two lanthanides are often difficult to conjugate. Fluidigm robustly validate their antibodies and are a reliable source so it was not necessary to check these antibodies although these were subsequently checked during antibody titration.

Using the compensation beads only allows the assessment of whether the metal lanthanide has been conjugated successfully to the antibody. Whether the conjugated antibody specifically binds to the epitope of interest and thus identifies the correct cell population would need to be further using cells.



**Figure 3-12| Validation of each lanthanide conjugated antibody for mass cytometry using Universal Comp beads.**

*(please see figure legend on next page)*

A histogram for each metal labelled antibody with the median intensity for the lanthanide metal and bead count shown for each antibody that was conjugated in-house at UCB Pharma, Slough. Beads were used instead of cells as an initial check to ensure the lanthanide had bound to the antibody. If this initial conjugation had not worked, the conjugation would have to be repeated. The X axis label depicts the antigen marker and the Y axis shows the bead count number. For each marker, the median intensity is shown which quantifies the signal strength of the conjugated antibody recorded in its appropriate channel. A single peak represents a signal has been detected in the channel. Where there is a bimodal peak distribution this indicates that the signal, though present in the appropriate channel, may indicate that part of the lanthanide is chemically dissociating from the marker. However, this is not a concern as the signal can be observed in the appropriate channel and no markers have dissociated completely from their lanthanides. Whilst this analysis confirms that the antibodies have been successfully conjugated, it does not confirm epitope specificity, which will be demonstrated in the next figure.

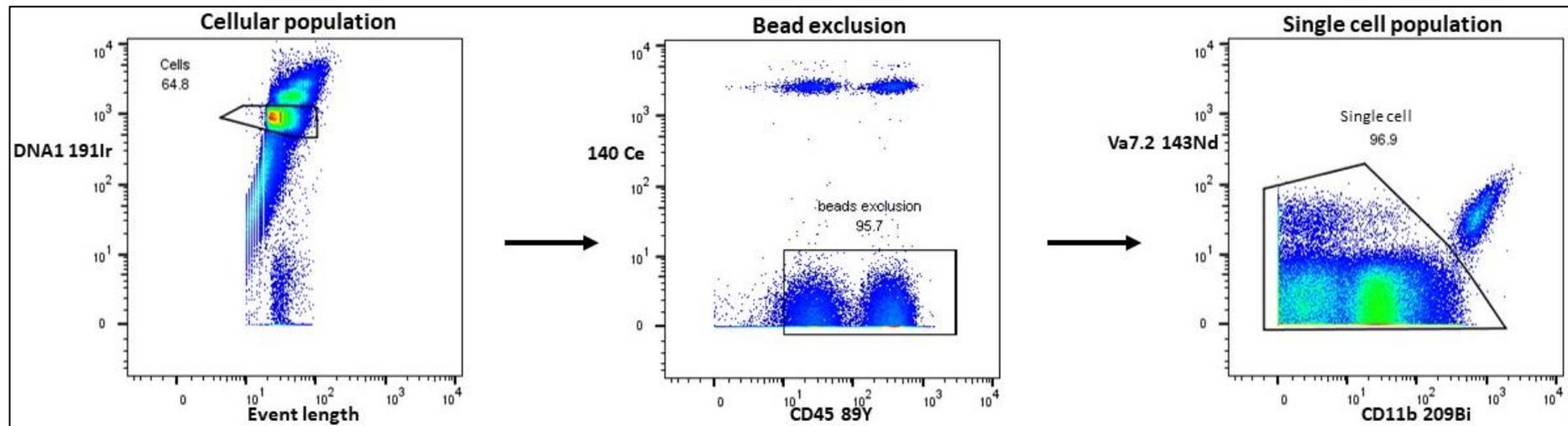
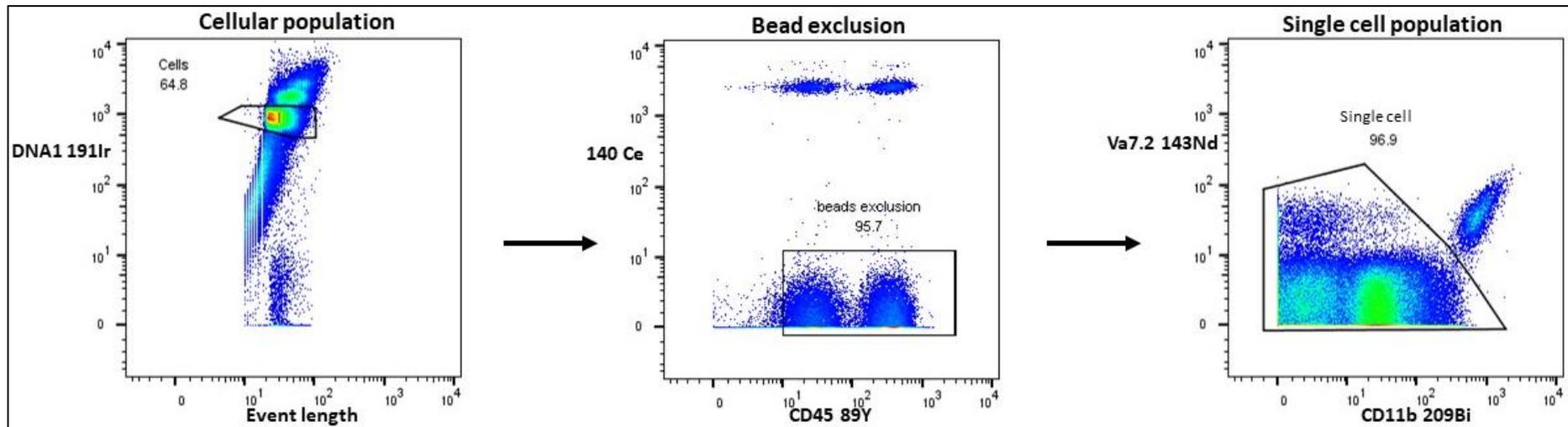


Figure 3-13 depicts the gating strategy used to obtain a single cell population where the DNA intercalator Iridium and event length are the initial gates drawn to broadly identify the cell population. To reduce the number of doublets, events were collected between 300-500/s but not exceeding 500 events. Increasing the number of events not only increases the formation of doublets but also potentially leads to blocking of the nebulizer resulting in longer run times. The bead population are removed using the metal channel 140 Ce as this channel was not used for any of the markers and was the best at distinguishing the normalisation beads and gating on the CD45+ cells. Lastly, to remove the contaminating platelets, a gate using the 143Nd and 209Bi lanthanide channels was drawn to eliminate this population.

Once all the antibodies for the panel had been conjugated and validated, antibodies for the mass cytometry panel were titrated at 5 $\mu$ g, 2.5 $\mu$ g, 1.25 $\mu$ g, 0.625 $\mu$ g, 0.3125 $\mu$ g and 0.16 $\mu$ g to determine the optimum volume which would distinguish the negative and positive cell populations. Figure 3-14, Figure 3-15, Figure 3-16, Figure 3-17 and Figure 3-18 show 1.25 $\mu$ g of each antibody run in one panel to determine the staining of the particular marker against CD45. Figure 3-19 shows the gating strategy and staining pattern of CD138 which is a marker for plasma cells present in low frequencies in healthy blood. Therefore B cells were differentiated into plasma cells (method highlighted in chapter 2) and staining of CD138 was confirmed by gating for plasma cells which were CD19 negative and CD138 positive.

### 3.2.5 Titration and validation of lanthanide conjugated antibodies for mass cytometry



**Figure 3-13 | Manual gating strategy for mass cytometry to obtain a single cell population**

To check the antibody staining by mass cytometry, a simple gating strategy was performed. To identify the cell population, a gate was drawn on the events which were positive for the DNA intercalator Iridium (Ir) against event length. Events that stain highly for the DNA intercalator are doublets and were not included in the gate. Beads were excluded from the cell events by gating on the metal channel 140 Ce and gating on the CD45+ events 140Ce negative. Beads and cells are clearly separated in this biaxial gating dot plot. The last gate was to remove the platelet population from the event count which was best achieved against the metal channel 143Nd and 209Bi. A single cell population can now be obtained and further gated to identify antibody staining.

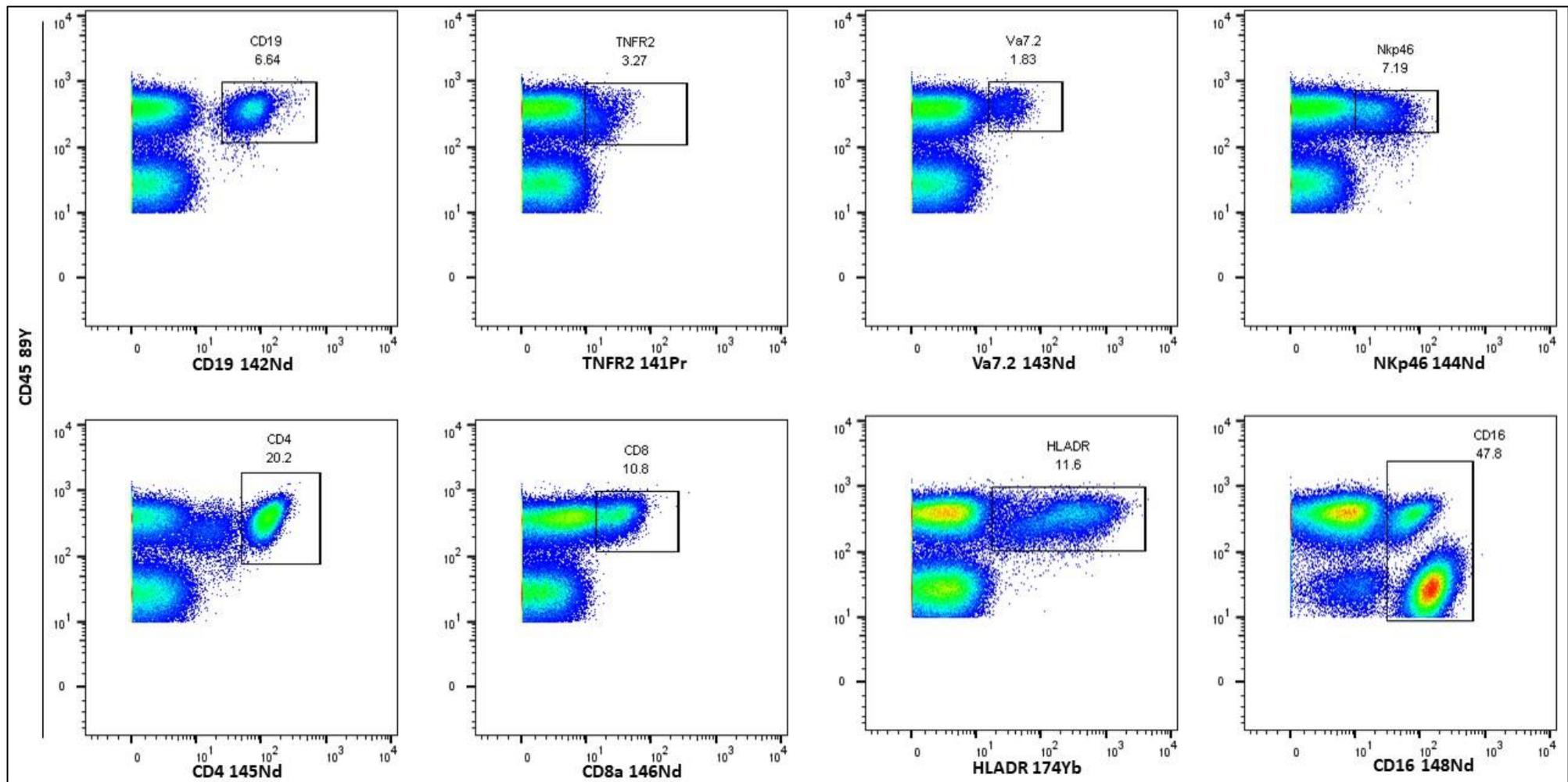
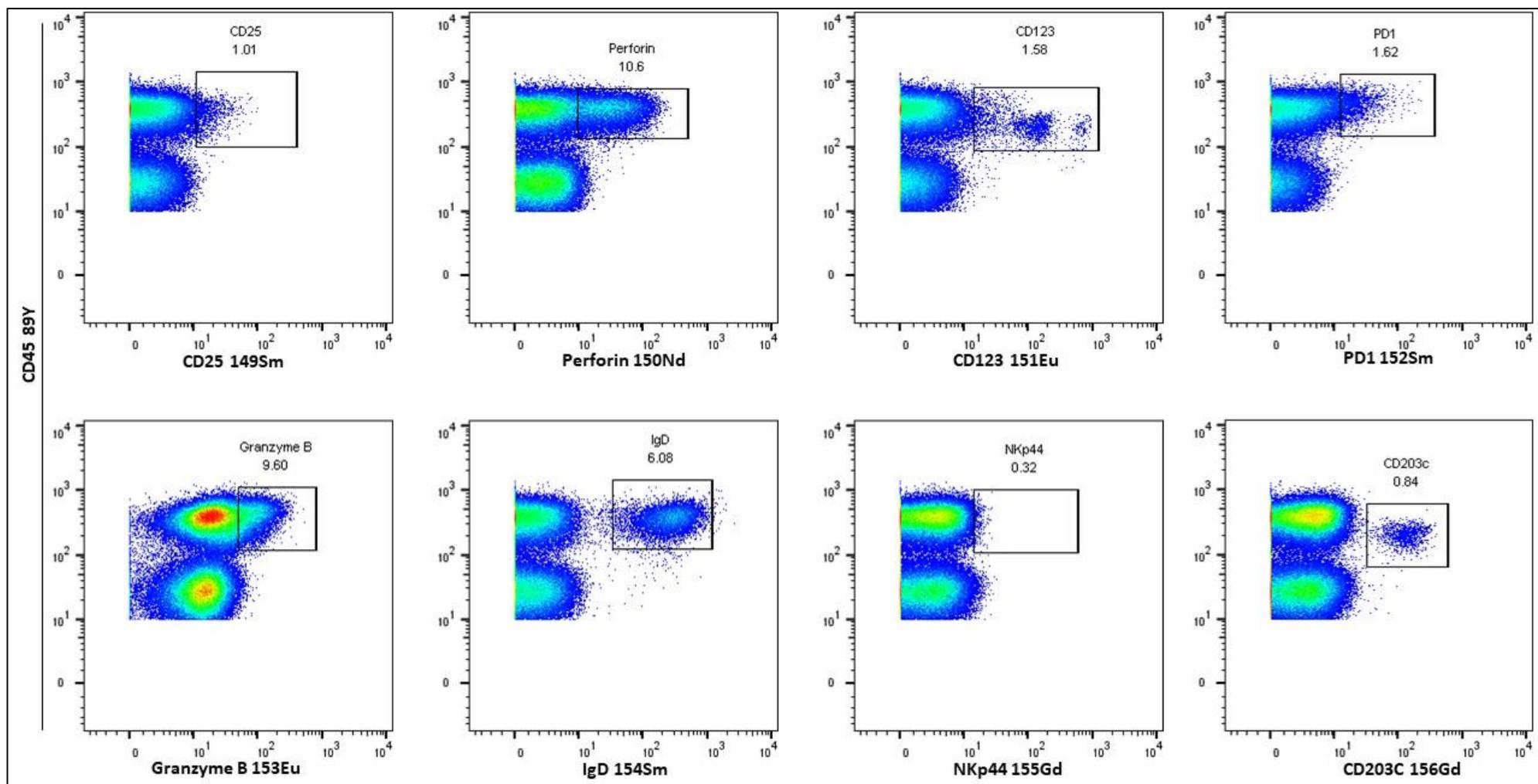


Figure 3-14 | Validation of lanthanide conjugated antibodies for mass cytometry using 1.25µg of antibody

After the gating strategy shown in figure 3-13, biaxial plots were drawn for each marker on the x-axis to assess the staining of the marker against CD45 on the Y axis. On the dot plot, each dot represents a single cell and density of a collection of single cells which equate to a population are denoted by blue, yellow, green and red colouration.



**Figure 3-15 | Validation of lanthanide conjugated antibodies for mass cytometry using 1.25µg of antibody**

After the gating strategy shown in figure 3-13, biaxial plots were drawn for each marker on the x-axis to assess the staining of the marker against CD45 on the Y axis. On the dot plot, each dot represents a single cell and density of a collection of single cells which equate to a population are denoted by blue, yellow, green and red colouration.

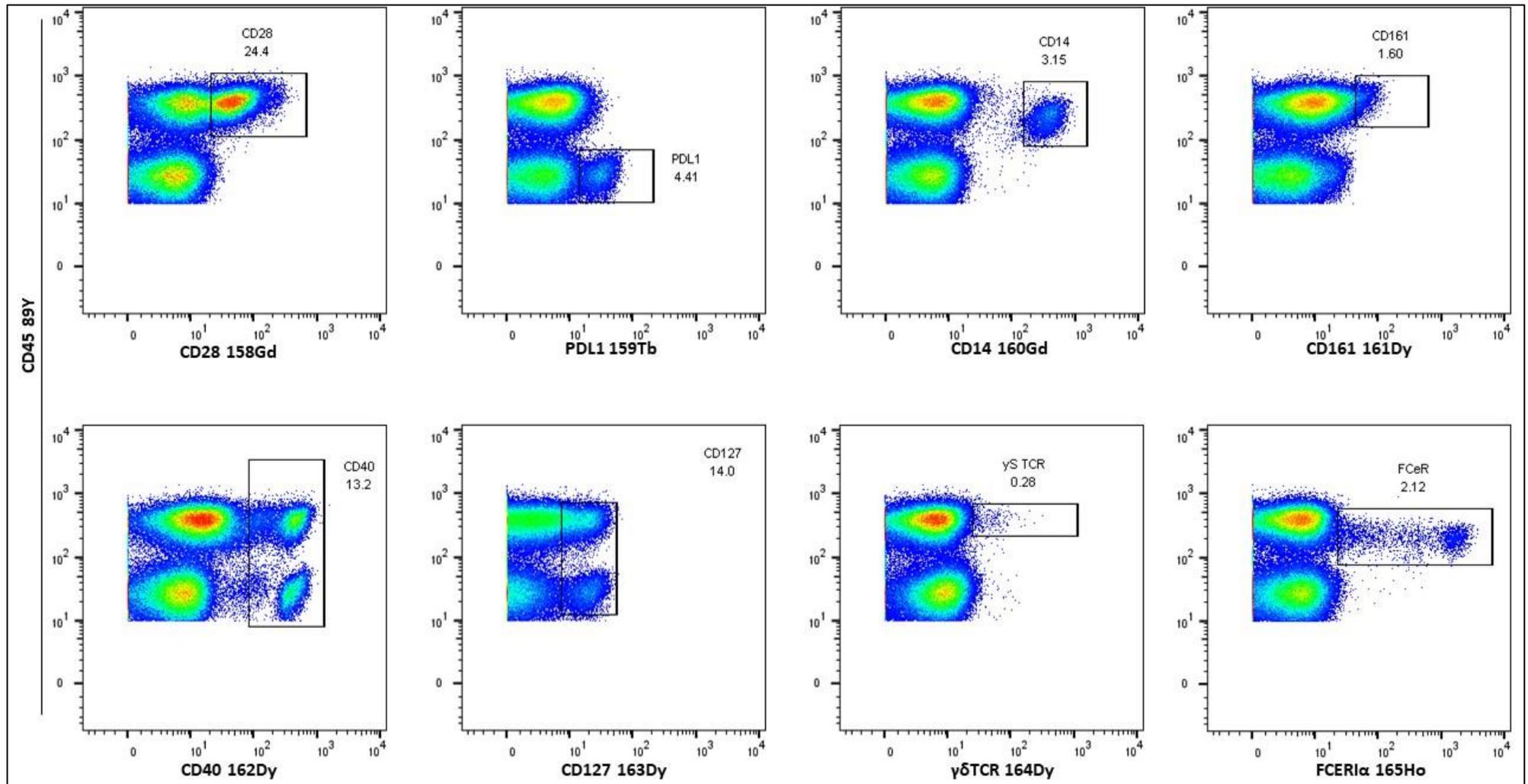
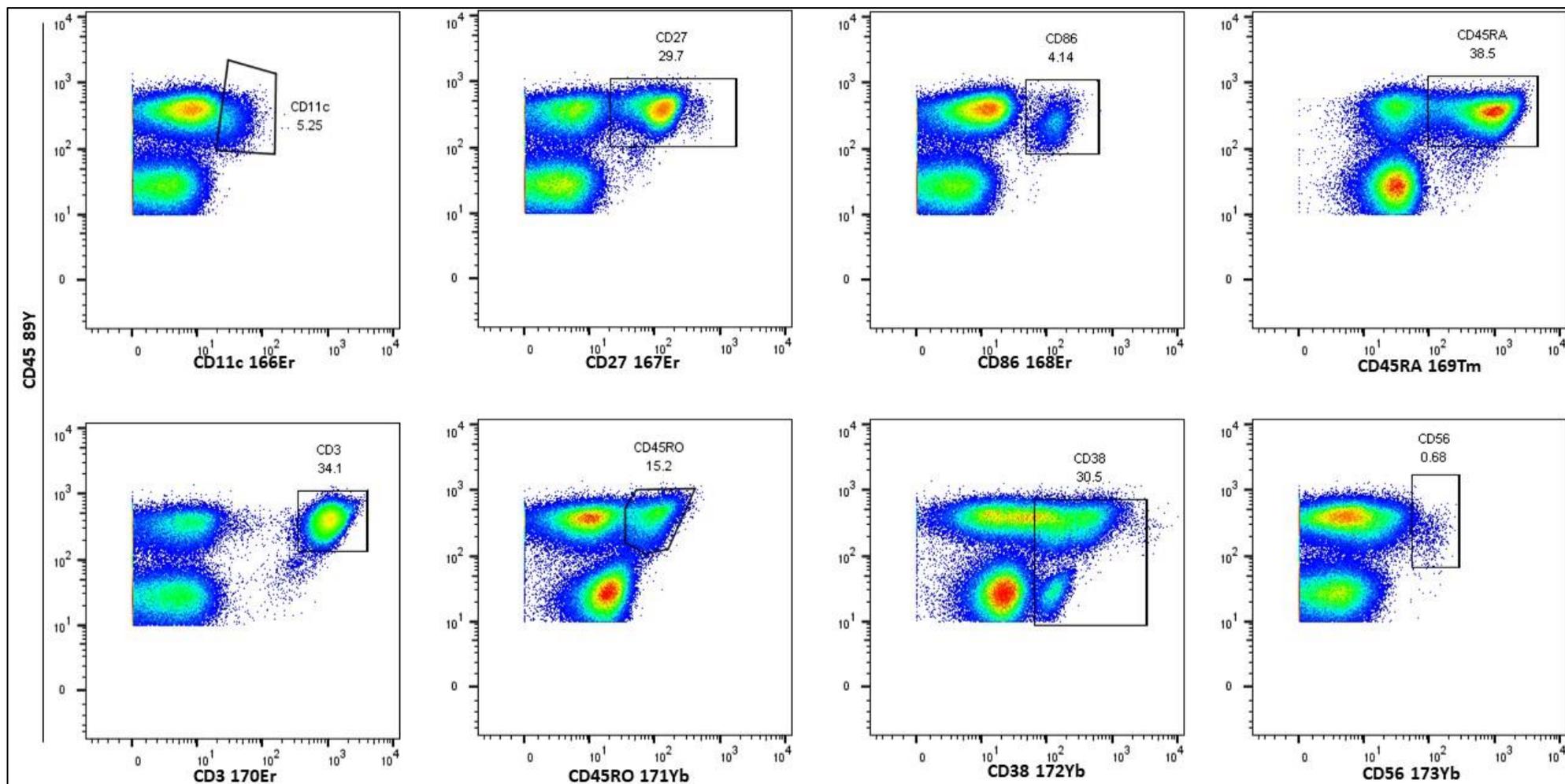


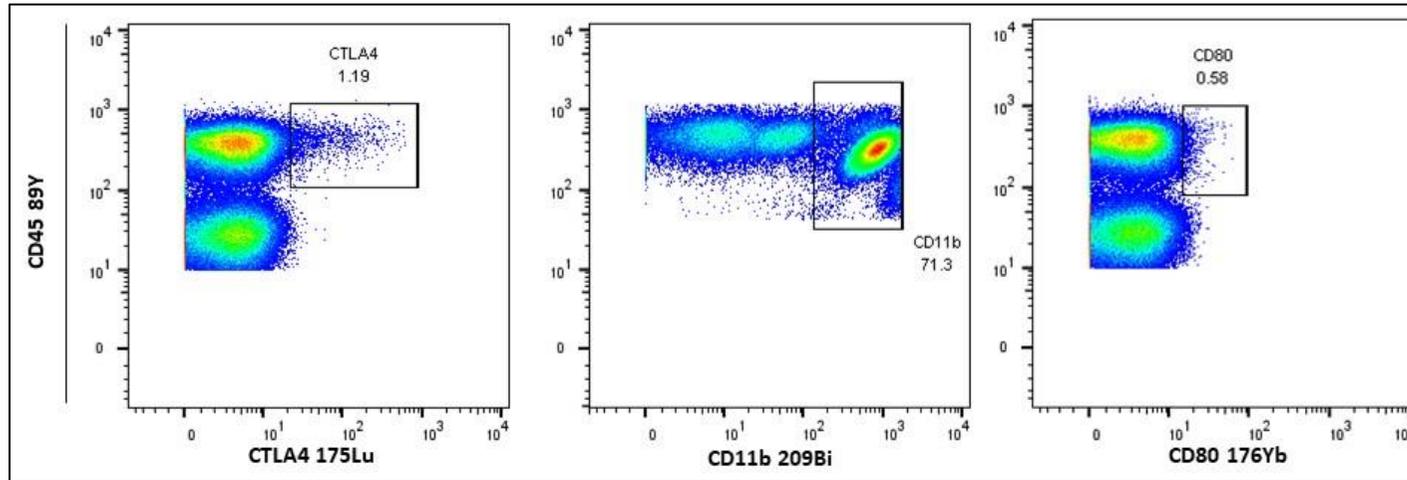
Figure 3-16| Validation of lanthanide conjugated antibodies for mass cytometry using 1.25μg of antibody

After the gating strategy shown in figure 3-13, biaxial plots were drawn for each marker on the x-axis to assess the staining of the marker against CD45 on the Y axis. On the dot plot, each dot represents a single cell and density of a collection of single cells which equate to a population are denoted by blue, yellow, green and red colouration.



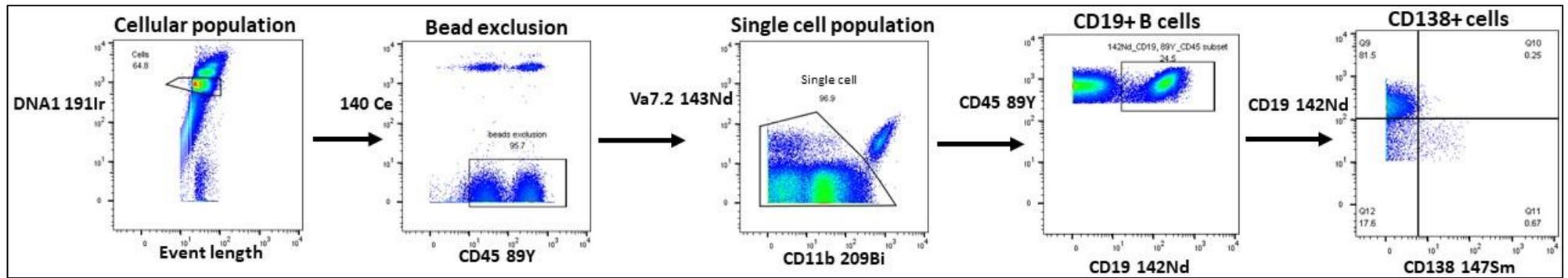
**Figure 3-17 | Validation of lanthanide conjugated antibodies for mass cytometry using 1.25µg of antibody**

After the gating strategy shown in figure 3-13, biaxial plots were drawn for each marker on the x-axis to assess the staining of the marker against CD45 on the Y axis. On the dot plot, each dot represents a single cell and density of a collection of single cells which equate to a population are denoted by blue, yellow, green and red colouration.



**Figure 3-18 | Validation of lanthanide conjugated antibodies for mass cytometry using 1.25 $\mu$ g of antibody**

After the gating strategy shown in figure 3-13, biaxial plots were drawn for each marker on the x-axis to assess the staining of the marker against CD45 on the Y axis. On the dot plot, each dot represents a single cell and density of a collection of single cells which equate to a population are denoted by blue, yellow, green and red colouration.



**Figure 3-19| Gating strategy for marker CD138 expression on plasma cells using 1.25µg of antibody**

Identifying the plasmablast cell population using the DNA1 191Ir and Event length parameters, followed by bead exclusion using the CD45 89Y and 140Ce parameters. A single cell population was obtained by gating against the CD11b 209Bi and Va7.2 143Nd parameters which eliminated non-specific cells. CD19 142Nd and CD45 89Y parameters defined the B cell population with plasmablasts being characterised as CD19 negative and CD138 positive as depicted by the final gate.

### 3.3 Discussion

Ensuring that the antibody clones selected for mass cytometry are compatible with the technique is a critical step in ensuring that the data obtained can be reliable. The aim of this chapter was to demonstrate that the clone of antibody selected was specific for the epitope on the antigen. Furthermore, it was demonstrated that the antibody clones selected for mass cytometry were capable of binding to the epitope under fix freeze conditions as confirmed by flow cytometry. A further verification step was carried out that ensured that staining profiles resembled the technical specification sheet as shown on the vendor's website and also published literature where these clones have been used in previous immunophenotyping experiments. This provides a reference profile for the expected staining to assess that staining profiles using mass cytometry are comparable to staining profiles achieved from flow cytometry and that irrespective of technology, staining profiles and thus downstream analysis are comparable. For mass cytometry, protocols provided by Fluidigm recommend validating antibodies prior to applying the panel to cohort experiments.

To maintain a consistent approach, the same donor blood was used in both fresh and fixed freeze blood processing to compare the effects of the different processing methods. If it were logistically possible, fresh sample processing yields the best data as the cell populations have had minimal handling. Fresh sample processing has the advantage of representing the immune cell population setting in its natural state. Fresh sampling would also allow for a viability dye to be added in mass cytometry before the application of fix freeze, however this was not possible as the samples were immediately fixed and frozen upon collection. However, the strength of the fix freeze protocol allows for a uniform and consistent method across all samples. Whole blood samples allow for neutrophils to be retained which are a cell population often removed due to their presumed short-lived status and abundance. However, having access to technology such as mass cytometry which allows for deep immunophenotyping of multiple cell populations, it would under utilise the opportunity if cell populations were to be removed.

The biaxial gating strategy performed in FlowJo, was deliberately simplified to assess epitope specificity. Data analysis from mass cytometry is reliant upon robust markers given that cellular profiles cannot be detected using light scatter properties. Sensitivity was measured by titration to ensure the optimal amount of antibody was used to detect a positive signal and thus separate the positive and negative populations. All markers were analysed against CD45, a marker of white blood cells, even for those markers which traditionally would subset a population for example CD4, which would typically be shown in a biaxial plot against a CD3 T cell marker. This broad gating strategy was applied to demonstrate the robust specificity of the markers but also anticipating the analysis approach used for high throughput cytometry data such as unsupervised clustering. Unsupervised clustering algorithms assess staining profiles of markers and according to similarity of which cell/s stain for particular markers, will place them in close proximity and thus cluster cells that are similar in staining. Clustering algorithms look across the whole panel of markers and place all markers against one another which overcomes the reductionist and limited approach of biaxial gating. This also enables identification of novel cell populations which may be overlooked from a biaxial gating approach. Thus it was deemed a holistic approach to assess each marker against CD45 to identify the entire staining profile of that marker in whole blood. It was not possible to run all 37 markers in one panel by flow cytometry so smaller panels were created.

The 37 markers which form the mass cytometry panel are those which have been well studied by cytometric analysis. However, some of the markers selected are based on their functional read out but these markers also require an activating stimulus to obtain a signal in a healthy donor. These markers include NKp44, PD-1, PDL1 and CD80 which are expressed at low levels in healthy unstimulated cells but increase upon activation in response to a stimulus or pathology.

Ascertaining the clones for each antibody marker was important to ensure the data obtained is reliable. In addition checking the conjugation of antibodies to their metal lanthanides and titration of these antibodies to determine the optimum signal to ensure that negative and positive populations could be distinguished are important for downstream analysis.

The next chapter will detail the methodology for analysing data from mass cytometry using bioinformatics.

## Chapter 4 Methodology for analysing mass cytometry data

### 4.1 Introduction

#### 4.1.1 High throughput data acquired by mass cytometry requires bioinformatics pipelines

The aim of this chapter is to describe and justify the analysis approach for data acquired from mass cytometry. The biggest challenge in any high throughput data and in this case, cytometry data, is how best to approach and select the most appropriate, efficient and replicable method(s) suited for the data. As the field moves away from traditional manual gating approaches<sup>162,237</sup> towards artificial intelligence and computer algorithms to model cellular subsets, the complexity of analysis becomes ever more challenging due to the continuously evolving and vast array of bioinformatics tools to select from. With little consistency in the field as to the best practices or guidelines for selecting the optimum approach, this can prove to be an overwhelming task for the researcher<sup>238</sup>. As quickly as the field is evolving, this too raises the issue of older algorithms and approaches becoming redundant; requiring a certain degree of flexibility to explore the field.

Before considering which algorithmic tools to use, it was important to check the quality of the antibody staining and to ensure that the data clean-up process obtained a single cell population of cells. This is a crucial step as a population of cells containing debris or doublets will distort the interpretation of the data by fluctuating the antibody staining. This experimental design included a batch control in each barcoded run to ascertain the antibody staining quality. Unanticipated technical challenges did arise during this project which resulted in antibody staining discrepancies across the batches. To overcome this, a normalisation algorithm tool developed by Sofie van Gassen<sup>175</sup> was performed through the R Studio platform to normalise the data which will be detailed in this chapter.

As data sets become increasingly complex, comprising of both large patient cohorts and immunophenotyping panels, this project being no exception, robust analysis pipelines are required. To date, there are no published reports that have acquired a clinical cohort as large as 187 barcoded samples on the mass cytometer. It was anticipated that the analysis of this data would take a significant amount of time, which was confirmed by the length of time it did take, 12 months, to understand the relevant findings from the data.

When faced with the dilemma of where to begin analysis of a complex data set, it can be of some reassurance to approach this initially by unsupervised learning. Unsupervised learning consists of computer algorithmic approaches which can compute independently of human supervision. The data set is uploaded into the algorithm and by machine learning, mines through the data, looking for cellular relationships irrespective of the data labels. An unsupervised learning approach is often combined with a supervised learning approach which consist of traditional manual gating techniques including Boolean gating and biaxial plots. The supervised component of the analysis is often used to verify the findings from the unsupervised approach to ensure that the cellular subsets are not an artefact of the algorithm but are bona fide populations. Dimensionality reduction, clustering and density mapping are all features of unsupervised learning. These allow for the researcher to obtain 'signposts' as to which cellular subsets or cellular immunophenotypes may be of interest to investigate further. This was especially useful for this project where the population of cells were

heterogeneous in the samples. The panel was designed to capture this heterogeneity and so an unsupervised learning in this instance was appropriate to perform a global analysis across the data. Furthermore, the blood and synovial fluid samples were left unstimulated; this was to capture the immunological landscape as close to its natural state in vivo and thus neutrophils were retained in the sample preparations.

The evolution of analysis for cytometry data has undergone extensive progress. As mentioned before, traditionally manual gating approaches through Flowjo or similar software have allowed for comprehensive analysis of single cell analysis most commonly depicted by biaxial plots. A considerable amount of our existing knowledge about the cellular immunophenotyping has come from this approach and it is an approach that is still currently used. Moreover, analysis through FlowJo has become the default approach when validating other approaches to ensure that automated algorithms yield comparable results to those achieved by manual analysis. However, the limitations of manual approaches are outweighing the benefits and these become more apparent when handling a large cohort with multiparameter dimensions. A second but equally major limitation of biaxial gating is that it does not readily convey the relationship amongst immune cell populations and thus struggles to capture the global immune landscape in a particular sample representing either disease or steady state i.e. health.

With this in mind, Cytobank was developed as a cloud-based platform that offers machine-learning assisted analysis of high dimensional single cell data <sup>239</sup>. This platform allows for multiple cytometric FCS files to be uploaded and accessed by anyone within a laboratory research group to explore the data in an intuitive manner. Whilst manual gating can be performed within Cytobank, the main attractive feature is interactive approach with data sets from single cell experiments. The result of this exploration leads to visually aesthetic plots, the most common being a tSNE or viSNE plot which conveys all the cell populations interrogated by the immunophenotyping panel in a globe shaped graph. These plots have become almost synonymous with cytometry data as they convey the cell populations that have been included in a quick, accessible format. A viSNE plot does not represent the relationship between cell populations and is thus only 2 dimensional in its output. Cytobank offers a number of clustering and visualisation tools to observe data including SPADE, flowSOM and Citrus which vary in terms of output but are all essentially clustering tools to visualise cell populations.

Cytobank has been regarded as an important data analysis platform however, in a rapidly evolving field it is no longer considered the gold standard of cytometric analysis. The need for visual clustering platforms whilst important, in essence only capture part of the cytometric analysis journey. To be able to interrogate which immune cell populations, if any have changed in a given condition is not necessarily apparent by clustering algorithms particularly if these changes are in rare cell populations or functional changes which may not be easy to discern from a global analysis. Additionally, Cytobank does not allow more than 1.3-2 million cells to be analysed at any one time, thus restricting the scope of analysis in large data sets. Another major drawback is that a license needs to be purchased if Cytobank is to be used frequently which for smaller research facilities can be a limitation. Despite these drawbacks, Cytobank has been adopted by many research groups due to its intuitive interface and cloud-based platform and its accessibility has increased in part due to it being recently purchased by the company Beckman Coulter.

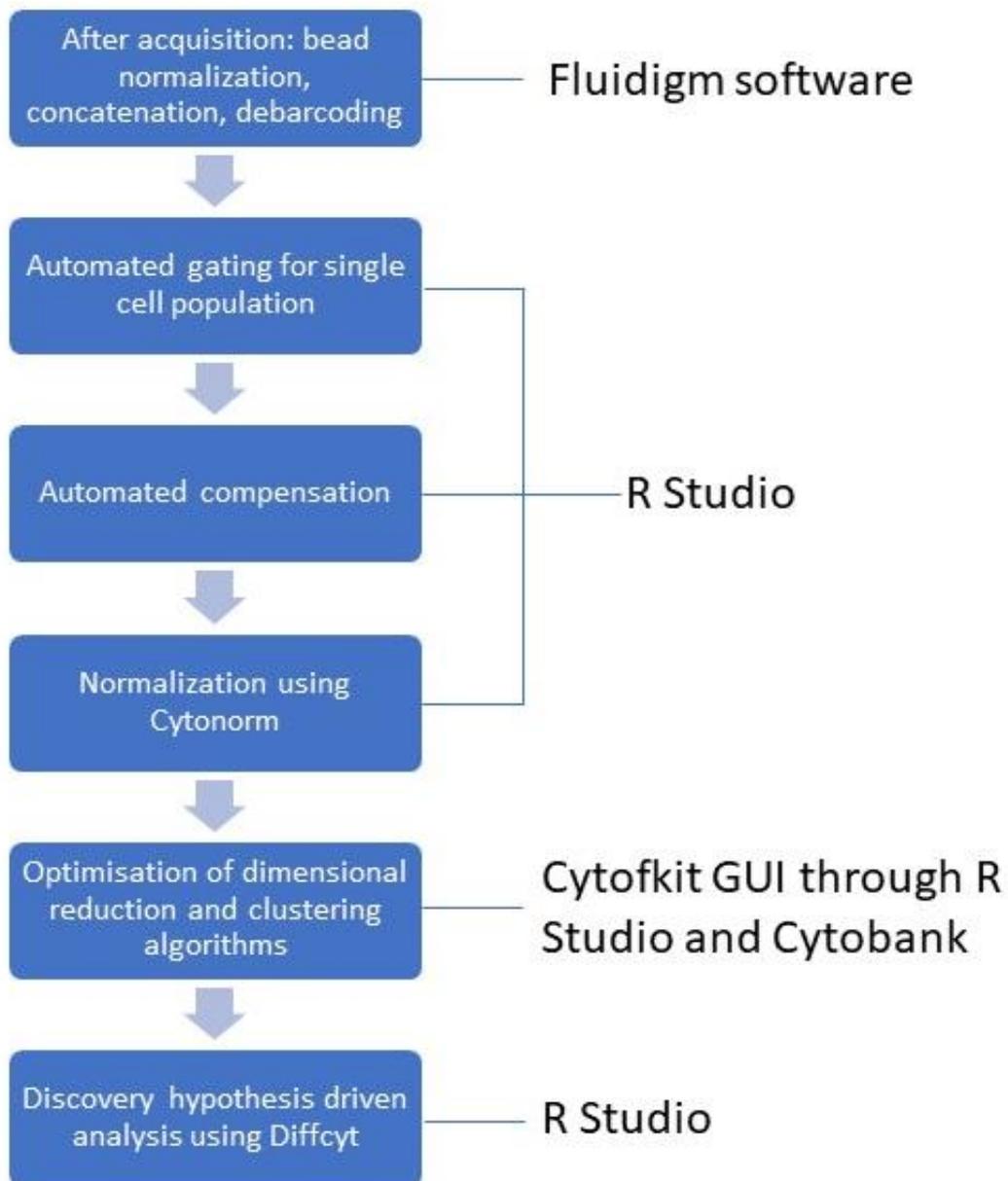
Cytobank is not the only software that allows for clustering algorithms with increasing options becoming available <sup>126,240,241</sup>. For example an extension of viSNE analysis is U-map which allows for the relationship between any single cell point to be assessed in the context of another cell point and

can also broadly capture cellular differentiation stages<sup>163,242</sup>. U-map analysis is currently not a feature of Cytobank and the seemingly endless options of algorithms can make it difficult to choose which one is most suitable for the data set in question. An automated machine learning tool SIMON (sequential iterative modelling 'overnight') allows for researchers to compare results from 128 different algorithms and is particularly useful for datasets which contain many missing values<sup>243</sup>. It is becoming evident that a combination of approaches may be required as one software or algorithm may not provide all the solutions. Original softwares such as Flowjo have improved with recent versions incorporating clustering features such as tSNE which also interacts with the traditional biaxial gating features enabling the user to precisely map where the gated population is on a high dimensional clustering plot.

The resources available to a research group influence the bioinformatics approach chosen. There are a number of considerations including the availability of in-house bioinformaticians, statisticians, computing power i.e. the availability of supercomputers or high specification computing processors able to compute a large number of events quickly.

Optimising the clustering algorithms took the most time in the whole analysis process. This is important because if a population is over or under clustered it will affect the interpretation of the cellular subsets and their relationships. I explored the clustering algorithms extensively, optimising cell number, markers and computing time which are detailed further in this chapter. Once the clustering optimisation was completed, I analysed the data through a scripted pipeline devised by the Robinson Bioinformatics group based at the University of Zurich, Switzerland called Diffcyt<sup>166,176</sup>. The analysis methodology for this data is summarised in Figure 4-1.

It is hoped that in a vast field of both bioinformatics and cytometry, this chapter will describe and explain the rationale in the approach for analysing this dataset, whilst additionally providing a starting platform or template for future (mass) cytometry analysis.



**Figure 4-1 | Methodology of analysis applied to data from mass cytometry.**

Flow diagram sequentially outlines the data analysis approach and the software used.

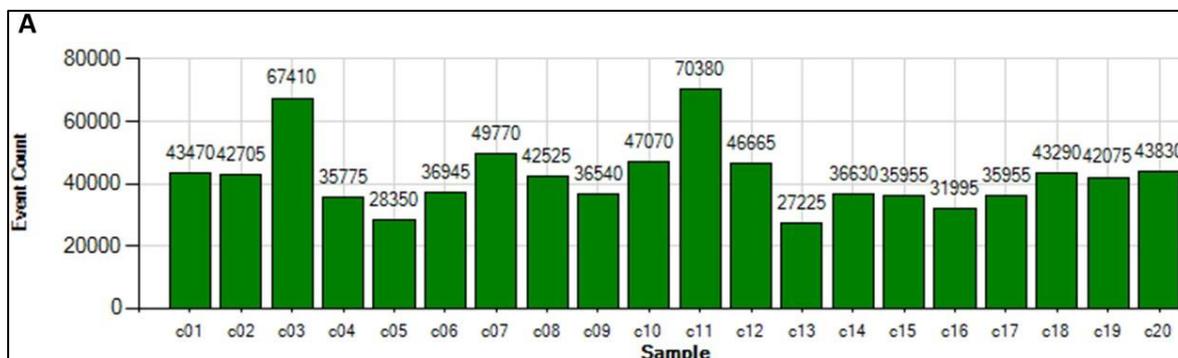
## 4.2 Results

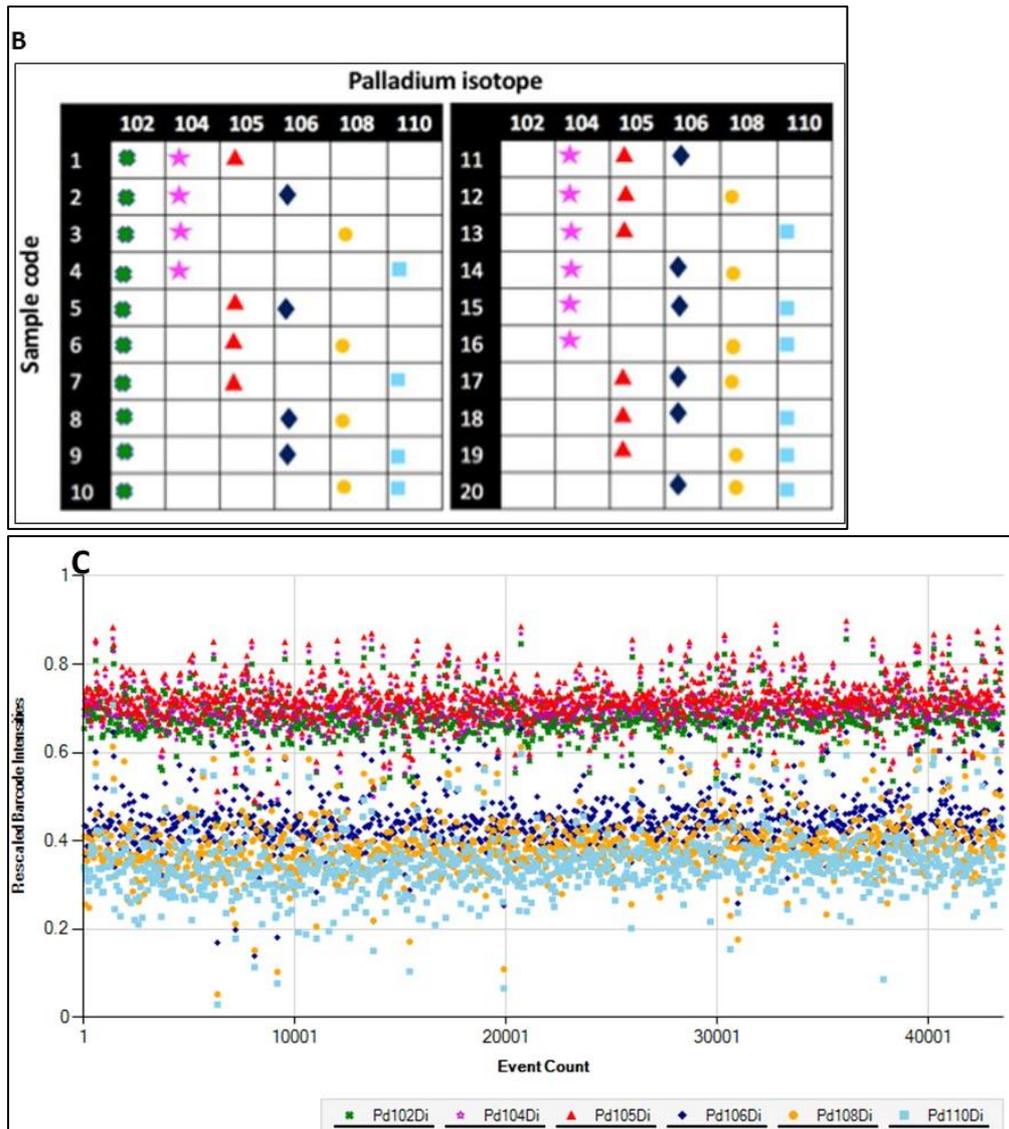
### 4.2.1 Data pre-processing

#### 4.2.1.1 Debarcoding samples after acquisition on mass cytometer

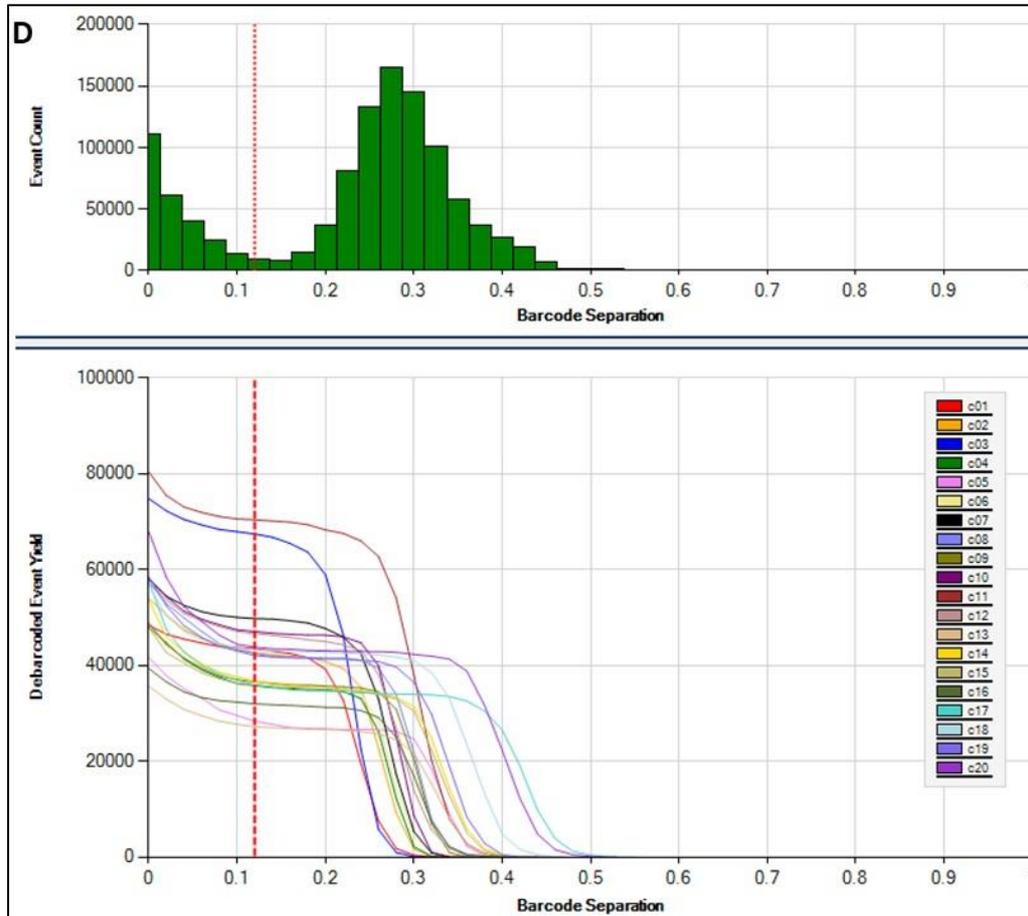
Barcoding has provided an opportunity to ensure uniformity across samples at a previously unmatched level. The limitation however, is that only 20 samples can be pooled together and barcoded simultaneously using a set combination of palladium isotopes. Efforts have focussed on alternative lanthanides such as tellurium which can potentially generate up to 35 unique barcodes<sup>244</sup>. Barcoding is relatively new method to mass cytometry although is used frequently in cohort studies run on flow cytometers. However, the barcoding technology for mass cytometry has proven to be a relatively straight-forward concept and Fluidigm have provided the methods for wet lab and also an automated software which allows for the debarcoding of individual samples.

To prepare samples for barcoding, 3 million cells are counted from each sample which is an important step to ensure that during acquisition, the number of cells acquired on the mass cytometer are equal for all samples. If the number of cells are disproportionate between samples this will distort the collection of events and result in uneven collection of events across the samples. In Figure 4-2A, the Fluidigm debarcoding software allows the researcher to analyse the number of events collected for each sample in a histogram plot with the barcode sample number across the x axis and the event count on the y axis. Figure 4-2B shows the 20 combinations of palladium isotopes for labelling 20 individual samples. The palladium isotopes are provided in a combination of three different palladium isotopes for each barcode.





*(figure legend on next page)*



**Figure 4-2 | Fluidigm debarcoding software allows samples to be individually identified.**

The figures (A-D) are from a pilot experiment with 20 barcoded patient samples. The software is provided with the Mass cytometry machine and can be downloaded from the Fluidigm website <https://dvsscience.com/> **A** The event count for each sample within a barcoded batch of 20 samples (C1-C20 along the X-axis) can be visualised to assess uniformity of event collection across the samples. The uniformity of the cell number is variable due to the technical steps which can result in cell loss. In addition, during acquisition, cell event loss can occur due to machine blockages. **B** Palladium isotopes are used to barcode up to 20 samples which equals one batch. The schematic demonstrates the combination of 3 isotopes used for barcoding each sample as prepared by Fluidigm and for subsequent debarcoding. **C** The zoomed-in dot plot shows the barcoding staining intensity. The three palladium isotopes that are used to barcode each sample can be identified between 0.6 and 0.8 on the Y axis and can be assessed for each individual cell collected for that sample to ensure consistent staining. This dot plot shows for event counts between 10001 and 40001 a uniform barcode stain. Events that have picked up non-specific staining can be viewed below 0.6 on the rescaled barcode intensities axis and are removed from the file. **D** The histogram depicts the Barcode separation cut off which is set at 0.12 to remove debris and capture cell events. The number of cells (yield) obtained after setting the barcode separation cut off is shown below.

When debarcoding each sample using the debarcoding tool, it is important to analyse the specificity of the barcode staining. For each sample, the intensity of the barcode staining can be observed by zooming into the event count for each sample and ensuring that the events included for the sample are stained by the specific palladium isotope combination. In Figure 4-2C, the barcode intensity staining is shown for events collected between 1 to 40,001. The combination of three palladium isotopes used to stain the sample can be precisely analysed where any contaminating isotopes can be removed. Each event can be analysed if deemed necessary but the precision of the barcode staining is high so to check, every 10,000<sup>th</sup> event was analysed to assess precision of staining intensity for events.

To further aid identification of events for each sample, the barcode separation histogram can be adjusted (Figure 4-2D-E). In this investigation, it was unnecessary to go further than the barcode separation index which was set at 0.12 for all the barcoded samples. This is similar to that of a meltcurve for real time polymerase chain reaction experiment, except the barcode separator distinguishes between debris and the start of events. There is always debris at the beginning to account for the ions and particles that are present before the sample events are collected and these are efficiently removed by the barcode separation tool. If the barcode separation index is insufficient or the sample is particularly contaminated by non specific barcodes, the Mahalanobis distance can be applied to filter outliers taking into the account the covariance of the barcode populations <sup>245</sup>. The Mahalanobis distance was not applied to this data as the barcode staining had few outliers after the barcode separation index was set.

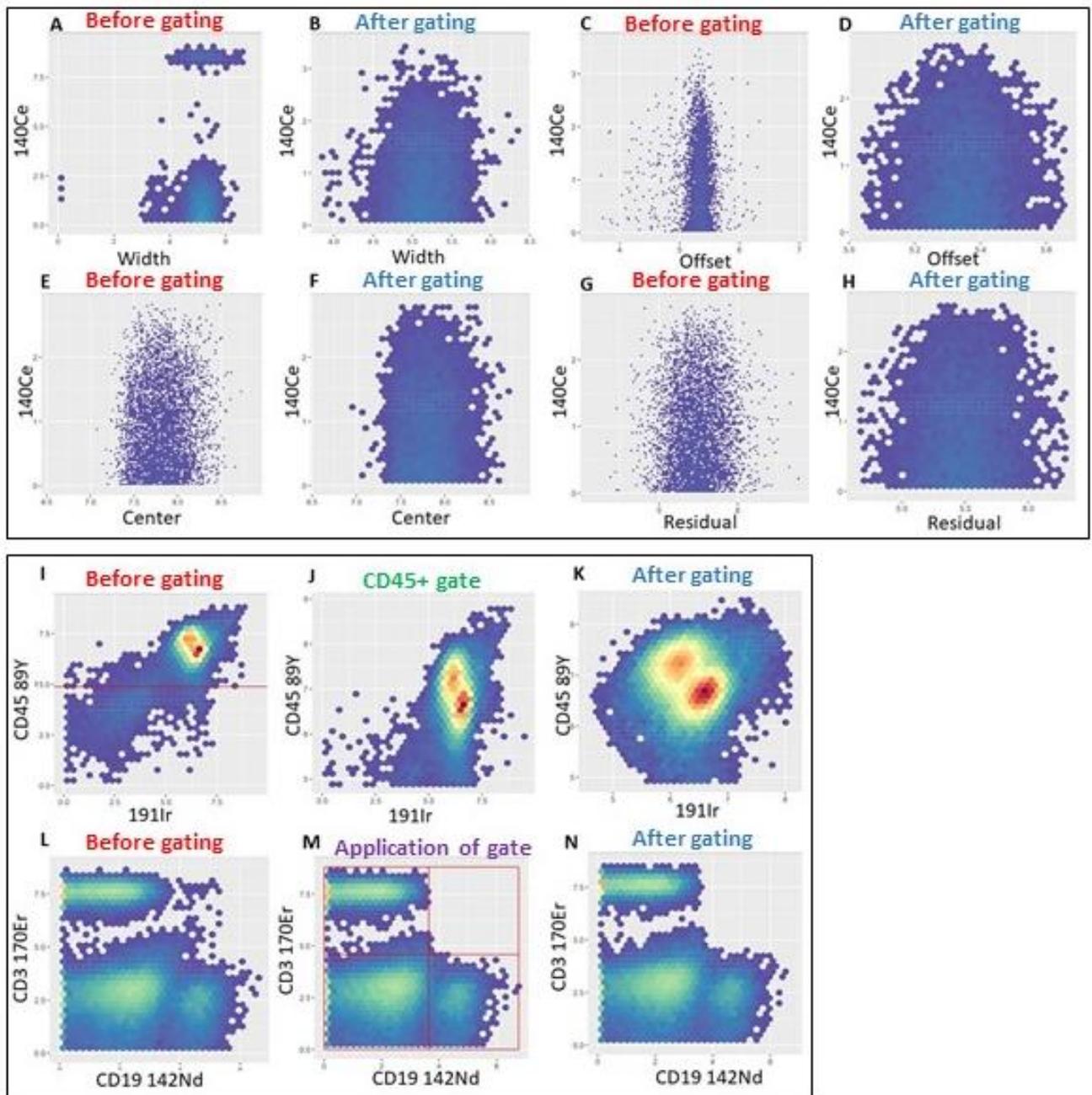
#### **4.2.1.2 Identifying and obtaining single cells using Gaussian parameters for mass cytometry**

Whilst flow and mass cytometry are technologies both designed to interrogate the single cell population, the principles of both technologies are fundamentally different. This includes the acquisition parameters for identifying the cell population. As a starting point, identifying the general cellular population on a flow cytometer normally includes the forward and side scatter parameters which is a redundant principle in mass cytometry. However, defining a 'cell' in mass cytometry requires additional parameters to be used as part of the data cleanup process to begin the analysis of a single cell population. Whilst identifying a single cell population was broadly described in Figure 3-13, the process of data cleanup has become more complex including the removal of debris and dead cells (a viability marker was not included in this study), normalisation beads and coincidental ion clouds from raw data. When cells are injected into the hot argon torch, the plasma formed strips the electrons off the atoms. In turn the positively charged ions from the lanthanide labelled antibodies collide with the ion detector to form electron pulses that are then converted into voltage pulses <sup>246</sup>. The event length parameter is able to filter out pulses that are either incomplete or have an abnormally long lag time between signal pulses. In brief, the event length parameter is capable of eliminating unwanted noise <sup>246</sup>. A cut off point is calculated that an internal threshold less than 10 would not be included in the event length and no more than 150 consecutive pushes of an event would be calculated <sup>246</sup>. Event length is usually measured in combination with DNA1 and/or DNA2 with researchers often using both DNA channels to avoid any discrepancies between the signals. DNA1 and DNA2 are derived from the cationic double-stranded nucleic acid intercalators that have a natural abundance of iridium 191Ir and 193Ir <sup>246</sup>. When optimising the gating strategy for this data set, no difference was observed between DNA1 and DNA2 and therefore gates were drawn using DNA1.

Whilst the event length and DNA intercalator parameters are robust parameters for identifying single cells, four additional measurements have been identified that can improve the clean up of the data and eliminate signal noise more accurately. These four parameters are collectively known as the Gaussian parameters: width, offset, center and residual (Figure 4-3). These parameters are used to discriminate between noise and signal ratio of the pulse recorded by the mass cytometer which also helps in double removal. Gaussian parameters reduce the bias of gating strategies that are adopted by the individual researcher for example performing numerous biaxial plots with different combinations of markers to eliminate doublets. These Gaussian parameters have been validated and verified by two approaches: the software Gemstone <sup>247</sup> and Fluidigm. In the automated gating script applied to this dataset, width was the first Gaussian parameter to be incorporated which

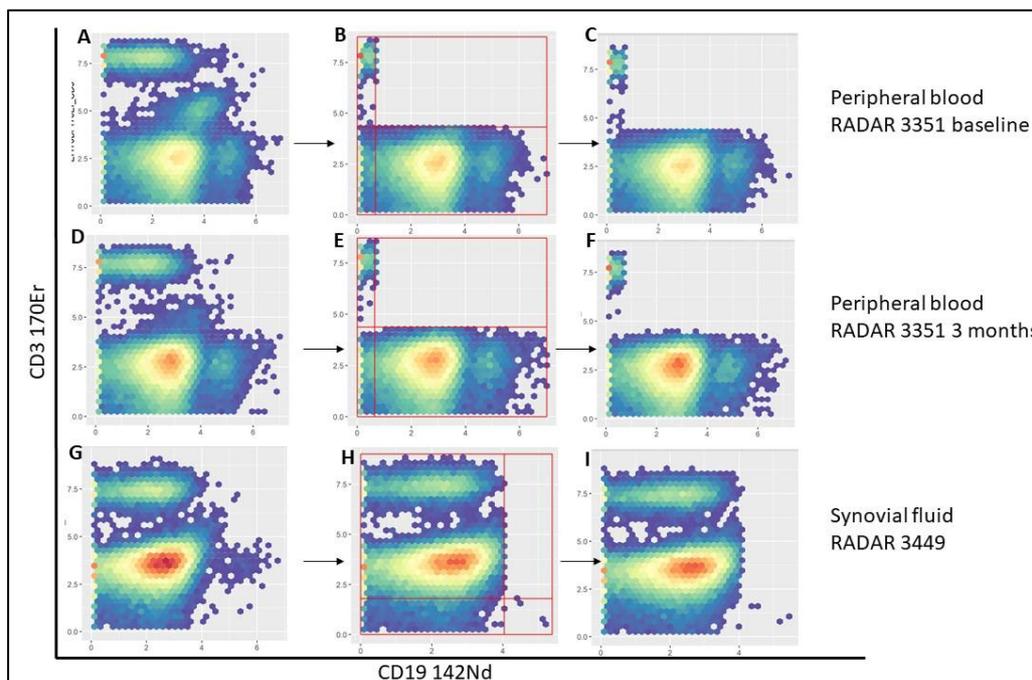
distinguishes the occurrence of ion clouds. The offset parameter determines the presence of multiple-peak pulses and center defines the Gaussian fit of the signal pulse<sup>246</sup>. The last Gaussian parameter applied to this dataset measures the differences between the Gaussian model and the pulse and thus assesses the normal distribution of the pulse emitted<sup>246</sup>.

The automated gating script was applied to all 182 samples to be analysed through R studio of which only 3 samples failed shown in Figure 4-4. Two samples were from the same barcode batch 3 and were taken from the patient and the third sample was synovial fluid. The samples passed all of the automated gating script except the last step which was to draw the final doublet exclusion gate. In that gate, for all three samples, the CD19+ B cell population were gated out in the final .fcs output file. It is unknown why this error occurred but it was decided to remove these three samples from further analysis to ensure consistency and uniformity downstream of the analysis process.



### Figure 4-3 | Automated gating strategy using Gaussian parameters for identifying single cell population

The above image sequentially details the automated gating approach (A-K) through the R Studio platform for obtaining a single cell population using Gaussian parameters width, offset, center and residual against the parameter 140Ce for one patient whole blood sample. These four additional parameters are integrated within the mass cytometer software and calculates the measurements that provide information about the quality of the total ion current pulse. These parameters are stored as FCS 3.0 measurements. Offset is a Gaussian discrimination parameter that is low for multiple-peak pulses. Width is a Gaussian discrimination parameter that is generally low for coincident ion clouds. Center is related to the mean of the Gaussian fit of the signal pulse. If a pulse has two peaks due to coincident ion clouds, where the first is the highest, center is relatively low, whereas if the second peak in the pulse is highest, then the center is high. This gating approach was applied to all samples simultaneously following acquisition on the mass cytometer and subsequent debarcoding using the Fluidigm software. **A** shows the width Gaussian parameter before it has been applied to the sample and **B** shows the cell population once the width parameter has been applied. It can be observed that the bead population present in **A** has been removed resulting in a more defined peak where the cut-off was drawn between 4 and 6 for the width parameter and at 3 on the 140Ce parameter resulting in a defined peak removing unspecific events. **C** shows the offset Gaussian parameter prior to being applied to the sample. A solid peak is visible between 5 and 6 on the offset scale and 2.5 on the 140Ce parameter where events on the outside of this peak will be removed due to them being part of the ion cloud. Events outside the peak represent ion clouds **D** shows the cell population once the offset parameter has been applied to the sample **E** shows the cell population before the center parameter has been applied. It is evident here that the peak is 'cleaner' meaning that ion clouds or unspecific events are not visible compared to that observed for Width and Offset parameters **F** shows the cell population once the center parameter has been applied. Cut-off was defined between 7 and 8.5 on the Center parameter and between 0 and 2.5 on 140Ce parameter **G** shows the cell population prior to the residual parameter being applied. The cut-off was determined between 5 and 6 on the residual scale and between 0 and 2.5 on the 140Ce scale **H** shows the cell population once the residual parameter has been applied to the sample **I** After defining the single cell population using the Gaussian parameters, the CD45+ cell population against the DNA intercalator 191 Iridium (Ir) channel was defined. The cut-off to remove debris was set at 5 on the CD45 89Y axis and **J** shows the CD45+ 191Ir+ cell population once the CD45 cut-off has been applied and **K** shows the final CD45+ 191Ir+ cell population when the 191Ir cut-off determined at 5 **L** shows doublets in the CD3 170Er CD19 142Nd double positive gate. **M** To remove the doublet population identified in image L, CD19+ population cut-off gate was determined at 3.8 and the CD3 cut-off gate was determined at 4.8. This removes the doublet population whilst also identifying a single population of T cells (determined by CD3) and a single population of B cells (determined by CD19). The CD3 negative and CD19 negative populations are also present in the lower left quadrant of the plot. **N** The gates calculated in images L and M were applied to remove the CD3+ CD19+ doublet population to obtain a single cell population for downstream analysis.



#### **Figure 4-4 | Automated gating for obtaining a single cell population can be applied on a large data set**

Automated gating was applied simultaneously on 187 patient samples obtained on the mass cytometer. This was successful excluding three samples **A-C** RADAR 3351 blood sample baseline and **D-F** RADAR 3351 blood sample follow up time point at 3 months and **G-I** 1 synovial fluid sample did not pass automated gating. The gating step that failed for all 3 samples was at the final step when applying the gate for removing doublets using the CD3 170Er and CD19 142Nd axes. An incorrect gate is drawn resulting in the removal of the CD19 B cell population. These 3 samples were removed from further analysis.

#### **4.2.1.3 Compensation for mass cytometry data**

The concept of compensation is closely associated with flow cytometry where the fluorescent signal is not only measured in the primary channel but also in neighbouring channels due to overlapping excitation and emission spectra of fluorescent dyes<sup>248</sup>. This concept is known as 'spillover' where signal is recorded non it should be in adjacent channels and can be mathematically corrected by correlating the original signal in an approximately linear manner known as compensation<sup>248</sup>. Software to correct compensation is easily accessible such as FlowJo however, panels that consist of more than 15 markers become harder to compensate. Flow cytometers with the capacity to phenotype more than 28 markers have been developed but currently not widely available<sup>249</sup>. Therefore the advent of mass cytometry has provided not only the prospect of immunophenotyping 40 markers but also significantly reducing spillover.

However, using mass cytometry does not eliminate spillover altogether as whilst fluorescence is not a factor to consider, spillover can still occur by three main sources: abundance sensitivity, oxidation and isotopic impurities<sup>250</sup>. Abundance sensitivity is the spillover results from imprecisions in ion detection caused by asynchronous movement of identical ions (e.g. 145Nd ions detected in 146Nd or vice versa) at the initiation of acceleration thus resulting in metal  $\pm 1$ <sup>250</sup>. Abundance sensitivity is checked during the calibration set up time of the mass cytometer prior to acquisition. The cut off threshold for abundance sensitivity is 0.3% for 159Tb of a primary signal. If the abundance sensitivity accelerates the threshold the calibration would need to be repeated until 0.3% or less was achieved. Oxidation is an unavoidable as it occurs due to the plasma ionization of isotopes and results in metal+16 interference<sup>250</sup>. As oxidation is undesirable, it is also measured during the daily instrument calibration prior to sample acquisition. If the oxidation interference is less than 3% of signal measured from the reference isotope 139La isotopes (as these are easily oxidised) in tuning solution the mass cytometer is considered to have minimal oxidation levels and thus does not need to be recalibrated<sup>250</sup>. Isotopic impurity occurs as a result of isotopes not being 100% monoisotopic. Challenges in elemental enrichment can exist as it is not always possible to enrich every isotope to 100% purity<sup>250</sup>. Lastly, isotopic impurity can be measured accurately by conventional inductively coupled plasma mass spectrometry which can detect signal contamination which is the most prominent source of overlap in mass cytometry assays.

Chevrier et al., in 2018 designed an algorithm to address spillover issues from mass cytometry data through the CATALYST R/Bioconductor package and an interactive Shiny-based web algorithm which can be used to correct for spillover<sup>251</sup>. Using polystyrene compensation beads, each conjugated antibody was added separately to these beads and then mixed together and acquired on the mass cytometer. The algorithm incorporates the non-negative least-squares (NNLS) approach which corrects for spillover that occurs in empty channels without changing the data structure and applies semi-automatic spillover corrections<sup>251</sup>. Each bead is assigned to a specific population based on the dominant signal in the corresponding lanthanide channel and sample cut-offs are automatically

applied. A spillover matrix is generated based on the single stained populations and takes into account abundance sensitivity, oxidation and isotopic impurity<sup>251</sup>. The spillover matrix is then applied to the beads and samples to correct for compensation (Figure 4-5 A and B). When applying the compensation matrix to the samples from this experiment, there was not much difference between uncompensated data and compensated data (Figure 4-6 A, B, C, D) as evidenced by manual gating in FlowJo. This was expected as compensation issues in mass cytometry are significantly reduced but as part of an optimised protocol for the analysis of this data, compensation correction was included.

Designing the panel and assigning the metal lanthanide to each marker requires careful consideration to reduce the impact of factors affecting compensation. For metals that have multiple isotopes such as neodymium (Nd) which has 5 stable isotopes, markers should be carefully assigned where the expression pattern is well characterised. Furthermore markers which can have intermediate levels of expression such as HLA-DR should not be placed on channels that are more susceptible to oxidation or contamination as this could negatively influence the interpretation of the data. For this study, it was not a difficult task to assign markers to channels meaning that the data was largely unaffected by compensation issues. However, for studies that incorporate more markers which may require the use of channels which are more susceptible to oxidation, compensation becomes an essential part of the data clean up process.



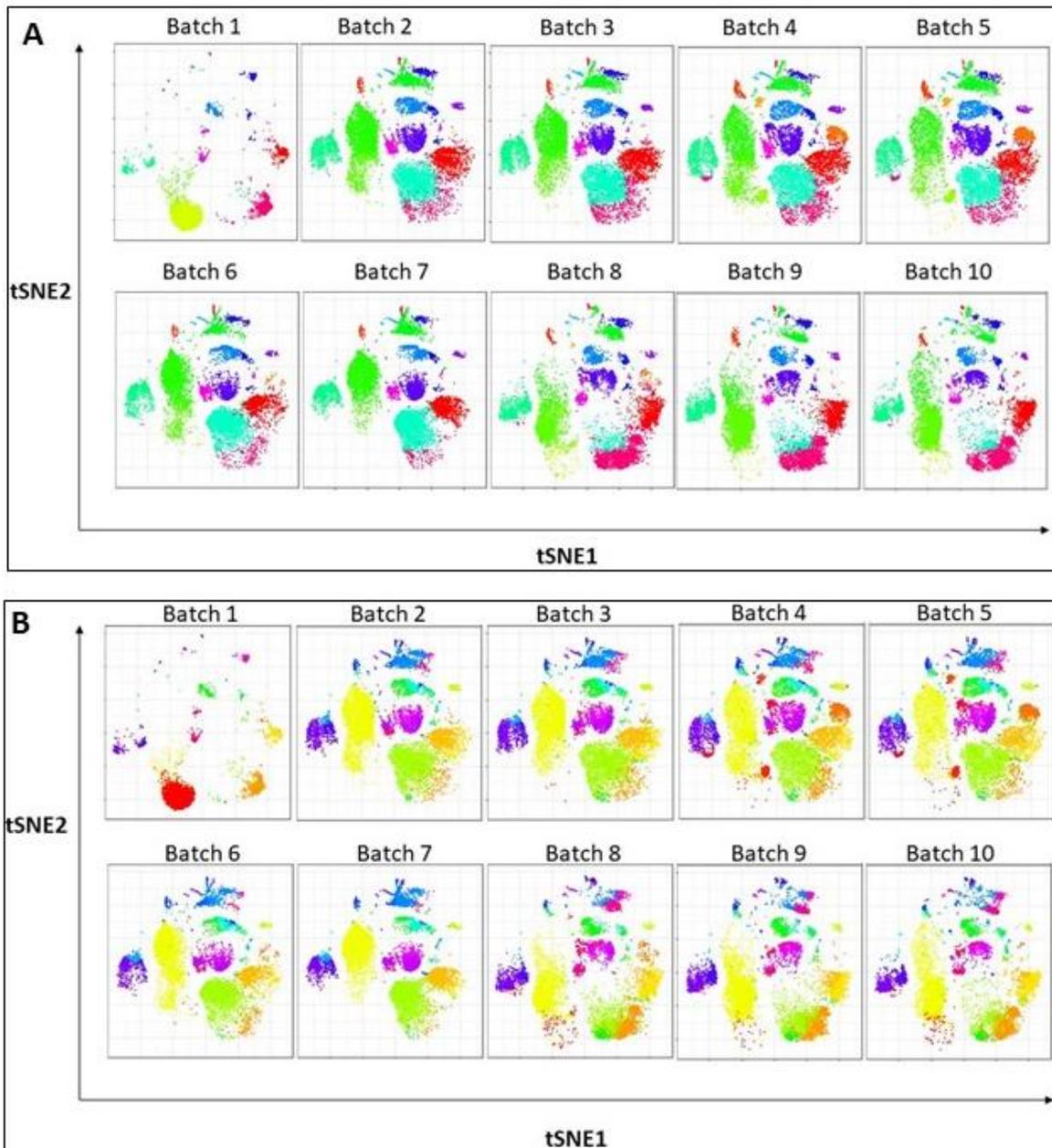
## Figure 4-6 | No difference can be observed between uncompensated and compensated channels

The compensation matrix shown in the previous figure was applied through R studio to all the samples in this study. Compensation was performed by adding each antibody to OneComp compensation beads and pooling these together in a single tube and acquiring 10,000 events on the mass cytometer. The compensation matrix algorithm initially applied a semi-automatic spill over correction for mass cytometry created using the R/Bioconductor package, CATALYST<sup>251</sup>. In this step, the FCS files are deconvoluted to identify the single antibody-positive beads and each bead is assigned to a specific population based on the dominant signal, and the purity of the bead populations is further increased by automatically applying sample-specific cutoffs<sup>251</sup>. The second part of the algorithm looked at determining the spill over at the single-bead level and by default it considers interference between channels expected to interact based on abundance sensitivity, metal impurity and oxidation but can also check for ad hoc spill over that has occurred during the acquisition. In the third and final step, the compensation matrix from the solved linear system and applies this to the bead and cell samples to remove unwanted signal spill over<sup>251</sup>. To look at the effect of spill over on the channels, manual gating checks were performed. As an example, **A** and **B** show the uncompensated and compensated CD3 and CD19 channels respectively and **C** and **D** show the uncompensated and compensated CD4 and CD8 channels respectively. No differences can be detected in the expression profiles between uncompensated and compensated plots. This was expected as the panel was designed to minimise/avoid spill over including placing markers with different expression profiles (as informed by literature) on isotopes of the same metal. Furthermore, the lack of spill over supports the main advantage of using a mass cytometer in that compensation is a modest issue and large immunophenotyping panels can be successfully designed without the cumbersome task of compensation.

### 4.2.1.4 Clustering algorithms show batch variation

Before running the data set through further analysis for hypothesis driven testing, it was important to ensure that batch effects from variation of antibody staining were undetectable. 10 barcoded batches, washed, counted, stained and frozen on the same day were prepared with each plate containing a consistent internal batch control from the same healthy donor to account for variation. Using the cytofkit graphical user interface (GUI) through a new Bioconductor package in R Studio, which requires only two lines of code<sup>252</sup>. Other software packages are available to visualise data from mass cytometry namely the cloud based platform Cytobank<sup>253</sup>. However, as mentioned at the beginning of this chapter, Cytobank has limitations chiefly in terms of the limited number of events that can be processed simultaneously, inflexibility to tailor algorithms for the dataset in question and the need to purchase a license annually. Cytofkit allows for high dimensionality data to be clustered into cell populations based on similar antibody staining patterns and machine learning which trains the algorithm to recognise specific patterns and cluster a heterogeneous population of cells appropriately. The result is that these clusters are presented in a visual format which can be readily interpreted in terms of the cell populations present in a given sample.

For comparison, two different unsupervised clustering algorithms were chosen in Cytofkit, Phenograph and FlowSOM (self-organising map). The initial analysis was to observe whether there was any batch variation as a direct consequence of antibody staining variability by observing the internal whole blood batch control (same donor) which was included in each barcoded batch (labelled 1-10) in Figure 4-7. The Phenograph algorithm uses a graph-based partitioning method which is efficient not only in detecting cell populations but also in identifying subpopulations<sup>252,254,255</sup>. The Phenograph algorithm initially constructs a nearest-neighbour graph which looks at the populations of cells and measures their phenotypic relatedness and accordingly applies a graph partition using the Louvain algorithm<sup>252</sup>. FlowSOM is another unsupervised technique for clustering where similar cells are assigned to the same node thus creating a minimum spanning tree. These nodes are then grouped into metaclusters using hierarchical clustering effectively grouping the data into cell populations<sup>256</sup>.



**Figure 4-7 | Batch variations are observed by two different clustering algorithms**

A semi-automated interface using the cytofkit GUI Shiny App in R Studio, clustering algorithms, **A** Phenograph and **B** FlowSOM clustering algorithms identify batch variations using the internal batch control which was present for each of the 10 barcode batches. Different colours represent different cell populations where clustering of many single points (which represent a cell) create a cluster as determined by the staining intensity calculated by the algorithm. The main purpose of this figure is to convey that the batch control which was identically processed shows variation across the batches.

Both Phenograph and FlowSOM clustering algorithms, using 10,000 cells from each batch control, showed that antibody staining variations were detectable across the batches, particularly in batch 1 where the cell populations are vastly different compared to those in batches 2-10. All batches were prepared simultaneously and subsequently cryopreserved at  $-80^{\circ}\text{C}$  so it is not considered that significant batch variation was introduced at these stages. However, due to the logistics of acquisition, each batch had to be thawed and acquired individually on the mass cytometer. Whilst the acquisition of the batches should not vastly differ, it was observed that during the acquisition of

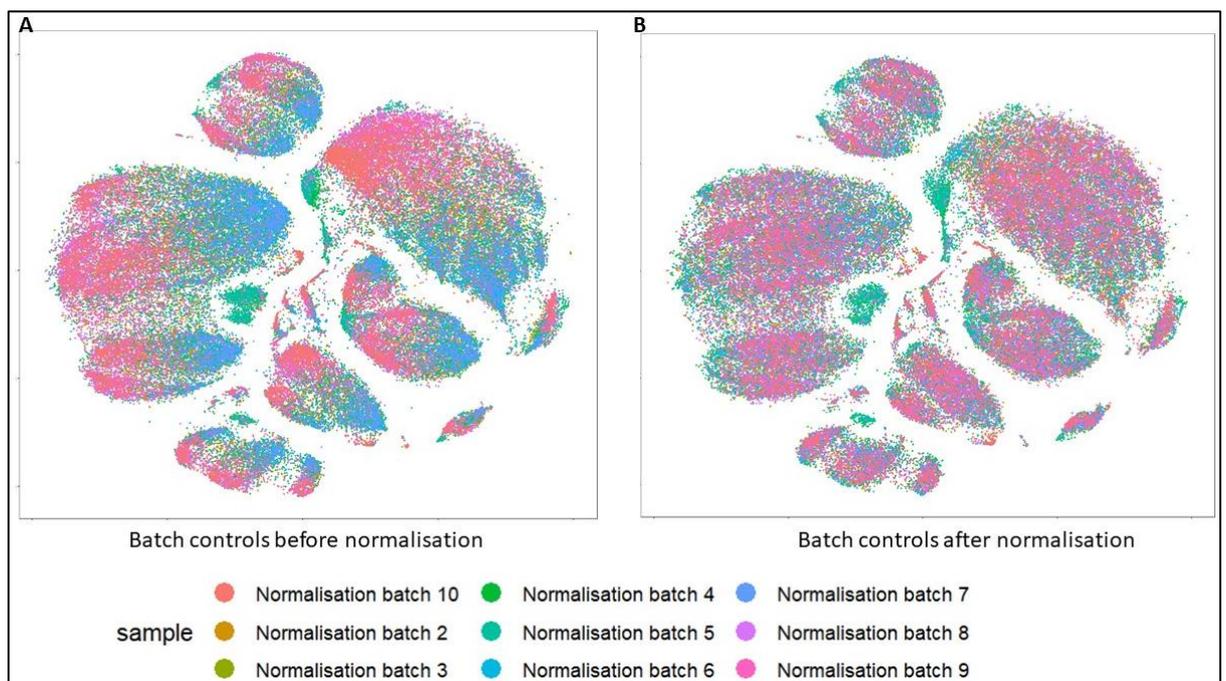
batch 1, the sample loop consistently blocked, increasing acquisition time and intermittent sample collection due to the delay of unblocking the cytometer. This is a problem that is not uncommon in mass cytometry and anecdotally reported across research groups as a frustrating occurrence of the technology. Blockages during cytometry acquisition is also formally acknowledged with technical reports from expert cytometrists on how to reduce this occurrence<sup>257</sup>. However, during the acquisition of this cohort, there was no obvious indication as to which batches would be susceptible to blocking the mass cytometer. The mass cytometer that was used to acquire this cohort was consistently maintained and calibrated prior to acquisition and recalibrated using normalisation beads when the machine became blocked. Although significant effort was made to minimise the effect of variation as much as possible, collection time, freezing, thawing, preparation and washing of samples, antibody staining and instrument-dependent effects can all have an impact on the data which would not be captured by normalisation beads. For the purpose of this analysis, all samples which were present in batch 1 were removed from further analysis, as this batch was deemed vastly different compared to those in batches 2-10. Removing batch 1 from this analysis did not mean that it could not be analysed at all, as batch 1 contained follow-up time points for patients with established RA which can be analysed independently. The variation observed in batches 2-10 required correcting prior to further analysis to prevent false interpretation of the data.

#### **4.2.1.5 CytoNorm algorithm for normalising antibody staining across sample batches**

Variation in staining is an unavoidable occurrence in cytometry particularly if introduced by sample collection or machine signal drift. However, as this is a phenomenon experienced by other groups, algorithms to correct for variation have been developed. Van Gassen and colleagues have acknowledged the existence of staining variability and devised the CytoNorm normalisation script which can be run through the R Studio platform<sup>258</sup>. The advantage of CytoNorm is that it is specifically developed for data obtained from mass cytometry and thus is appropriate for use on single cell data. With support from UCB Pharma, this CytoNorm script was adapted to use for this data set. CytoNorm ensures internal consistency between clinical samples based on shared controls across various batches and for this dataset, each internal batch control from batches 2-10 was used to initially train the normalisation algorithm and model the batch effects. The principle of the normalization is as follows. Initially, data from the shared controls (batch controls) is used to learn the appropriate transformations for each batch<sup>258</sup>. These technical variations can be protein marker specific where certain cell populations express a higher level of the protein resulting in several population-specific transformations to normalize cells<sup>258</sup>. Firstly 25,000 cells in each batch control sample was clustered with all 37 markers using the FlowSOM algorithm to identify cell populations.

FlowSOM is a commonly used clustering tool and was used in identifying batch variations in section 4.2.1.4 as it is fast and efficient in detecting populations of various sizes and shapes without being computationally expensive<sup>256,258,259</sup>. Here, Van Gassen and colleagues made an assumption that whilst the variation across the batches exist, the differences between the cell populations is bigger and thus FlowSOM can distinguish cell populations without being detrimentally affected by batch variation. However, a cut-off point of a coefficient variation of  $\leq 2$  was considered to be acceptable, calculated by Van Gassen et al., with anything larger than this attributed to batch effect<sup>258</sup>. Moreover, for this dataset, batch 1, where the biggest variation was identified in the cell populations was removed from this analysis which eliminated the possibility of confounding results. Once the FlowSOM algorithm run was completed, normalisation was applied per cluster and an overview of the distributions for each control sample per marker, per cluster was calculated. Once the distributions were determined for each batch control, modelling the transformation was calculated

to determine the actual distribution to the goal distribution <sup>258</sup>. As defined in the paper detailing the CytoNorm algorithm, Van Gassen et al., modelled the transformation by using splines <sup>258</sup>. A spline is a piece-wise defined function which is used to calculate the difference between given points in a dataset but still retains the structure of the data. In this instance, using splines allows for the translation of original marker values to the new marker values and thus be as close as possible to the goal distribution and results in an individual spline for each pair of clusters and markers per control sample <sup>258</sup>. This pre-calculated FlowSOM model and splines were applied to the samples by mapping the new samples onto the training samples (controls) FlowSOM clusters and splines. Each sample that has been normalised by the CytoNorm algorithm is then converted into a .fcs file to use in the analysis pipeline. Each sample that has been normalised by the CytoNorm algorithm can be subsequently validated in FlowJo to assess the outcome of normalisation by aligning each marker in each sample.



**Figure 4-8|FlowSOM analysis shows that once normalisation through the CytoNorm script has been applied, batches control staining variation becomes consistent.**

**A** Prior to the application of CytoNorm to normalise the samples from all batches, the FlowSOM clustering analysis performed using the cytofkit GUI through the R Studio platform, shows that the cell populations in batch controls 2-10 have a heterogeneous staining pattern (depicted by the colours observed which are specific for each batch) despite being from the same donor. The expectation here would be that all the cell populations would superimpose with little discrepancy between the batches but this was not observed. **B** After application of CytoNorm on the batch controls the staining pattern within the clusters across the batches are more consistent. However, batch 4 (green) still appears to not conform to the homogenous staining and therefore samples in batch 4 were analysed separately downstream of the analysis pipeline.

As batch variation was apparent across the batch controls, it was important to assess whether normalisation could improve staining consistency without having detrimental impact on the final interpretation. The batch controls were used to train the algorithm and this was then applied to the batch controls. Figure 4-8 shows the batch controls before and after normalisation using FlowSOM clustering in cytofkit to initially observe how the batch controls align in terms of cell population clusters. In section 4.2.1.4, this was already identified but what becomes apparent is when 10,000

cells from each batch control were concatenated to create one FlowSOM plot, the batch controls do not form homogeneous clusters and each batch control (denoted by colour) sits distinctly apart. Although the batches form clusters in close proximity to one another, the expectation would be that as they are all derived from the same donor, processed on the same day and simultaneously prepared that the clusters should be close to if not exactly identical. However, once the batch controls have undergone normalisation, the cell populations clustered again using FlowSOM, are now more homogeneous in staining as each control now blends in with the other controls. Batch 4 (depicted in green) however, can be observed to sit apart from the other batch controls. Due to this unresolved normalisation, batch 4 was thus removed from analysis downstream as it was decided that the inconsistencies observed even after normalisation would influence interpretation of the data and that batch 4 could be analysed separately so as to limit the affect of batch variation on data analysis.

To achieve consistency in the interpretation of the data, samples were normalised through the CytoNorm in terms of the hypothesis in question. This was decided because inconsistencies in normalisation across donors from different batches became apparent when the samples were analysed in FlowJo. This inconsistency was particularly noticeable in the markers CD11b and CD16, where across the normalised samples, alignment was still vastly different. This could be either due to donor variation but this was thought to be unlikely as the staining profile for CD11b and CD16 were consistent within the batch but inconsistent even after normalisation across the batches. Certain markers have been shown by other groups including Van Gassen et al., observed that CD15 and CD66 fluctuated across the batches but not within the batches. Whilst Van Gassen et al., were able to resolve the problem of fluctuating expression by changing the spline function from a linear to a non-linear transformation. It is worth observing here that the markers in this data set CD11b and CD16 and the markers identified CD15 and CD66 are markers that are typically used to identify myeloid cells in particular the neutrophil population. The neutrophil population of cells that showed the most variation in this dataset. As neutrophils are generally a short lived cell population in the blood, it is likely that by the time the samples were processed, some cell loss had occurred due to the nature of the cells and the logistics involved between collection of sample to processing of sample. It is also accepted that variation in cell populations including within the neutrophils will occur due to donor variation but where the CD11b and CD16 expression deviated due to obvious batch effects which could not be corrected by the CytoNorm algorithm, these samples were removed. In doing so, it was felt that the samples included for data analysis downstream would be reliable and yield data that could be trusted to inform interpretation and draw conclusions.

## **4.2.2 Validation of Diffcyt pipeline for mass cytometry data analysis**

### **4.2.2.1 Preparing the metadata and marker data frame to import into the Diffcyt pipeline**

Automated analytical pipelines are required to interrogate high dimensional cytometry data to analyse data in an efficient and replicable manner. Whilst many automated algorithms exist for analysing single cell transcriptomics data, this has not been the case for cytometry data until recently. Many of the algorithms developed for cytometry data have been informed by existing approaches in transcriptomics analysis and adapting these algorithms, or directly incorporating them for cytometry analysis has become a rapidly developing field in bioinformatics.

Whilst clustering plots such as those performed in FlowSOM or Phenograph have become a familiar presence in conveying cytometry cell populations, an approach that can integrate both unsupervised

clustering and supervised statistical analyses to detect cell populations or functional states such as activation status to associate with a specific outcome would prove invaluable for cytometry data analysis. Weber et al., have provided a solution for this problem by developing a new computational framework known as Diffcyt, which incorporates FlowSOM clustering to define cell populations, and empirical Bayes moderated tests including *edge R*, *limma* and *voom* which have been adapted from differential analyses of transcriptomics data <sup>166,176</sup>. Differential analyses testing in the Diffcyt workflow also includes alternative methods which have been adapted from the classic regression framework <sup>166</sup>.

The complexity of this dataset meant that a robust approach was required which could test several different hypotheses. Coupled with the complexity of having no prior bioinformatics experience meant that I had to decide on how best to answer the hypotheses with the Diffcyt framework whilst acknowledging time constraints. With significant help from UCB Pharma, I was able to adapt the script in R Studio to interrogate the data. Before commencing analysis using the Diffcyt pipeline, two Excel spreadsheets were prepared. The first spreadsheet details the sample file names to be read into the script pipeline and the conditions assigned to each sample e.g. health or RA. Applying the conditions to each sample allows for patterns (if any) to be observed easily without having to refer back to the identity of each individual sample. Patients and condition names do not bias the Diffcyt pipeline as the algorithms included assess staining patterns of the antibody markers and not labels included by the researcher. Each spreadsheet was designed to include samples that best answered the hypothesis in question. Although the Diffcyt algorithm should be able to handle multiple patient identifiers and conditions, this became difficult for statistical analysis so spreadsheets were simplified to include sample name and up to three conditions (e.g. health, early RA and established RA).

The second spreadsheet contains the markers which were included in the panel to be analysed and the assignment of each protein marker to 'lineage' and/or 'functional' categories. Table 4-1 shows the markers used for analysis and how the markers were characterised for analysis through Diffcyt. Lineage refers to the markers that are associated with defining specific cell populations as identified by the literature and are well established for example CD3 as a marker for T cells and CD19 as a marker for B cells. Defining a lineage in the context of the Diffcyt script also includes protein markers which are not conventionally associated with a functional state of the cell for example activation status. Often markers however, have the ability to function as both lineage markers and functional markers and the definition of a lineage and functional marker can be oversimplistic. The Diffcyt pipeline anticipates this and therefore allows the researcher to assign protein markers as lineage and functional if deemed appropriate. Markers are denoted as '0' if they are not lineage or functional and '1' if they are lineage or functional in the appropriate column. Where markers are deemed to be both lineage and functional, this is represented by a '1' in both columns and where markers are deemed as neither lineage or functional and therefore not included in the analysis, this is marked as '0' in both columns. Lineage markers are used to define clusters which represent cell populations and can be tested for differential abundance represented as percentages of the number of cells analysed and presented as a boxplot. For analysis of functional states of the cell, the median expression of the signal per cluster are used to calculate differential states within populations. The differential state can only be used where the changes in expression are not subtle however, the advantage of this analysis is that the results can be directly associated with cell types or populations of interest <sup>260</sup>.

Three markers were excluded in the Diffcyt analysis pipeline including CD123 and CD56 where marker expression was absent due to no expression of these markers being detected. CD123 was primarily included to identify dendritic cells and CD56 to identify NK cells. This did not hinder the process of identifying dendritic and NK cells as the FlowSOM clustering algorithm was able to detect

both populations using additional markers and these cell populations automatically segregate. IgD was incorrectly labelled during acquisition as '1gd' which was incompatible with the script due to incorrect notation, although expression can be analysed through traditional gating methods using software such as FlowJo. The exclusion of these markers can be observed in Table 4-1 where 0 has been placed under the column headings 'lineage' and 'functional'. The assignment of markers to 'lineage' or 'functional' was extensively optimised by multiple iterations of the FlowSOM algorithm to identify clusters and was applied to all the analysis runs in the same format each time to retain consistency.

**Table 4-1| Classification of lineage and functional protein markers for Diffcyt analysis**

0 denotes the marker is not classified as lineage and/or functional and 1 denotes that the marker is either lineage and/or functional. Protein markers which are canonical lineage markers as established by literature were assigned 1 in lineage and 0 in the functional category. Lineage markers included: CD11b, CD161, CD127, TCRy $\delta$ , CD11c, CD3, CD203c, CD14, FCeRI, CD19, va7.2, CD4, CD8a, CD16, CD138, CD25, CD45RA, CD45, and CD45RO. Functional protein markers were assigned 1 and 0 for lineage included: Granzyme B, NKp44, CD28, CTLA4, Perforin, TNFR2, PDL1 and CD80. Protein markers which were determined as both lineage and functional were assigned 1 to each column included: CD40, CD27, CD86, NKp46, CD38 and HLA-DR. These assigned labels were optimised prior to the final analysis.

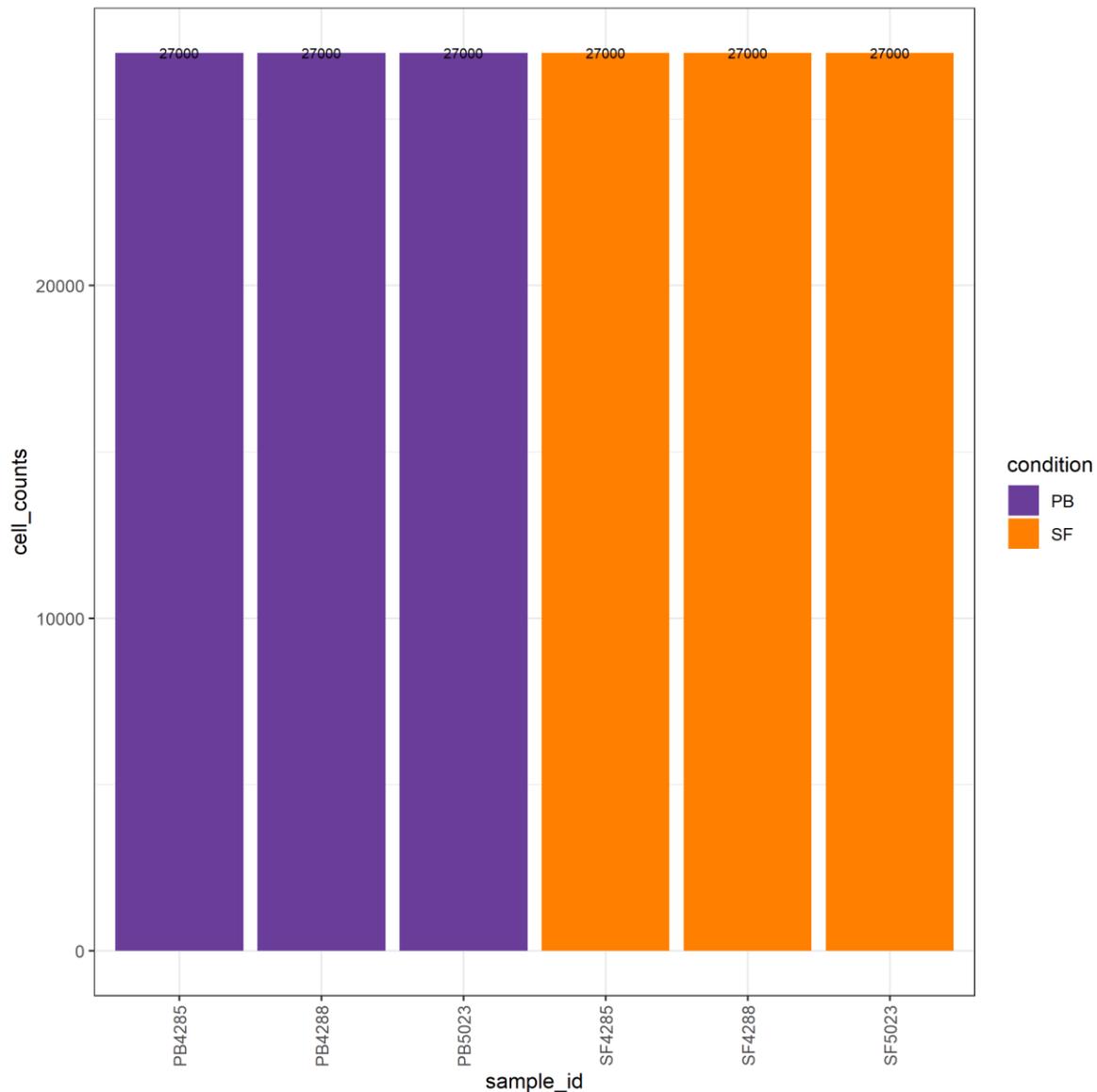
CD56, CD123 and IgD were not included in the phenotypical analysis.

Antigen	Lineage	Functional
CD11b	1	0
CD161	1	0
CD40	1	1
CD127	1	0
TCRy $\delta$	1	0
CD11c	1	0
CD27	1	1
CD86	1	1
CD3	1	0
CD123	0	0
Granzyme B	0	1
NKp44	0	1
CD203c	1	0
CD28	0	1
CD14	1	0
FceRI	1	0
CTLA4	0	1
CD19	1	0
va7.2	1	0
NKp46	1	1
CD4	1	0
CD8a	1	0
CD16	1	0
Perforin	0	1
TNFR2	0	1
CD138	1	0
CD25	1	0
PD-1	1	1

IgD	0	0
PDL1	0	1
CD45RA	1	0
CD45	1	0
CD45RO	1	0
CD38	1	1
CD56	0	0
HLA-DR	1	1
CD80	0	1

#### 4.2.2.2 Equal number of events from each sample were analysed by downsampling

Diffcyt is suited for complex experimental designs but both computing capacity and time placed a restraint on the analysis and it was decided that samples should be analysed according to which hypothesis they were best suited to answer therefore reducing the number of samples in each analysis. Run time took between 6-8 hours per day, with the most time attributed to FlowSOM clustering. Once the metadata and marker files had been read into the script, samples were downsampled equally to 27,000 events to accommodate run time and ensure that equal number of events were being analysed from each sample. Downsampling was done through R studio for each analysis run and new .fcs files were written out for subsequent analysis. 27,000 events were selected as this was the lowest number of events identified in one sample. Extensive optimisation was undertaken to ensure that 27,000 events was enough to capture any significant changes within the immune cell populations that may be taking place. Increasing the number of events to 86,000 (data not shown) did not affect the number of clusters as these were already set within FlowSOM. Another advantage of high resolution clustering is rare populations are easily identified and unlikely to be merged with larger clusters even with a low number of events. Once samples have been downsampled and new .fcs files produced, a bar graph is produced to represent and confirm the amount of cells included from each sample (Figure 4-9). This is useful especially to ascertain that an equal amount of cells have been taken from each sample and confirming the samples have been downsampled successfully. It was observed that downsampling did not negatively impact the final conclusion drawn in this investigation.

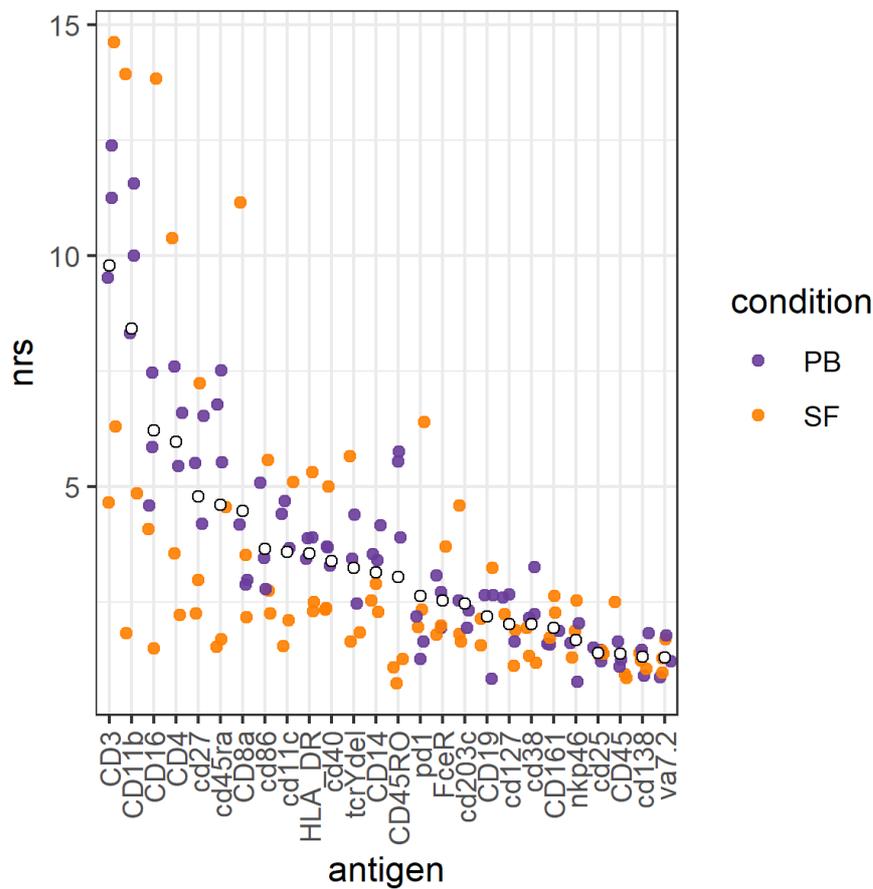


**Figure 4-9 | Bar graph shows 27,000 cells from each sample (3 synovial fluid (SF) and 3 peripheral blood (PB)) included in the Diffcyt analysis after downsampling in R Studio.**

#### **4.2.2.3 Marker expression intensities and expression levels can be assessed across all samples**

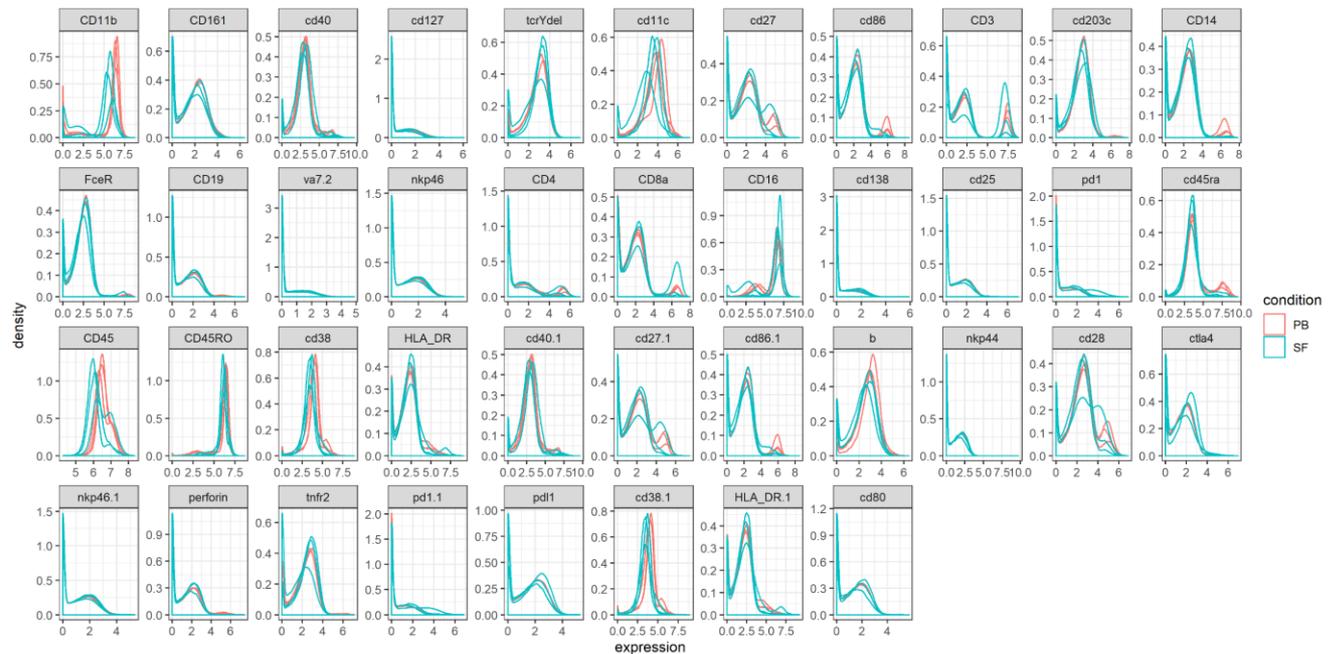
Initial analysis through the Diffcyt script looks at expression levels of markers providing an opportunity for the researcher an opportunity to check staining expression is as anticipated. Observing expression profiles across the samples can help to identify any anomalous staining that is contrary to expectation. In this case, analysis for this dataset was stringent with batch effects already addressed and therefore only samples that gave complete confidence were included. Individual markers can be analysed in each sample and the variance can be observed by using a non-redundancy score (NRS) principal component analysis (Figure 4-10). This can indicate the amount of diversity between samples with the most diverse marker in analysis of synovial fluid and peripheral whole blood as CD3 (Figure 4-10). The NRS plot shows markers in descending order of variability on the x-axis with va7.2 identified as the least variable marker in Figure 4-10.

Another check of marker expression globally across the samples is also performed using a smooth densities plot (Figure 4-11) which presents each marker as a ‘smoothed’ histogram plot with each sample overlaid to observe the profiles for lineage and functional markers. The samples are coloured according to the conditions they represent which in Figure 4-11 is peripheral whole blood and synovial fluid. Density refers to the visualisation of the distribution of the marker expression and can be useful if batch effects or unexpected staining have not been corrected for prior to analysis. These will stand out and affect downstream analysis where it becomes computationally expensive. Whilst the script provided an opportunity to double check the marker expression, this was performed extensively (as shown throughout this chapter) to ensure marker expression was optimum and corresponded with existing biological understanding.



**Figure 4-10 | Non-redundancy score (nrs) plot shows the antigens defined as ‘lineage’ markers for each sample included in the analysis.**

The NRS calculates the variability of markers in each sample. The higher the nrs for a marker, the higher the variability with the CD3 marker showing the highest variability from analysis comparing 3 paired peripheral blood and synovial fluid samples. The nrs for synovial fluid is more variable compared to peripheral blood suggesting that heterogeneity of RA as a disease is better detected nearer to the site of pathology. The solid coloured circles represent peripheral blood (PB) and synovial fluid (SF) scores on the non-redundancy scale and the white circles indicate the mean nrs from all the samples. The markers on the X- axis were used to inform FlowSOM clustering to identify cell populations and are arranged in order of highest variability to lowest.



**Figure 4-11 | Smooth densities of marker expression for each sample is calculated for lineage and functional markers**

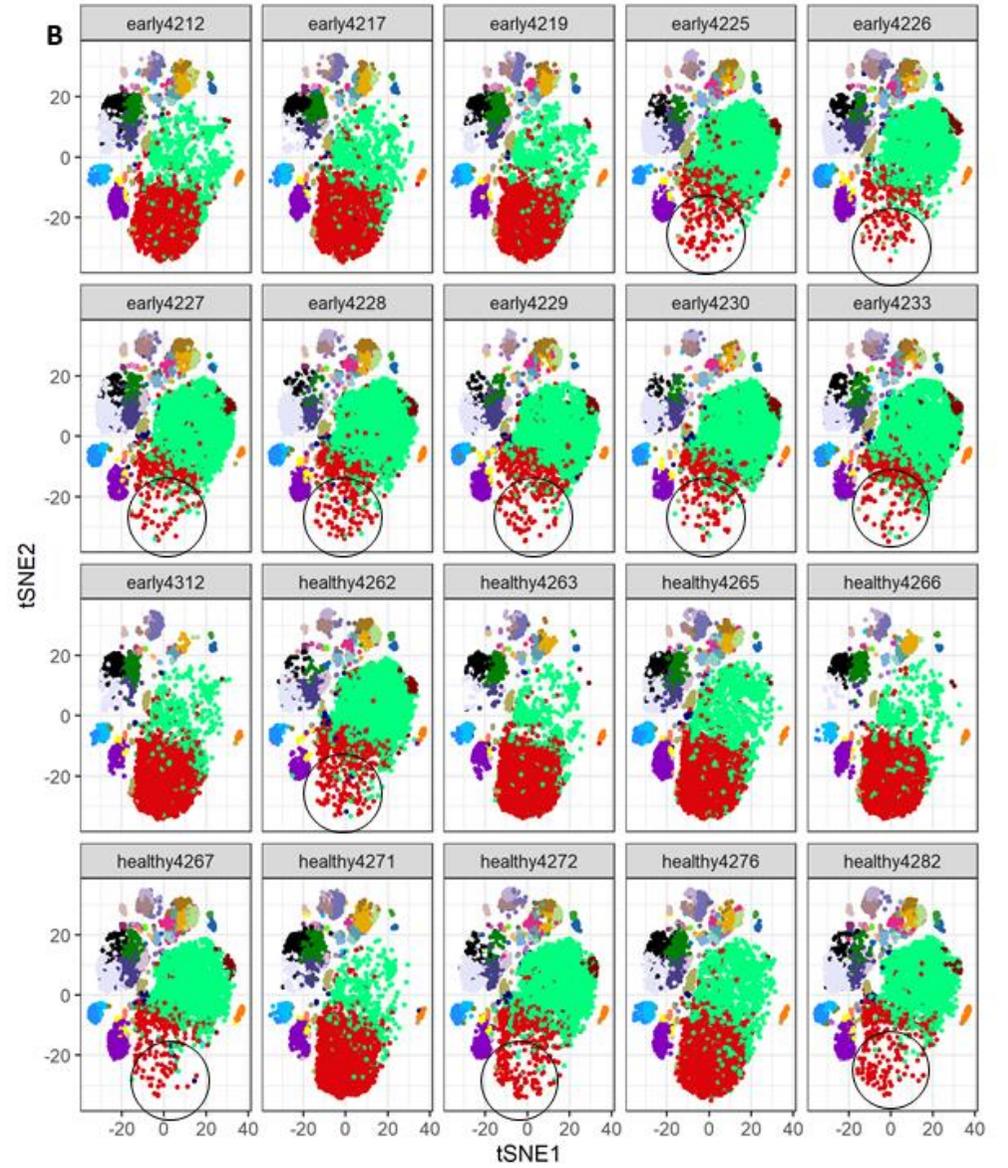
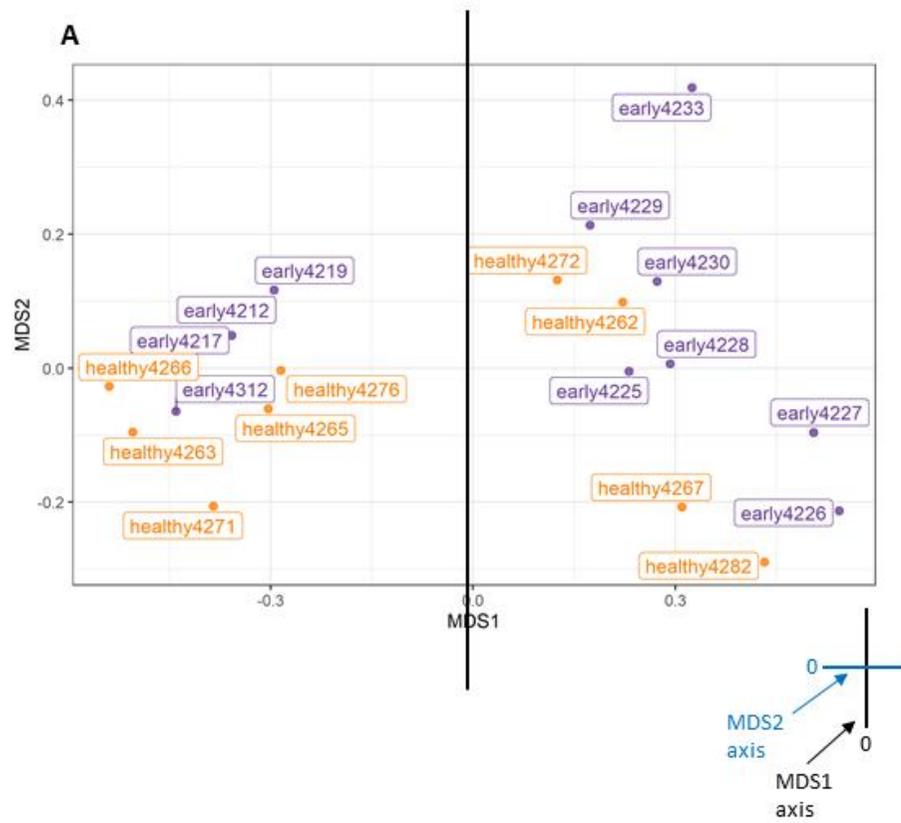
Lineage markers and functional markers were analysed in 3 paired peripheral blood and synovial fluid samples to assess the distribution of expression. Where lineage markers were also classified as functional markers, these have been labelled with .1. 'b' refers to Granzyme B.

#### 4.2.2.4 Multi-dimensional scaling plot measures similarities in marker differential expression across samples and between conditions

Multi-dimensional scaling (MDS) plot analysis is similar to principal component analysis commonly used in transcriptomics, and allows for unsupervised analysis which assesses similarities in differential marker expression between samples and conditions to be explored before conducting formal statistical analysis. Samples can cluster well according to condition provided that there is enough difference in marker differential expression between the conditions and this is plotted in the first dimension (MDS1). If samples are similar, they will be in close proximity in terms of distance on the plot and samples which are not similar will be further apart. The second dimension, MDS2 looks at the difference between patients which can be assessed within a condition as well as between conditions.

When initially optimising the Diffcyt pipeline, MDS analysis was applied to a small sample size. 11 patients with early RA and 8 healthy donors were selected across different batches and the expectation was that early RA and health would separate into two distinct groups. On the MDS1 axis five healthy donors and 4 patients with early RA cluster to the left and 7 patients with early RA and 4 healthy donors cluster to the right (Figure 4-12 A). Further analysis by FlowSOM clustering confirmed that batch effects existed within the neutrophil population (coloured green and red) where the red cluster varies across the individual samples. However, a distinct batch effect pattern can be observed within this cell population (Figure 4-12 B, red cluster), and thus driving the distinction between the two conditions. Samples which showed a prominent cell population (red cluster) were grouped to the left of the MDS1 axis and where samples had a reduced cell population

(red clusters highlighted by circles) were grouped to the right of the MDS1 axis. Whilst all the samples had been normalised using the CytoNorm algorithm, initially the algorithm did not work within the neutrophil population. The splines in the CytoNorm algorithm, split the neutrophil population into two populations: CD16<sup>lo</sup> CD11b<sup>lo</sup> and a CD16<sup>high</sup> and CD11b<sup>high</sup>. The neutrophil population however, should be a single population but discrepancies in CD16 and CD11b marker staining shown by FlowSOM clustering demonstrates how this can bias data analysis. Therefore, the MDS plot was a useful visualisation plot when examining the relationship between conditions and determining whether this is an artefact of staining or to pursue further statistical analysis.



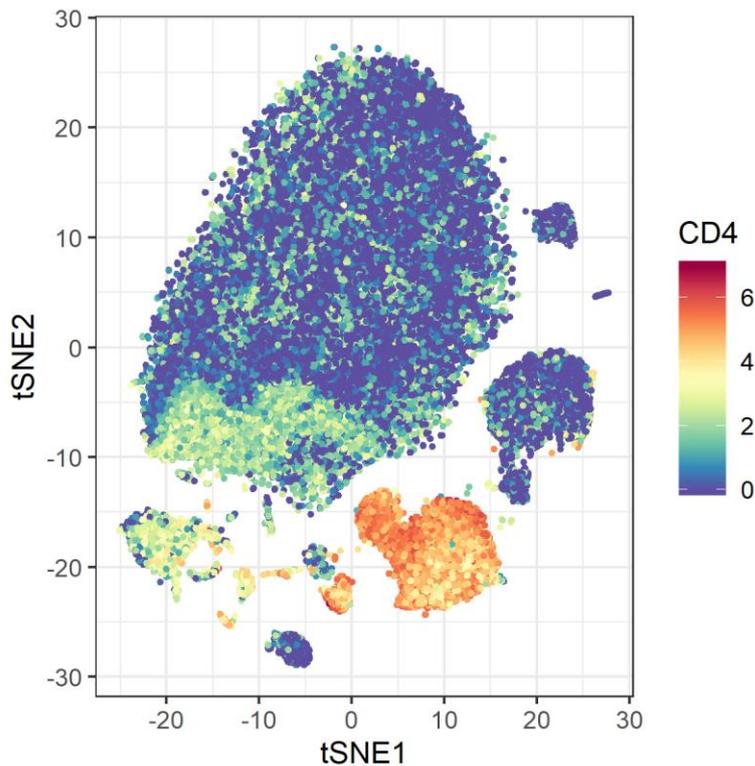
#### **Figure 4-12 | Optimising the Diffcyt pipeline identified batch effects across the neutrophil population when analysing health and early RA**

**A** Multi-dimensional scaling (MDS) analysis did not show health and early RA form two distinct groups when assessing similarities across median marker expressions. The MDS1 dimension calculates the similarity between conditions but this was not evident. **B** FlowSOM clustering performed within the Diffcyt workflow identified the neutrophil population (depicted by green and red colouring) had inconsistent staining within the population which influenced the MDS plot. The neutrophil population was split into two populations during the normalisation step which was not initially identified and this could be detected across the batches: batches 3-10 and demonstrates the benefit of marker expression analysis which is incorporated into the workflow.

#### 4.2.2.5 FlowSOM clustering and identification of immune cell populations for statistical differential testing

In section 4.2.1.4, clustering algorithms Phenograph and FlowSOM using the Cytokit GUI Shinyapp were used to assess immune cell populations in the batch controls included in each of the barcoded batches in which batch effects were identified. Once batch effects were normalised by applying the CytoNorm algorithm, including those described in section 4.2.2.4, cell populations were defined by using the FlowSOM clustering algorithm and the *Consensusclusterplus* packages incorporated in the Diffcyt workflow as these were considered to have optimum performance when clustering high dimensional cytometry data<sup>259</sup>. Interestingly, when experimenting with different clustering algorithms and options, FlowSOM was faster using the Cytobank cloud software compared to using the in built FlowSOM algorithm built into the Diffcyt script or in the Cytokit GUI. However, within the Diffcyt script, I tested the FlowSOM algorithm by repeatedly running the clustering with different number of cells, and changing the seed to ascertain the robustness and replicability of the populations identified by FlowSOM analysis. Seed refers to the starting point of the analysis of all the cells to be analysed in a dataset. When the seed is set randomly, it means that each run begins at a different starting point of the dataset. A consistent seed was set at 1234, as already defined in the script and applied to each run to limit any inconsistencies and maintain reproducibility. FlowSOM can be run on a heterogeneous population of cells such as the samples used in this data and therefore individual cell subsets do not need to be subsetted prior to running in FlowSOM. To confirm this, analysis was performed on two subsetted populations of cells: CD3+ T cells and CD11b+ CD16+ neutrophils to determine whether an increased number of cells from one cell population gave more insight into that specific cell type or whether the results were comparable to a heterogeneous population of cells. Encouragingly, no difference could be observed in differential abundance of T cells or neutrophils in a heterogeneous population of cells or a single subsetted population of cells (data not shown).

The FlowSOM algorithm performs clustering by using three main steps. The first step is to build a self-organizing map where cells are assigned according to similarities based on markers and features to 100 grid points. Based on where cells are assigned to within this grid, this forms the basis of a self-organising map. From this initial grid based self organising map, a minimum spanning tree is created which is similar in appearance to a dendrogram. This minimal spanning tree clusters groups of cells resulting in metaclusters which are represented on a 2D FlowSOM plot on tSNE axes. The analysis clusters all cells from all samples which provides a consistent approach and in principle means that there is no need to downsample events from each sample. To ensure that the FlowSOM clustering has worked, individual markers can be checked to determine whether expression is as expected as with an example shown in Figure 4-13 for CD4 marker expression.



**Figure 4-13 | tSNE plot shows the expression level of the CD4 marker.**

To assess specificity of the FlowSOM clustering, each marker can be individually assessed to observe the staining expression in the clusters which are formed by pooling all the samples in the Diffcyt analysis and randomly selecting 1000 cells from each sample (n=6). The orange-red cluster shows cells staining positive for CD4 as indicated by the colour scale where negative expression is 0 (blue) and positive expression ranges from 4-6 depending on intensity.

FlowSOM is designed so that it over clusters. 40 clusters allowed for the identification of the main cell populations by observing median expression of all the markers categorised as 'lineage'. Markers that have informed the cluster are shown by a colour scale where orange-red shows medium to high median expression intensity and blue/yellow show that those markers are not used to define that cluster. The FlowSOM algorithm has been deliberately designed so that the researcher takes the time to manually go through the heatmap and define and label the clusters. This provides an opportunity to double check that populations identified are biologically meaningful and not an artefact of the algorithm. The most time of running the script is spent here and for each dataset the cluster numbers change so need to be checked and relabelled for consistency particularly when making inferences from different analysis runs.

In Figure 4-14, the heatmap from an analysis on peripheral blood is shown. The heatmap shows the 40 clusters as defined by FlowSOM and after I analysed these clusters manually, cluster merging revealed 17 populations of immune cells. Percentage of cells for each cluster are shown in brackets next to the cluster number on the heatmap and as before, the colour scale measures the intensity of the marker expression: negative 0 (blue) through to positive red (8). The median expression is also quantified in each square.

The dendrogram on the left of the heatmap helps to identify the hierarchy of similarity amongst the clusters and can assist in deciding which clusters are similar. For example, cluster 2 clearly has the largest abundance of cells (59.27%) and when observing the markers, they are characteristic of the neutrophil population: CD45, CD11b, CD16, CD38, CD11c and CD45RO. On further inspection,

clusters 1, 12 and 31 are very similar to cluster 2 but weakly express CD11c which the algorithm clustered separately. I chose to group these four clusters together and Figure 4-15 shows the heatmap of the immune cell populations after the clusters have been merged and their population percentage. One population called 'cells' was unable to be verified and was excluded from statistical analysis. The annotated populations can be visualised in a tSNE plot with cluster size proportional to the percentage of cells in each population. The markers used to define each population are listed in Table 4-2.

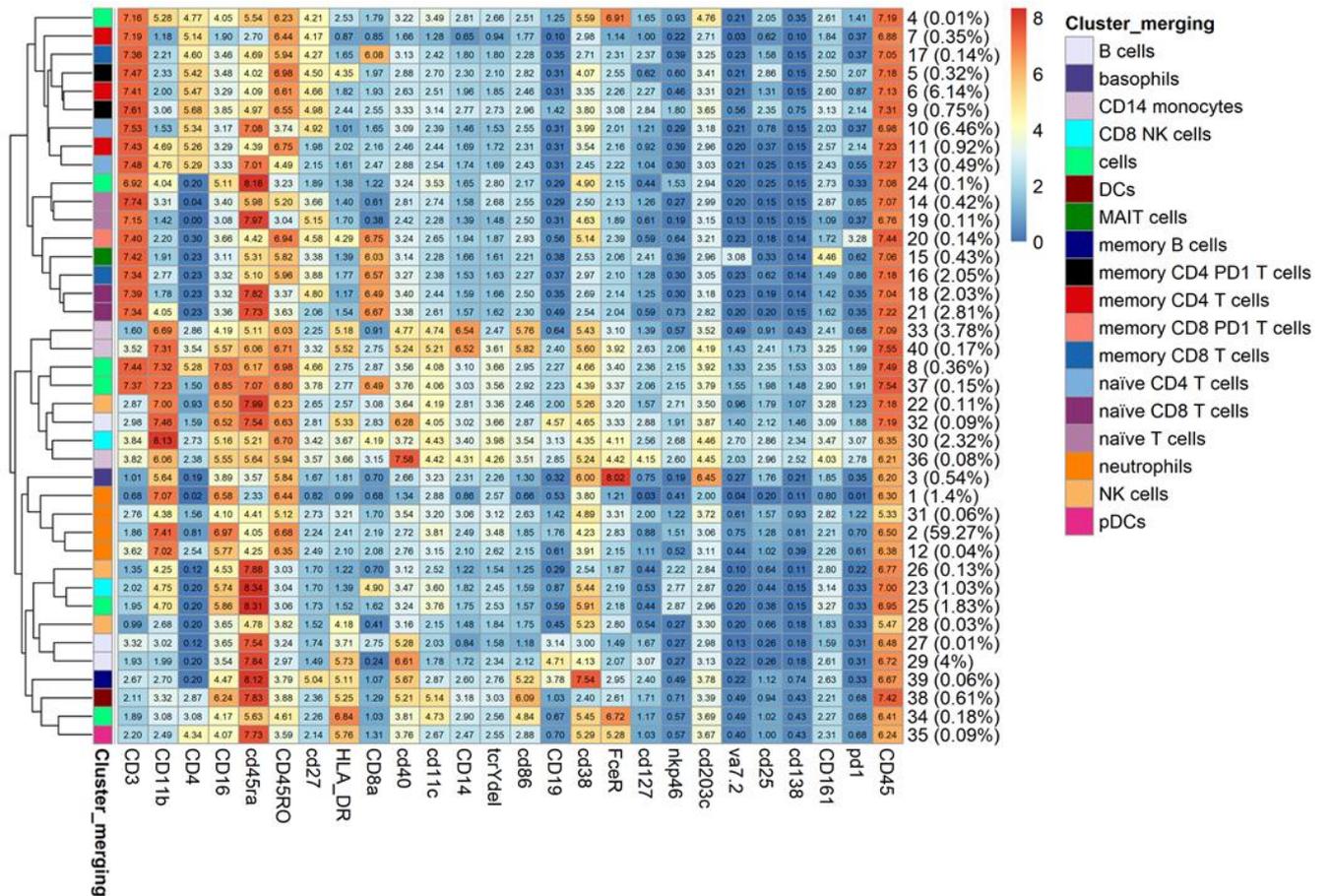
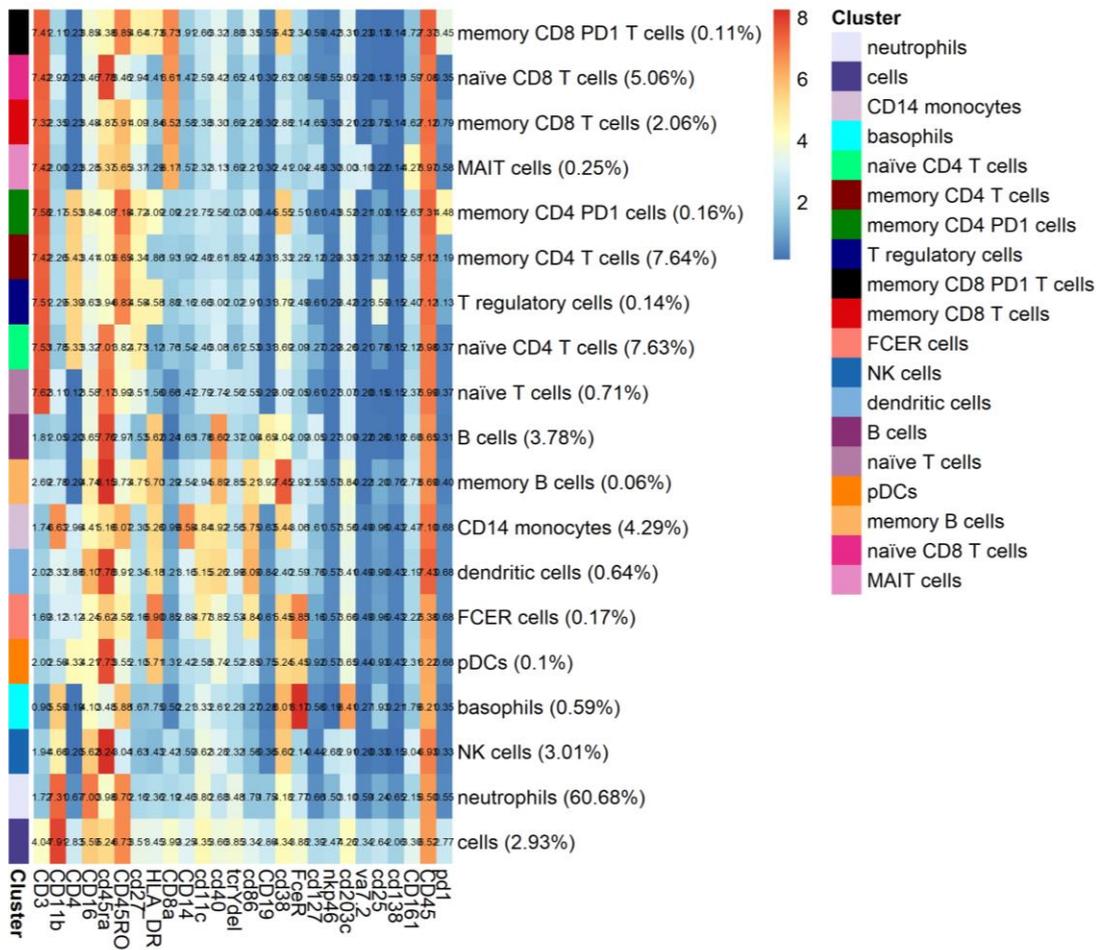


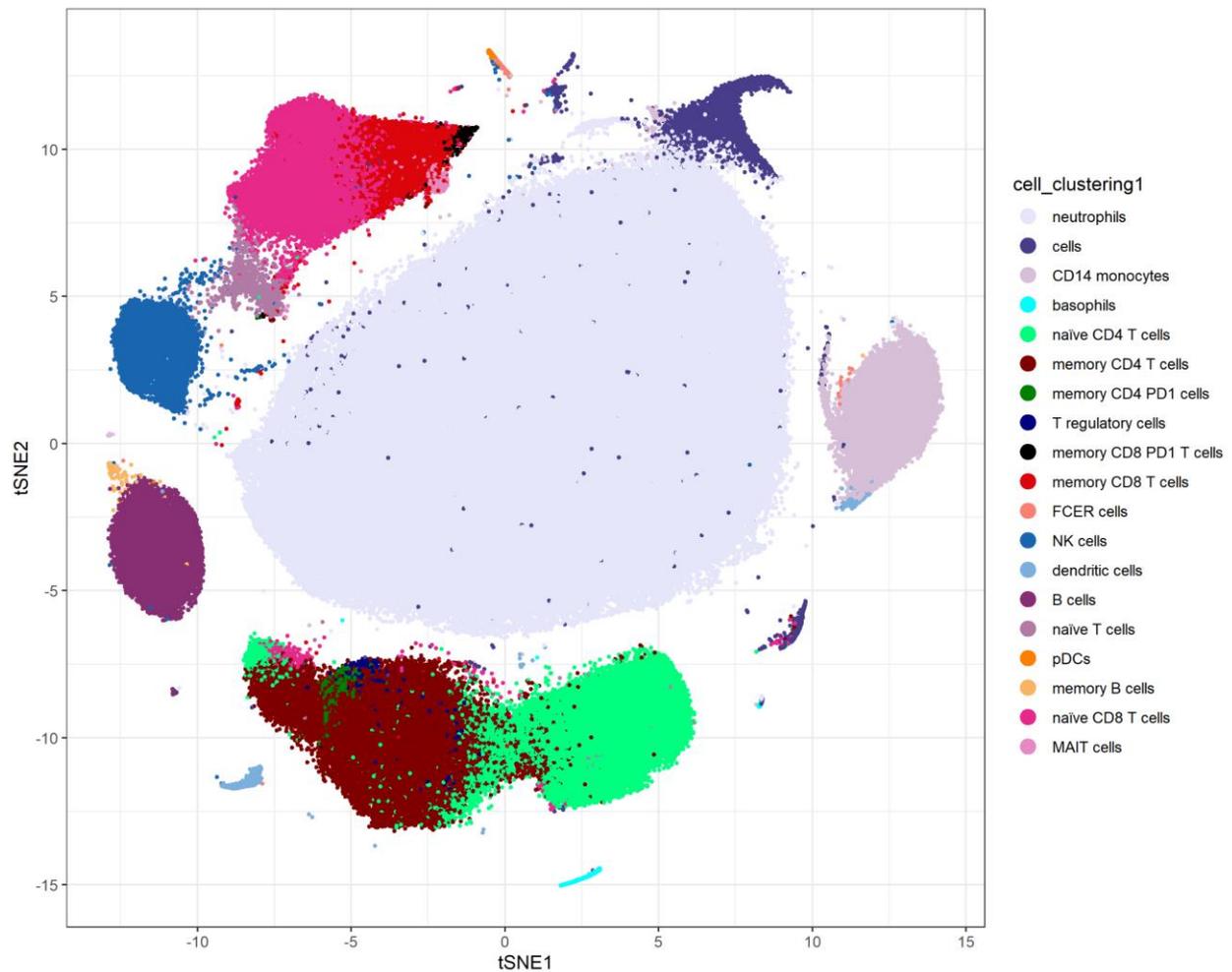
Figure 4-14 | Heatmap shows 40 clusters identified by FlowSOM and the merging of these clusters into defined cell populations.

The FlowSOM algorithm is deliberately designed to be semi supervised during the clustering process. This means that whilst the clusters are computationally identified, it is for the researcher to define the immune cell populations. FlowSOM is designed to over cluster, and 40 clusters are presented with the percentage of cells that form that cluster in brackets. The number of clusters were estimated based on previous iterations of FlowSOM however, it is recommended with the FlowSOM algorithm to over estimate the number of clusters and afterwards combine clusters that are phenotypically similar, together. Based on median expression intensity for lineage markers, coded by a colour scale blue = 0 (no expression) to red = 8 (high expression) and quantified in cell of the heatmap, clusters can be merged. The hierarchy of similarity between the clusters is represented on the left of the heatmap by the dendrogram. 17 populations in peripheral whole blood were identified in this clustering analysis. One population labelled as 'cells' was unidentifiable and as a result was excluded from further analysis.



**Figure 4-15 | Heatmap of immune cell populations identified in peripheral whole blood after cluster merging**

Cluster merging analysis was performed by using the heatmap produced in Figure 4-14. Briefly, the heatmap in Figure 4-14 showed the intensities of each marker within a particular cluster. Clusters were merged where marker expression was deemed similar across all the markers and referring to the tSNE plot to identify the location of these clusters. Merging the clusters provides a more accurate insight into the number of cell populations that are actually present within the biological sample without including the same populations multiple times due to slight changes in the marker expression. The immune cell populations identified in peripheral whole blood are shown with the percentage of cells in each cluster shown in brackets median expression intensity for lineage markers, coded by a colour scale blue = 0 (no expression) to red = 8 (high expression). The population labelled ‘cells’ was a non-specific population of cells which could not be identified and hence was not analysed further.



**Figure 4-16| Immune cell populations, after clustering analysis can be visualised in a FlowSOM tSNE plot**

The FlowSOM plot shows the immune cell populations denoted by the colours on the right hand side of the image. The largest cell population is neutrophils. The visual size of the populations correlate with the amount of cells that make up the population.

**Table 4-2| Immune cell markers used to define cell populations in FlowSOM**

Each cell population and the protein markers used to define the individual populations were informed by the preceding heatmap (using the percentages of marker intensity) as shown in Figure 4-15 and the tSNE plot shown in Figure 4-16.

Cell population name	Protein markers used to define the population
Basophils	CD45/CD11b/CD16/CD45RO/CD38/FCεRI/CD203c
B cells	CD45/CD19/CD45RA/HLA-DR/CD40
CD8 <sup>+</sup> Natural killer cells	CD45/CD3/CD8/CD11b/CD16/CD45RA/CD38
Dendritic cells	CD45/CD16/CD11c/CD45RA/HLA-DR/CD40/CD86
FCεRI <sup>+</sup> cells	CD45/CD45RA/CD11c/HLA-DR/CD86/CD38/FCεRI

MAIT cells	CD45/CD45RA/CD45RO/CD8/Va7.2/CD161
Memory B cells	CD45/CD19/CD16/CD45RA/CD27/HLA-DR/CD40/CD86/CD38
Memory CD4 <sup>+</sup> T cells	CD45/CD3/CD4/CD45RO/CD27
Memory CD8 <sup>+</sup> T cells	CD45/CD3/CD8/CD45RO/CD27
Memory CD4 <sup>+</sup> PD-1 <sup>+</sup> T cells	CD45/CD3/CD4/CD45RO/CD27/HLA-DR/CD38/PD-1
Memory CD8 <sup>+</sup> PD-1 <sup>+</sup> T cells	CD45/CD3/CD8/CD45RO/CD27/HLA-DR/CD38/PD-1
Monocytes	CD45/CD14/CD11b/CD16/CD45RA/CD45RO/CD11c/HLA-DR/CD40/CD86/CD38
Naïve T cells	CD45/CD3/CD45RA/CD27/CD38
Naïve CD4 <sup>+</sup> T cells	CD45/CD3/CD45RA/CD27/CD38
Naïve CD8 <sup>+</sup> T cells	CD45/CD3/CD45RA/CD27/CD38
Neutrophils	CD45/CD11b/CD16/CD45RO/CD11c/CD38
Natural killer cells	CD45/CD11b/CD16/CD45RO/CD11c/CD38/NKp46
Plasmacytoid dendritic cells	CD45/CD4/CD45RA/CD45RO/HLA-DR/CD40/CD38/FCeRI
T cells	CD45/CD3/CD45RA/CD45RO
T regulatory cells	CD45/CD3/CD4/CD25/CD45RA/CD45RO/CD27/HLA-DR

#### 4.2.2.6 Differential analysis and hypothesis testing

Before differential testing is performed, the MDS analysis (discussed in 4.2.2.4) can initially indicate whether significant differences between sample conditions will be identified. If the samples show a strong separation based on condition, this should give a gain in power to detect differences during differential analysis<sup>176</sup>.

Diffcyt is a new computational framework that allows for high-resolution unsupervised clustering together with supervised statistical analyses to detect cell populations or states associated with an outcome variable in high-dimensional cytometry data<sup>176</sup>. At its simplest, the null hypothesis for testing is that there is no difference in immune cell population frequency and immune cell function between the grouping of samples e.g. health compared to RA. To summarise this, essentially the null hypothesis is that there are no immunological differences (or at least within the scope of the protein markers selected in this investigation). Therefore the diffcyt methodology uses FlowSOM clustering to define cell populations, and empirical Bayes moderated tests which are adapted from transcriptomics analytical tools for differential analyses<sup>176</sup>. The model within Diffcyt allows for cytometry-measured features which are cell population abundances and median expression of cell state markers within the user defined cell populations. These two features are referred to as response variables and enable analysis of complex experimental designs typically attributed to

patient cohort studies. These variables include batch effects, paired designs and continuous covariates with linear contrasts enabling testing of a wide range of hypotheses<sup>176</sup>. Furthermore, the model considers batch effects and continuous covariates as fixed effects. Discovery of rare populations would be uncompromised as the framework includes a high-resolution clustering algorithm that will ensure that the rare populations do not merge within larger lineage clusters<sup>176</sup>.

For testing, the Diffcyt workflow used mixed models methodology using the general linear hypotheses function for arbitrary hypotheses using t-tests<sup>176</sup>. The p-value shows the probability that under the null hypothesis, the p-value calculated indicates the probability of observing as great, or greater difference between the two conditions<sup>176</sup>.

Differential testing is performed in two parts in Diffcyt. Differential abundance (DA) looks at each cluster and marker independently and based on this analysis, the proportion (percentage) of each immune cell population (as defined by cluster merging) is plotted on a boxplot, where condition is shown on the x-axis and percentage on the y-axis. The DA testing calculates the raw p-values and adjusts these p-values for each cell population. Significant populations are then presented in a heatmap. Differential expression (DE) looks at the functional markers, as identified at the beginning of the workflow, in each of the immune cell populations and again this is presented as a boxplot with functional markers on the x-axis and median expression on the y-axis. As with DA testing, significance in DE of functional markers within an immune cell population is presented in a heatmap. For both DA and DE boxplots, the y-axes are scaled to accommodate for the range of data plotted and the distribution of donors within a condition. To correct for multiple testing, the Benjamini-Hochberg adjustment was applied to each analysis and a false discovery rate was set at 5%<sup>176</sup>.

### **4.3 Discussion**

The aim of this chapter was to describe the methodology for analysing mass cytometry data generated in this work. Analysing cytometry data has increasingly relied upon automated pipelines to improve the reliability of data, expedite analysis and ultimately attain a level of replicability and consistency across research groups<sup>237,238</sup>.

Bioinformatics is a complex field and can be difficult to navigate. A lot of time is required to develop basic scripting skills and become confident in troubleshooting and repairing the script when it 'breaks' – a term used when the script has errors and further analysis cannot be conducted until the code is corrected. It is common practice in many research departments to have access to a bioinformatician/bioinformatics department which allows for data to be processed and returned to the researcher for further statistical analysis. Whilst a trained bioinformatician is indispensable, basic coding skills and ability to be able run scripts through the R Studio platform is becoming an important skill to develop. There are many advantages to becoming familiar with basic bioinformatics, chiefly allowing the researcher a certain autonomy in data analysis and not being reliant upon often inundated bioinformatics departments. However, the time required to familiarise with the algorithms described in this chapter should not be overlooked but it is hoped that researchers either looking for a prescribed methodology or direction for cytometry analysis can find some guidance from this chapter.

#### **4.3.1 Strengths and limitations of the Diffcyt workflow**

The Diffcyt workflow provided a unique opportunity to be able to visualise clustering of cell populations across different patient cohorts and also provide statistical quantification of cell

populations that were either decreased or increased between conditions and measure differential expression of functional markers stratified by immune cell populations. The Diffcyt workflow has incorporated many staining checks and the MDS plot in particular was useful in terms of identifying initial issues within the neutrophil population which influenced the grouping of samples. The 17 populations identified by FlowSOM were consistently observed in multiple FlowSOM runs which as far as this dataset was concerned, provided confidence in the replicability of this data. To maintain replicability, the same seed number was used so that the clusters were not vastly different each time. Diffcyt is not specific for any one type of analysis making it an adaptive pipeline for a variety of different clinical cohorts and hypotheses to be analysed. This dataset makes full use of this opportunity which will be demonstrated in the next two chapters and it is this flexibility that primarily influenced the use of this workflow.

The use of Diffcyt on data analysis from mass cytometry is not extensively reported as yet due to the pipeline being fully published in 2020<sup>166</sup>. It is anticipated that this will change as research groups become more familiar with the script and its availability. The researchers who created the Diffcyt workflow were able to test the script on a clinical cohort of peripheral blood obtained from patients with melanoma treated with anti-PD-1 therapy identified a small subpopulation of monocytes, were detected at baseline and able to predict response status<sup>166</sup>. This population of monocytes were at a low frequency and regarded as a rare population and was still able to be identified in the presence of batch effects, a strength of the Diffcyt pipeline as discussed by the authors<sup>166</sup>. Separately, the principle of Diffcyt is based on the same statistical approach reported by another research group working on mass cytometry and RA<sup>261</sup>. In the paper, Fonseka et al., approach data analysis using the 'Mixed effect modelling of Associations of Single cells' which like Diffcyt uses 'reverse association strategy' where it looks at association between cell population clusters and disease status at the single cell level<sup>261</sup>. The fact that two different approaches adopt the same statistical testing, is encouraging and data can be compared to see if similar results are obtained if a similar disease and immunophenotyping panel is being studied. Fortuitously, Fonseka et al., looked at peripheral blood from patients with RA and although the experimental approach is different, there is an overlap in the immunophenotyping panel and certain populations identified by Fonseka et al., have also been detected in this cohort, which will be discussed further in Chapter 5.

The clustering step in Diffcyt is a challenging part of the script. It is the most important part of the whole workflow as it informs the differential statistics. Whilst the clustering process is automated to produce the heatmap it is up to the individual to define the clusters. FlowSOM can cluster populations that are the same but may have subtle marker expression changes either due to different conditions, donor variability and/or batch effects. Marker expression variability as a result of batch effects were stringently addressed in addition to removing IgD, CD56 and CD123 from the final analysis due to technical error. This in itself should increase confidence in the dataset and the analysis. To identify clusters, a consistent approach was adopted, where 17 cell populations were identified each time FlowSOM clustering was applied. To further reduce the variability in clusters, 27,000 cells were analysed from each sample. This was selected based on the lowest amount of cells present in a sample and downsampling was performed on all samples. The merits of downsampling are mixed. Those who do not favour raise a valid point that by downsampling, rare populations of cells may be under represented or overlooked and incorporated into a larger (often more established) cell population. Another problem is that whilst automated clustering is robust and able to corroborate cell populations identified by manual gating approaches, clustering algorithms often do not hold up well across samples. For the Diffcyt workflow, cells from all samples are combined before clustering which does not bias the analysis further on. However, should there be a vast

discrepancy in the number of events analysed from each sample, this could bias the clustering as a sample containing more events would be disproportionately represented compared to a sample containing fewer events. Moreover whilst FlowSOM is generally thought to be fast, this should be put into context. To cluster 999,000 events from 37 samples, took just under 6.5 hours to run. Typically bioinformatics might require the use of higher computing power than the one used in this analysis but this is a true representation of analysis time and logistical situations meaning that higher computing power cannot be accessed in the given period of the research.

With the above considerations, it is no surprise that new clustering algorithms continue to be developed. The development of the partition-assisted clustering and multiple alignments of networks is capable of capturing immune cell populations that would be identified manually and aligning populations across samples to account for conditions <sup>262</sup>. Another clustering algorithm, QFMatch offers another solution to the problem of sample alignment in which it performs cluster matching in multiple samples post clustering <sup>263</sup>. QFMatch also locates populations which are present in some samples and absent in others and measures the dissimilarity score between cluster pairs <sup>263</sup>.

Thus in my opinion the Diffcyt workflow has offered an unparalleled opportunity to analyse a large clinical cohort with an approach that acknowledges the high-dimensionality of the data. The Diffcyt workflow may not be appropriate for all cytometry analyses but it is ideal for discovery analysis which is the premise of this study. It has allowed for an unbiased approach to the data analysis and not driven by the preference of the researcher but by what is actually detectable within the data. The next two chapters will discuss the findings of cell populations in RA and the implications these populations have. Furthermore, the cell populations identified will be discussed in the context of findings shown by other groups using mass cytometry in particular in RA.

### **4.3.2 Data pre-processing provides confidence for downstream differential analysis**

The best practice approach to guide data analysis was adopted <sup>264</sup>. Automated gating using Gaussian parameters provided and subsequent application of compensation meant that a high quality single cell population was obtained. Considerable time was spent analysing the quality of the staining across the barcoded batches and despite best efforts to reduce and preferably avoid batch effects, these unfortunately were identified. The batch effects identified in the batch controls were an example of careful experimental design in this study which are often missing from other mass cytometry datasets <sup>265</sup>. However, by identifying these early on in the data clean up and pre-processing phase, it gave greater confidence in the final data set for analysis. In this case, had controls not been included in the experimental design, it would have become a lot harder to identify staining variability. CytoNorm was a useful algorithm and able to correct for the staining variabilities although it was decided after the normalisation, that batches 1 and 4 would be removed from further grouped analysis as the staining still did not align with the remaining batches. Furthermore, batches 1 and 4 can be included in future analysis as these datasets contain patients where follow up time points were collected. It should also be mentioned that when designing the experiment, considerable thought was given to how best group the samples for barcoding. Barcoding has the limitation of only 20 samples to be included and this was limited to 19 samples as a batch control was included. Thus it was decided to group follow up time points for the same patients in one batch so that these could be analysed independently if necessary. However, for the purpose of consistency in this data analysis approach, these two batches were excluded as it was preferred to look at many

samples as possible across the batches. It was observed that batch effects can affect certain populations more than other populations and this was particularly apparent in the granulocyte population, specifically the neutrophil population. The markers CD11b and CD16 were particularly influenced by staining inconsistencies and correcting these was important to achieve a consistent neutrophil population across the batches and within samples. This is an acknowledged limitation of CytoNorm where an assumption is made that the batch effects are small enough that they do not impact the clustering<sup>258</sup>. When large batch effects do occur, biologically similar clusters across samples may become split as observed with the neutrophil population and as a result do not align, providing an opportunity for future normalisation pipelines to optimise the clustering step of the algorithm.

# Chapter 5 Immunophenotyping whole blood across the stages of Rheumatoid Arthritis

## 5.1 Introduction

### 5.1.1 Recent findings of pathologically expanded immune cell subsets in RA by single cell technologies

Immune dysregulation is apparent across all stages of RA and is a widely accepted concept amongst researchers within the field<sup>56</sup>. Whilst extensive research has been directed towards understanding the chronic inflammatory interactions at the cellular and molecular levels that occur in established RA, the possibility to extend this research to detect RA at its earliest opportunity, inform stratification of treatment, and/or identify individuals who may become refractory to treatment are unmet clinical needs which need to be resolved.

Flow cytometry has proven to be an invaluable tool for probing the immune cell populations and indeed continues to be the preferred tool as it is both widely accessible and an established technique. However, the major disadvantage of flow cytometry is the limited number of proteins markers which can be included thus restricting researchers to fully embrace the heterogeneity of cells present in biological samples. Therefore the need for new technologies that could incorporate multiple protein markers for comprehensive immune profiling was necessary to begin answering long-standing questions and indeed several reviews by experts encourage the adoption of novel single cell technologies to revolutionise and expedite discovery of pathological immune cell subsets in RA<sup>152,266–268</sup>. Single cell technologies targeting transcriptomic and proteomic profiling have provided an opportunity to interrogate the immune landscape at a high dimensional level. Large collaborations such as the AMP RA and SLE network which spans government, industry and non-profit organizations, aim to identify novel immune cell populations and molecular signatures which associate with RA pathology using cutting-edge single cell technologies<sup>269</sup>.

To date, significant immune cell populations have been reported by the AMP group including T peripheral helper (Tph) cells which were initially observed to be expanded in the rheumatoid joint<sup>157</sup>. The functions of Tph cells are discussed in detail in Chapter 6 but will be briefly mentioned here. Tph cells are characterised as PD-1<sup>hi</sup> CXCR5<sup>-</sup> memory CD4 T cells and do not display exhaustive traits but rather actively recruit T follicular helper cells and B cells to the inflamed synovium, through the production of chemokines IL-21 and CXCL13<sup>157</sup>. Tph cells were characterised across the three most commonly assayed compartments in RA: peripheral blood, synovial fluid and tissue. Tph cells did not directly differentiate seropositive patients from seronegative although were found to be more robustly increased in seropositive patients. Tph cells did not significantly correlate with many clinical parameters including age, gender, disease duration or treatment although a separate cohort of 23 seropositive patients were independently assessed before and after commencing new treatment and Tph cells were observed to decrease in frequency when disease activity reduced<sup>157</sup>.

The discovery of pathologically expanded Tph cells in RA demonstrates how high-dimensional immune profiling can reveal subtle immunophenotypic changes in the immune populations which can be overlooked by traditional approaches. However, in the case of Tph cells, they have also been identified in other autoimmune diseases including systemic sclerosis, IgG4, SLE, and psoriasis

vulgaris but not in seronegative RA and spondylarthritis which suggests a common mechanistic pathway shared by autoantibody mediated diseases <sup>270–273</sup>. However, whilst mass cytometry is an excellent tool for detecting multiple immune cell populations it is less useful when discerning which of these populations are directly responding to autoantigens and which are non-specifically recruited to inflamed tissue <sup>158</sup>. It is often postulated that by answering this question, eradication of RA may be possible but to date specific autoreactive T cells have proven difficult to identify. A landmark paper published in 2019 observed CD4<sup>+</sup> T cells in multiple autoimmune disease settings including celiac disease which is an ideal model to study autoreactive T cells as it is directed towards one specific antigen, gluten <sup>274</sup>. Antigen-specific lymphocytes can be specifically detected by using gluten-derived peptide-loaded MHC tetramers and Christophersen et al., demonstrated that gluten-specific CD4<sup>+</sup> T cells within the gut had a phenotype that closely resembled Tph cells identified in the RA joint <sup>157,158,274</sup>. The Tph cells in celiac gut tissue expressed high levels of HLA-DR, PD-1 and CXCR3 and analysis by bulk RNA sequencing, showed that this T cell subset had increased expression of genes *IL21* and *CXCL13*, important for B cell differentiation and chemoattractant, again consistent with the Tph cell phenotype <sup>274,158</sup>.

In a short Nature Rheumatology review article published in 2019, Rao puts forward a compelling explanation for the detection of Tph cells in other autoimmune settings including celiac disease suggesting a common mechanistic pathway in which Tph cells activate B cell function <sup>158</sup>. In turn, the pathologic expansion of Tph cells may be numerically useful as it may reflect the degree of T cell-B cell interaction taking place in patients which may correlate with disease activity or severity <sup>158</sup>. An intriguing follow up experiment would be to obtain RA gut tissue and observe whether Tph cells could provide a link between gut dysbiosis and onset of RA by assessing mucosal triggered antigen-specific immunity which cross-react with joint antigens <sup>275</sup>. Smiljanovic et al., showed that activated B and T cells in synovial tissue were associated with response to microbially activated monocytes and macrophages, with further analysis revealing elevated levels of CXCL13 protein expression in RA SF and serum <sup>275</sup>. It is unclear whether these findings can directly suggest the involvement of Tph cells in RA pathology but consistencies such as the detection of CXCL13 in RA samples continue to make the Tph cells a compelling population to investigate further.

In addition to the discovery and characterization of Tph cells, the AMP group have identified another CD4 memory T cell population defined as CD27<sup>-</sup> HLA-DR<sup>+</sup> were expanded in RA and demonstrated effector functions including production of IFN $\gamma$  and cytolytic factors which upon successful treatment, contracted <sup>261</sup>. This population of cells will be further detailed in Chapter 6 however, it is important to note that this subset of cells were distinct from Tph cells and were not characteristic of functional exhaustion but did suggest that CD27<sup>-</sup> HLA-DR<sup>+</sup> CD4 T cells were chronically activated <sup>261</sup>.

Mass cytometry and single cell transcriptomics have been particularly valuable when exploring populations within tissues. In RA, an increase in publications which focus on deep immunophenotyping of synovial tissue is becoming apparent. Excellent review articles on the progress being made in RA synovial research have been published highlighting the importance of this research <sup>13,14,146,276–280</sup>. As RA is a disease which primarily manifests in the joints, it is both obvious and important that attention has returned to the synovium with the benefit of advanced technology for more robust immune profiling. Both Tph cells and CD27<sup>-</sup> HLA-DR<sup>+</sup> CD4 T cells were identified in a small cohort of synovial biopsies (n=3 and 9 respectively). Presumably as both confidence in novel technology and clinical and laboratory techniques associated with these technologies become more established, larger biopsy cohorts are becoming possible as evidenced by Zhang et al., in 2018 <sup>161</sup>. 51 RA and OA synovial tissue biopsies were comprehensively immunophenotyped to detect which cell

populations perpetuated joint inflammation, by using a combination of single cell and bulk RNA sequencing and mass and flow cytometry to immunophenotype single sorted cell populations including T and B cells, monocytes and fibroblasts from synovial tissues<sup>161</sup>. The tissue architecture showed that OA tissue was largely composed of stromal fibroblasts and endothelial cells which was also observed in leukocyte poor RA tissue<sup>161</sup>. Analysis showed an expansion of specific immune cell immunophenotypes that associated with RA<sup>161</sup>. These included: IL1B<sup>+</sup> proinflammatory monocytes, CD11c<sup>+</sup> autoimmune B cells (which have also been detected in SLE<sup>281</sup>), and granzyme B and K producing CD8 T cell subsets<sup>161</sup>.

Whilst traditionally CD4 T cells have been extensively studied in RA pathology, renewed interest in the CD8 T cells and their implication in RA is gaining traction. Single cell RNA sequencing has revealed that CD8 T cells in the inflamed RA synovium are prolific producers of IFN $\gamma$  and TNF transcripts but have reduced cytotoxic potential detected by lower granzyme B and perforin transcript expression in seropositive RA patients<sup>161,282</sup>. Furthermore, when assessing the altered cytotoxic potential, it was observed that CD8 T cells had low expression of CD57 and CX3CR1<sup>282</sup>. The authors of the study postulate that CD8 T cells in the RA joint signal to synovial fibroblasts to produce IL-6 and other inflammatory mediators which in turn suggests that these CD8 T cells produce IFN $\gamma$  in an antigen independent stimulation<sup>282</sup>. It is also evident that regulation of CD8 T cells and their effector function is altered based on the upregulation of PD-1 on CD8 in the RA joint compared to peripheral blood. Contrary to expectation, these CD8<sup>+</sup> PD-1<sup>+</sup> cells in the joint were not considered exhausted but rather displayed a heightened proinflammatory state, indicating that checkpoint regulation within the joint is aberrant<sup>283</sup>.

Moreover, it was also observed that IL-6 producing THY1<sup>+</sup> HLA<sup>high</sup> sublining fibroblasts were expanded in RA synovial tissue compared to OA which was elaborated upon by Mizoguchi et al. in 2018 where fibroblast heterogeneity and subsets in RA synovial tissue were further detailed<sup>161,284</sup>.

### **5.1.2 Recent findings of pathologically expanded fibroblast subsets in RA by single cell technologies**

Findings by Mizoguchi et al., demonstrates how mass cytometry and single cell transcriptomics can be versatile and applied to non-immune cell populations. This has been particularly useful in the case of fibroblasts. Biologics used in RA treatment, target various aspects of the immune system, and whilst these are successful in reducing chronic inflammation, they fail to reverse or cure joint damage providing a compelling argument to investigate non-immune cells such as fibroblasts<sup>285</sup>. Whilst fibroblasts have often been functionally type casted as structural cells, studies have revealed that they are capable of producing chemokines and cytokines in response to environmental cues<sup>285</sup>. In their investigation, Mizoguchi et al., set out to thoroughly immunophenotype the fibroblast cell population by initially assessing the expression of four surface markers: podoplanin, THY1, cadherin-11 and CD34 having been selected from a wider immunophenotyping panel<sup>284</sup>. These markers were able to phenotypically distinguish 7 fibroblast subsets which was then further refined by single cell RNA sequencing to three major fibroblast subsets<sup>284</sup>. In addition, fibroblast subsets localized to specific regions in the synovium, with CD34<sup>-</sup> THY1<sup>+</sup> cells located deep in the sublining layer<sup>284</sup>. Conversely CD34<sup>-</sup> THY1<sup>-</sup> fibroblasts were observed in the lining layer and CD34<sup>+</sup> fibroblasts were found in both the lining and sublining layers of synovial tissue with the majority of fibroblasts expressing cadherin-11<sup>284</sup>. This observation reflects a study previously conducted in mice where FAP $\alpha$ <sup>+</sup> THY1<sup>+</sup> immune effector fibroblasts were detected in the synovial sublining whereas FAP $\alpha$ <sup>+</sup> THY1<sup>-</sup> destructive fibroblasts were restricted to the synovial lining layer<sup>286</sup>. Further evidence in

support of fibroblasts specifically locating within the synovium was observed using an inexpensive technique known as droplet-based single-cell RNA in 5 RA synovial tissue biopsies<sup>287</sup>. Here Stephenson et al., detected CD55<sup>+</sup> fibroblasts locating in the synovial lining and THY1<sup>+</sup> fibroblasts were localized to the sublining with THY1 capable of further segregating fibroblast populations into podoplanin<sup>+</sup> and podoplanin<sup>-</sup> fibroblasts populations<sup>287</sup>.

When comparing fibroblast composition between RA and OA, CD34<sup>-</sup> THY1<sup>+</sup> cells comprised a median of 22% of total fibroblasts compared to 8% in OA, however, CD34<sup>-</sup> THY1<sup>-</sup> cells were decreased in RA at 15% compared to 48% in OA tissue<sup>284</sup>. These findings provide evidence that fibroblasts may be functionally altered in RA compared to OA, and lend support to previous observations in mice, where adoptive transfer studies of fibroblast subsets within the joint showed that FAP $\alpha$ <sup>+</sup> THY1<sup>+</sup> and FAP $\alpha$ <sup>+</sup> THY1<sup>-</sup> fibroblast subsets had functionally different roles. FAP $\alpha$ <sup>+</sup> THY1<sup>+</sup> fibroblasts were implicated in chronic inflammation and had minimal effect on bone and cartilage destruction, whereas the inverse was true for FAP $\alpha$ <sup>+</sup> THY1<sup>-</sup> fibroblasts<sup>286</sup>. Therefore CD34<sup>-</sup> THY1<sup>+</sup> cells may associate with worsening clinical outcomes and this was reflected by clinical correlative data showing that patients with swollen joints had increased CD34<sup>-</sup> THY1<sup>+</sup> cells and CD34<sup>+</sup> fibroblasts<sup>284</sup>.

Recent data has reported the expansion of preinflammatory mesenchymal (PRIME) cells in blood have been detected two weeks prior to an RA flare in a longitudinal cohort of patients collected at 364 time points spanning over 4 years in four patients<sup>288</sup>. RNA sequencing revealed the expansion of CD45<sup>-</sup> CD31<sup>-</sup> podoplanin<sup>+</sup> PRIME cells which were also associated with an activated B cell profile prior to flare and in the presence of a flare PRIME cells were subsequently decreased<sup>288</sup>. PRIME cells demonstrate characteristics similar to fibroblasts and are detectable in blood is intriguing, and may support the hypothesis that fibroblasts may traffic from the blood to the synovium, and potentially precede the inflammatory events that take place in the joint<sup>288</sup>.

Treatments targeting fibroblasts have become a focus as they may provide another option for treating RA, particularly for those patients who have become resistant or refractory to existing treatments. Seliciclib is an oral cyclin-dependent kinase inhibitor which is capable of suppressing synovial fibroblast proliferation<sup>289</sup>. An initial clinical trial determining the maximum dose of seliciclib was conducted in 15 patients who were refractory to TNF blockade, showed no serious adverse events were observed thus paving the way for future clinical use<sup>289</sup>.

### 5.1.3 Aims

The overall aim of this chapter is to chronicle the immunophenotypes across the different stages of RA with particular focus on the early RA cohort of patients using the Diffcyt script. In addition, the chapter will aim to validate the robustness and sensitivity of the Diffcyt script to interrogate the immune landscape within this cohort. This chapter aims to contribute to the conceptual framework of 'stages' within RA; that is the understanding that RA is not a uniform disease but more akin to a syndrome. Whilst this concept is widely accepted, no definitive biomarker or immune signature has been consistently identified/correlated with a stage of RA. To provide earlier intervention in RA and thus have better treatment efficacy, understanding what distinguishes and differentiates RA stages will help clinical management of RA and thus improved patient care. Using the combination of a diverse cohort that spans different stages of RA, mass cytometry and an automated pipeline, this analysis will aim to comprehensively phenotype the immune cell populations in blood and observe whether any differences between the stages are detectable.

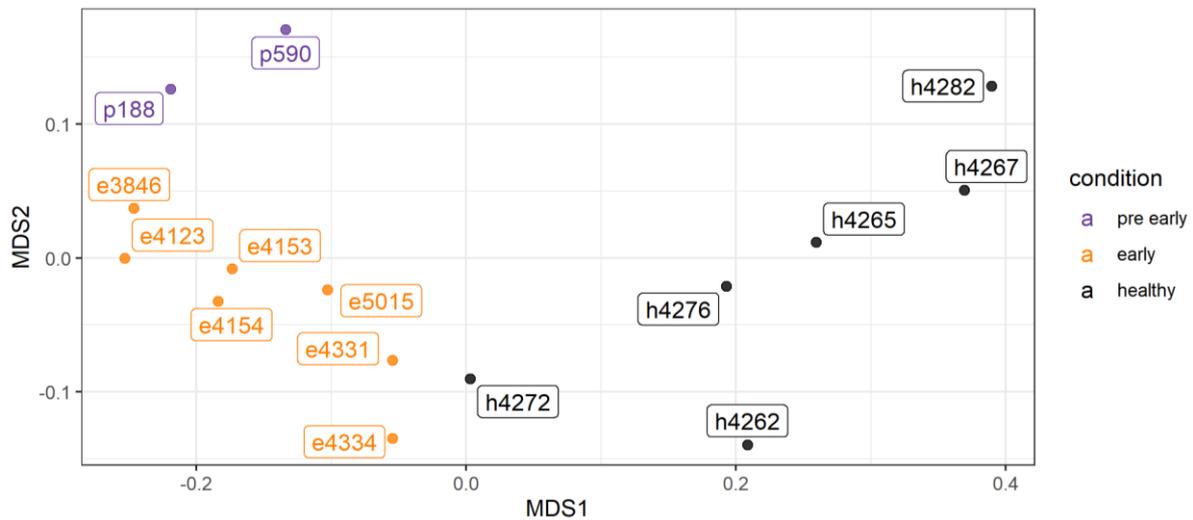
## 5.2 Results

The approach to analysis was designed to test both the validity of the Diffcyt script analysis pipeline and to explore the spectrums of the clinical stages within the RA cohort collected for this study. Due to batch variation as discussed in Chapter 4, not all samples were included for analysis and it was decided that only samples that successfully passed the data pre-processing steps should be included for analysis. Where possible, if samples originated from the same batch, these were analysed together, again removing inconsistent staining or signal during acquisition.

### 5.2.1 Evidence of immune dysregulation in two pre-clinical RA patients

The first aim of the analysis was to determine if any difference could be detected in any of the clinical stages within RA using whole blood. As these blood samples were not stimulated or activated in vitro, it was not known whether the mass cytometer and the panel designed, would both be sensitive enough to detect changes within RA and between RA and health. Therefore, Diffcyt analysis first looked at whether there was a difference between health, pre-RA and early RA and specifically whether immune dysregulation could be detected at the clinically earliest opportunity. Pre-RA refers to at risk patients identified by historical evidence and CCP<sup>+</sup> titer but were not diagnosed with RA. 2 patients formed the pre-RA cohort and were compared with 6 healthy donors and 7 early RA donors (with varied serology status – full clinical data in appendix) taken from different batches to ensure findings were not restricted to one batch and therefore applicable to patients in the same stages of RA.

Analysis began in the Diffcyt script by downsampling 27,000 events from each sample and a new .fcs file was created for each sample containing the downsampled events. 27,000 events were selected as this was the lowest number of events in one healthy sample included in this analysis. Eliminating the one sample and increasing the number of events to 80,000 did not affect the number of cell populations that were identified downstream and it did not change the outcome of the findings. This demonstrated that the Diffcyt pipeline was capable of processing a small number of events and identifying differences. The metadata placed all the samples for analysis into the three conditions pre-early RA, early RA and healthy. Evidence of a difference was initially observed in the MDS plot (Figure 5-1) which showed that pre-RA and early RA donors clustered together to the left of the MDS1 axis and were distinct from healthy donors. Interestingly both pre-RA donors were closer together on the MDS plot and were not within the early RA donors suggesting staining similarity. The MDS plot was calculated based on the median (arcsinh-transformed) marker expression of 26 lineage markers and 15 functional markers (as identified in Table 4-1 in Chapter 4) across all cells measured for each sample. Change in median represents distance between samples in the plot with those samples with similar medians, in closer proximity on the plot and where median marker expression is different, these are further distanced.

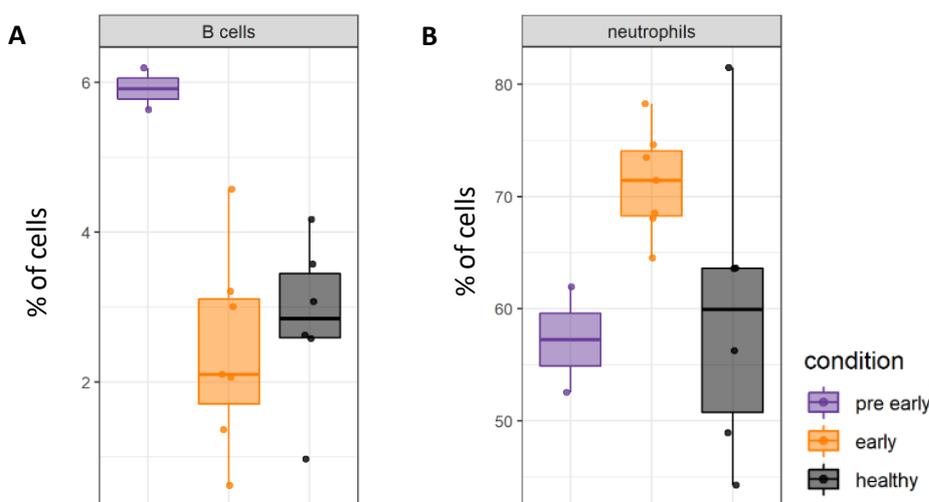


**Figure 5-1 | Multi dimensional scaling (MDS) plot observes the staining similarity globally across all samples included for analysis using 27,000 events from each sample.**

2 pre-clinical RA (denoted in purple and marked as (p) whole blood samples were analysed with 7 early RA (denoted in orange and marked as (e) whole blood samples run in the same barcoded batch and 6 healthy (denoted in black and marked as (h) whole blood samples. The numbers correspond to the cohort study number assigned to these samples. The MDS1 and MDS2 axes depict similarity between the samples represented by a dot and the proximity of this indicates similar patterns of staining. The 2 pre-clinical RA patients are in close proximity to one another and also align on a negative MDS1 axis whereas health separates from both pre-clinical and early RA patients on the positive MDS1 axis.

FlowSOM Clustering analysis detected 14 cell subsets. Differential abundance and differential expression of functional states testing did not reveal any significant differences although trends could be identified. Analysis of cell abundance across the populations revealed percentage increase in B cells in pre-RA compared to early RA and health (Figure ). Another interesting observation was a percentage increase in the neutrophil population in early RA compared to health and pre-RA (Figure ) which was analysed further.

### 5.2.1.1 Cell percentage increase in B cell population at the pre-clinical RA stage



**Figure 5-2 | Increase in percentage of B cells in pre-clinical RA in comparison to health and early RA.**

**A** shows an increase in percentage of B cells in preclinical RA (n=2) compared to early RA (n=7) and health (n=5). **B** Increase in the neutrophil population in patients with early RA compared to health and preclinical RA. Results not significant after adjusted p values. This comparison shows that in the pre-clinical phase of RA, immune dysregulation is observed and thus supports the hypothesis that across disease stages of RA, a change in immune cells is detectable.

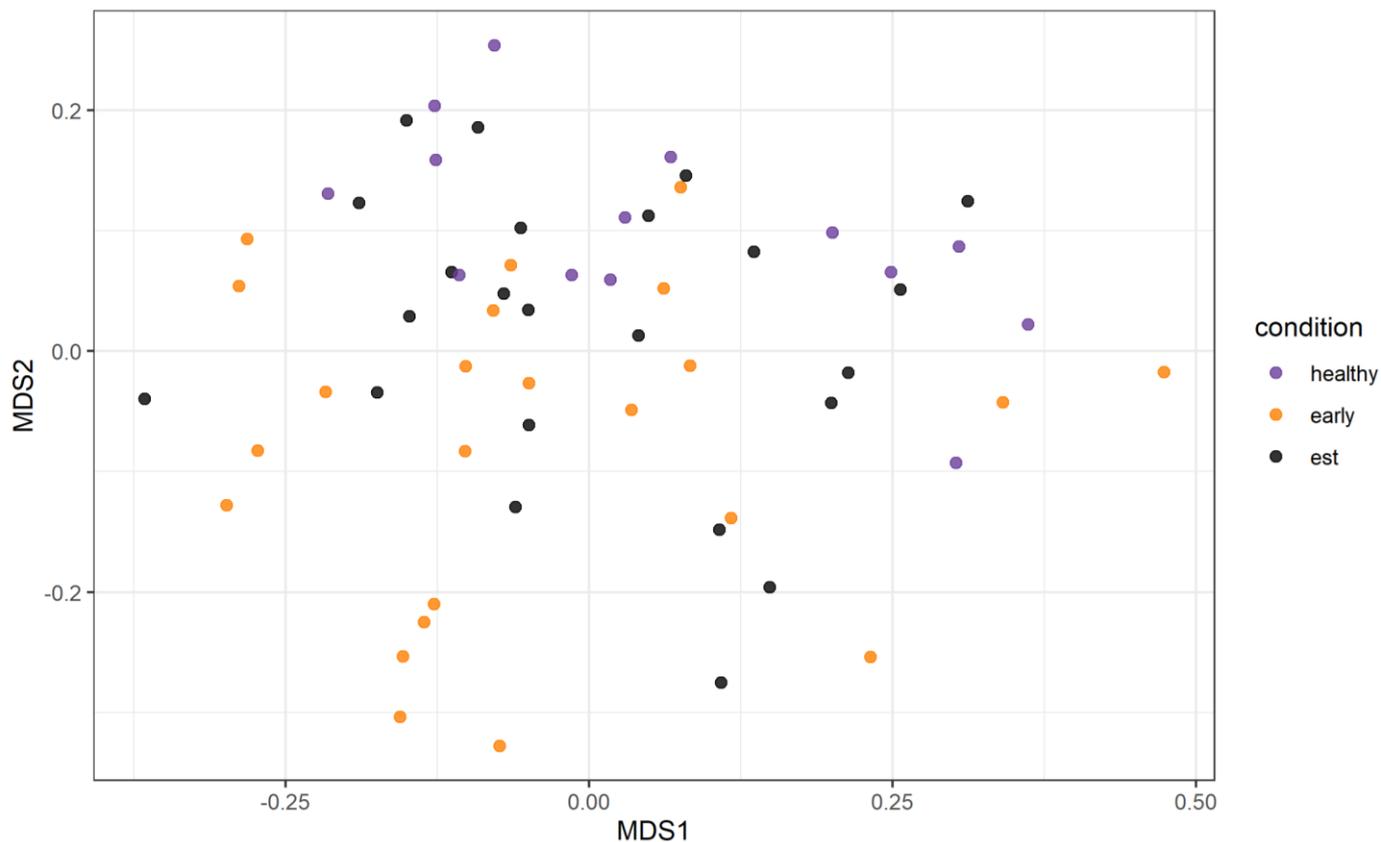
## **5.2.2 Analysing MDS plots to determine immune relationship between health, early RA and established RA**

Due to the batch effects incurred during acquisition of the cohort on the mass cytometer, it was not possible to run all samples simultaneously through the Diffcyt script as the script would be influenced by the staining variation and not by the underlying biology. Therefore the analysis could not be performed completely unbiased. It is also worth mentioning that supercomputing power at the time of this analysis was not available and furthermore it is unlikely the Diffcyt script would be able to cope with the complex data set presented here. Furthermore, should all samples be analysed together, the many parameters to this data set aside from just the multi-marker aspect including patient demographics, underlying immunology affected by treatment/progression/stage of the disease complicates analysis and can overwhelm the process of identifying what is a true representation of the biology and what is an artefact.

Each run performed through Diffcyt takes approximately 10-12 hours with the clustering step taking a significant amount of time. The clustering step can be considered as a two step process: the run time taken through FlowSOM to obtain the 40 clusters and the process of piecing together the information from the heatmap to make sense of the immune populations represented by the heatmap. This is time-consuming and requires careful judgement to decide whether to combine similar expression clusters together or whether these are bona fide phenotypes that represent a biological difference. This is much easier said than done particularly as each FlowSOM run changes the numbering of the cluster and can identify 40 different clusters each time FlowSOM runs. Therefore for this project, FlowSOM was run 5 times for each analysis to assess whether the same clusters were being detected with the same markers identifying the population. Another challenge is different clusters are identified in different patient cohorts even if the same lineage and functional markers have consistently been applied. Assessing which cell populations are common across all the analyses can therefore be another rate limiting step in the analysis process.

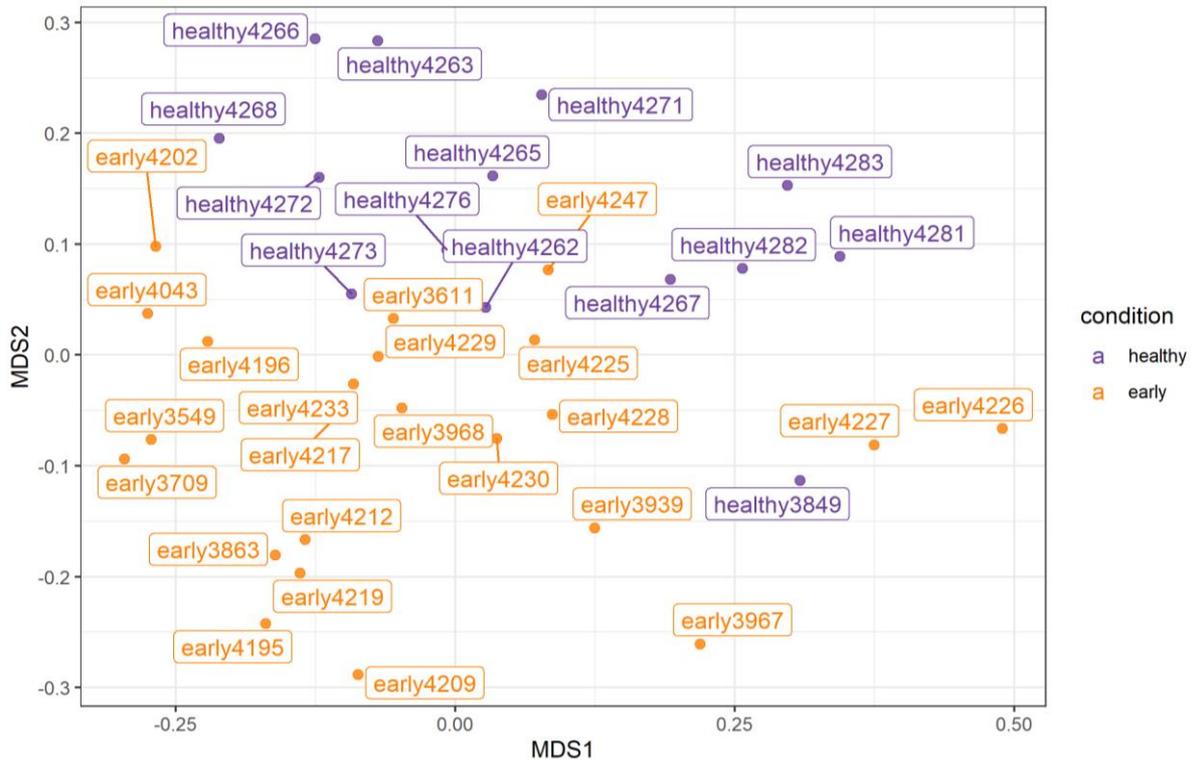
The MDS plot can be considered a global overview of the staining expression across samples and so therefore may overlook subtle underlying differences in staining expression. Also the MDS does not look at defined immune cell populations but rather staining expression which when analysed after clustering may better contextualise findings. However for this analysis, the MDS plot proved to be an invaluable tool to quickly ascertain whether the downstream analysis was likely to reveal significant immunological changes between the different conditions. An example of this was when deciding whether health, early and established RA would form distinct groups calculated by the MDS plot. Donors from all three conditions were taken from across all batches and had passed data pre-processing stages. Established RA referred to patients who had received treatment but did not discriminate in terms of type/amount of biologic/DMARDs received (full patient cohort details in the the Appendix). Figure 5-3 shows there was no clear distinction between healthy, early RA and established RA as determined by MDS plot however, when separated into health compared to early RA (Figure 5-4) and health compared to established RA (Figure 5-5), it is evident that the two groups are separated from health. An additional analysis which intended to observe whether early RA and established RA would separate into two distinct groups having removed the healthy controls.

Moreover, no distinction could be observed between early RA and established RA (Figure 5-6). For all comparisons the full Diffcyt script was run including the differential testing to ensure that no subtle differences were excluded but the MDS plots were informative. Therefore only early RA compared to health yielded significant results, which will be described further. The MDS analysis also suggested that whilst immune differences between RA irrespective of stage separate from health, this becomes less apparent between stages of RA. Given that early RA donors were newly diagnosed and had not received treatment and established RA patients included in the analysis were on treatment, it was thought that these two cohorts represented the polar ends of RA. However, whilst immunological differences may exist between these two groups of patients, the MDS analysis could not detect this and further downstream analysis did not identify significant changes between early RA and established RA.



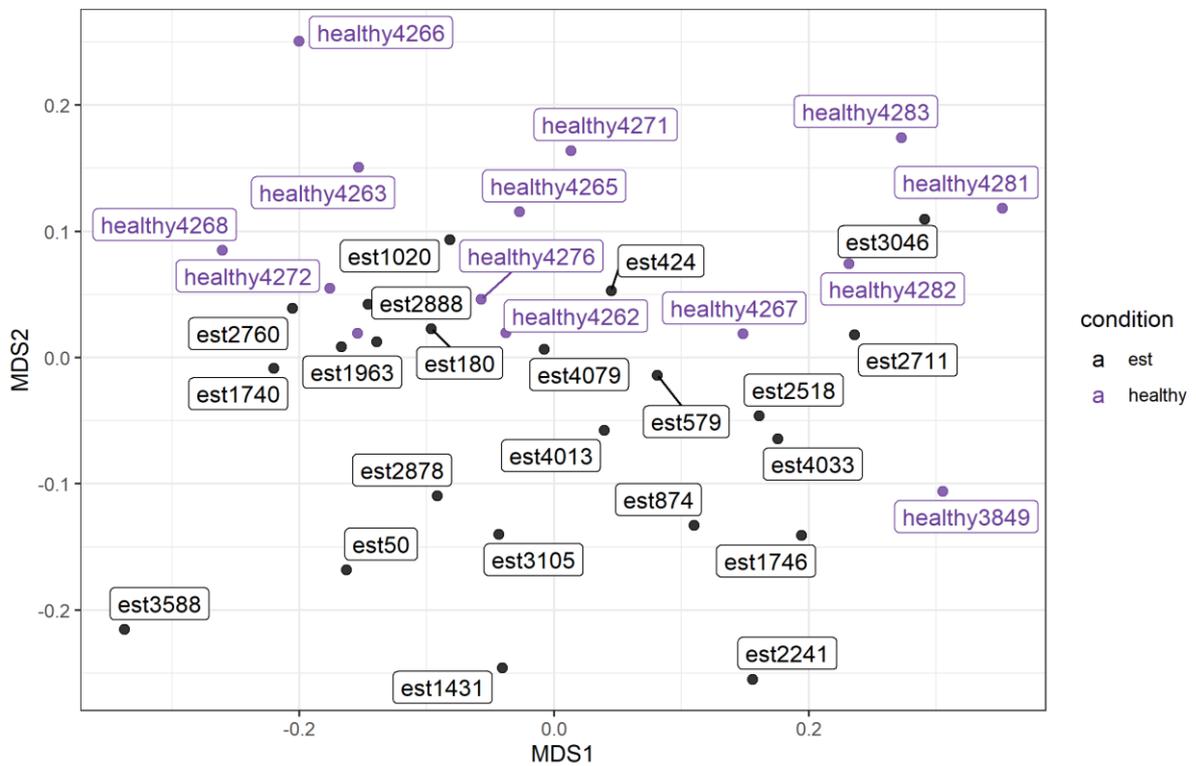
**Figure 5-3 | MDS plot analysis does not reveal 3 distinct groups of healthy, early RA and established RA**

Individual sample labels are removed to clearly see the relationship between the conditions. Multi dimensional scaling (MDS) plot does not show a clear separation of established RA (23), early RA (n=23) from health (n=14) by globally analysing the staining similarity across all samples using 27,000 events.



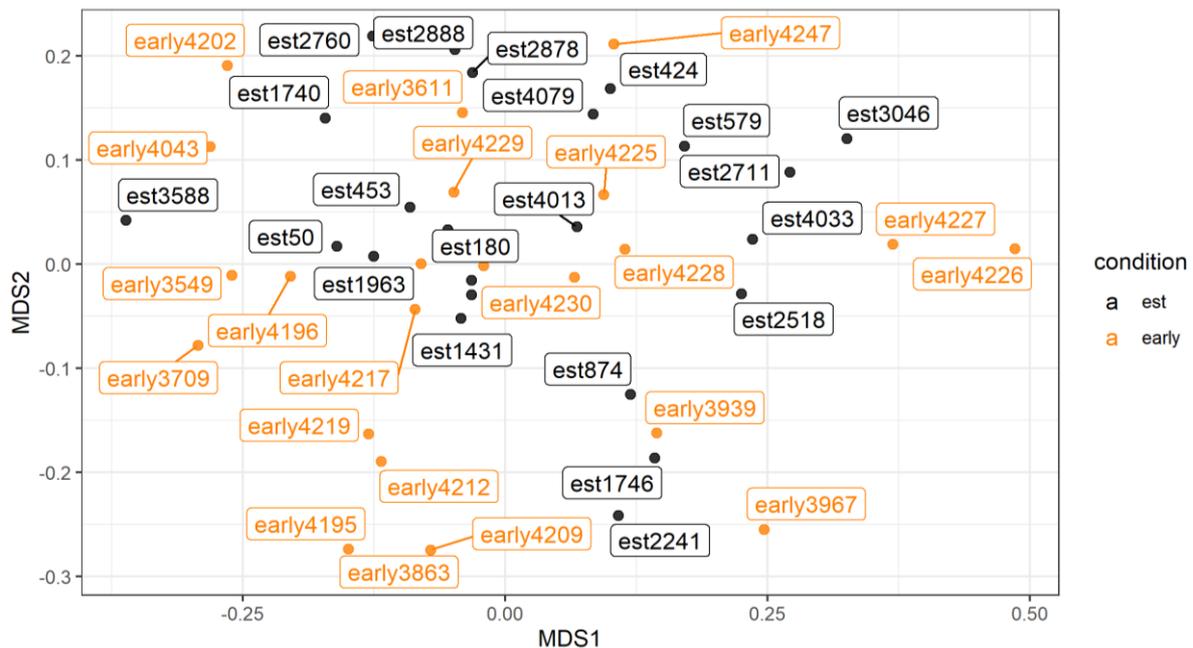
**Figure 5-4 | MDS analysis identifies early RA and health cluster into 2 distinct groups**

Multi dimensional scaling (MDS) plot shows separation of early RA (n=23) from health (n=14) by globally analysing the staining similarity across all samples using 27,000 events.



**Figure 5-5 | Distinction between health and established RA can be visualised by MDS plot**

Multi dimensional scaling (MDS) plot shows separation of established RA (n=23) from health (n=14) by globally analysing the staining similarity across all samples using 27,000 events.

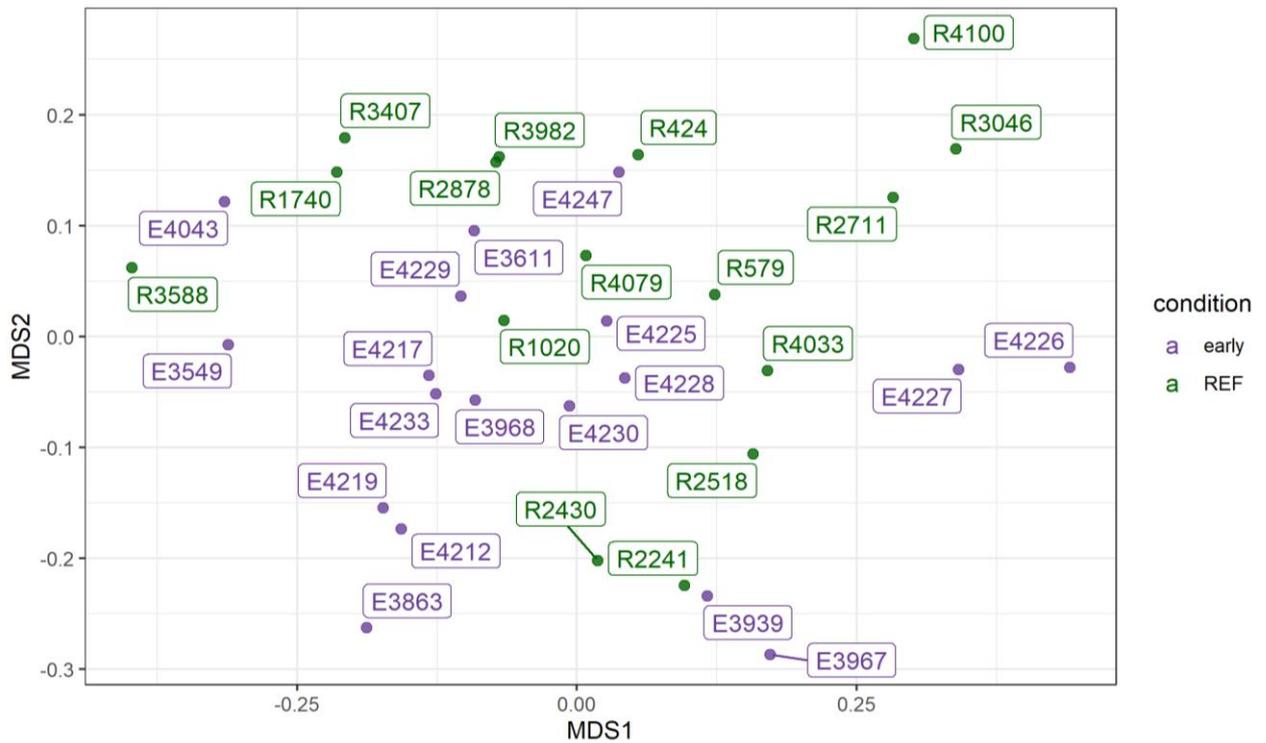


**Figure 5-6| No clear distinction between early RA and established RA by MDS analysis**

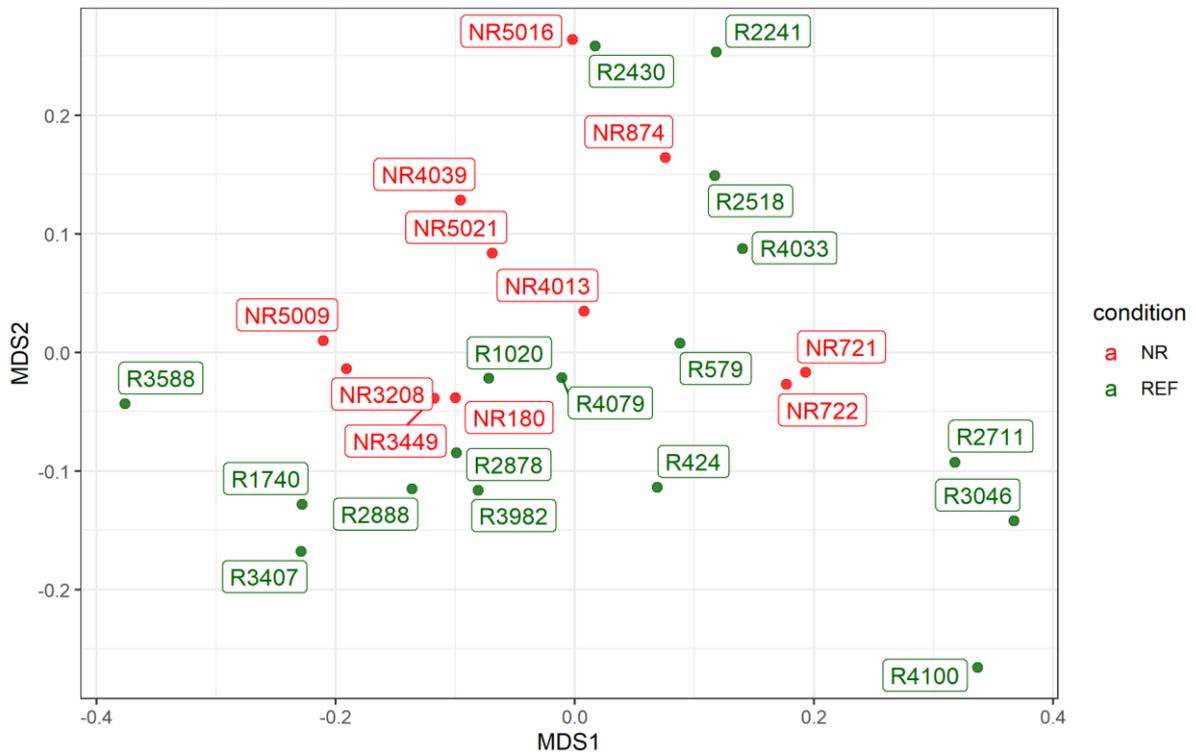
Multi dimensional scaling (MDS) plot shows early RA (n=23) does not distinctly separate from established (n=23) by globally analysing the staining similarity across all samples using 27,000 events.

To further stretch the polarity of the RA spectrum, another analysis was performed comparing early RA with refractory RA patients. Refractory RA patients for this analysis were defined using existing criteria<sup>290</sup> as those patients who had failed two mechanistically different biologics e.g. TNF and IL-6 blockade and were either in the ‘wash out’ phase before starting a third biologic or were already on treatment at the time of collection. MDS plot analysis revealed that no distinct separation between the two groups (Figure 5-7) was detectable suggesting that differences at the immune cell level become increasingly difficult to detect in more progressed RA. Another analysis comparing non-refractory patients (commenced initial biologic treatment or were treated with two different TNF blockers and refractory patients did not reveal any underlying immunological differences (Figure 5-8).

Therefore the MDS plots were useful when determining differences between stages of RA and whether these differences were likely to lead to significant changes within the immune populations. Although differential testing was applied to all the conditions run through the Diffcyt script, the MDS plots gave an accurate indication of whether significant changes were likely to be detected. Section 5.2.3 will detail the findings for differences between health and early RA.



**Figure 5-7 | Multi dimensional scaling (MDS) plot shows early RA (n=23) and refractory (REF) RA (n=23) are indistinguishable based on global staining similarity analysis.**



**Figure 5-8 | Distinction between non-refractory (NR) (n=11) and refractory (REF) (n=17) is not evident by multi dimensional scaling analysis**

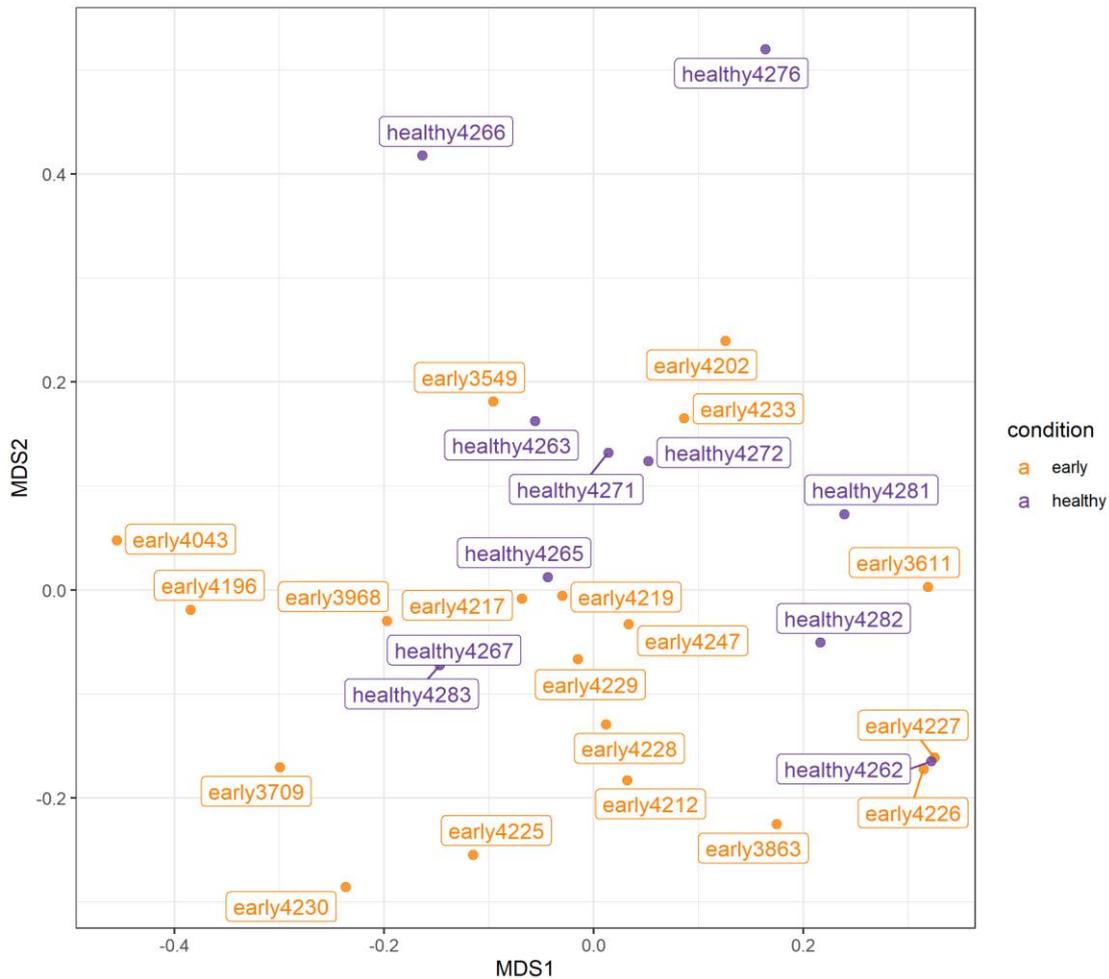
## 5.2.3 Distinct immune cell populations identify early RA from health

### 5.2.3.1 MDS plot analysis that granulocytes influence distinction between early RA and healthy donors

32 patients were collected in the early RA cohort, but only 23 were taken forward for analysis in Diffcyt to eliminate confounding results introduced by batch variations. As differences could be identified between pre-RA, early RA and health this provided evidence that differences were indeed present within the groups. To focus the analysis, healthy donors were compared with early RA donors to detect changes between the two groups. MDS analysis revealed that health and early RA donors broadly fall into two distinct groups when all immune cell populations are analysed (Figure 5-4).

To determine whether the absence/presence of immune cell populations impacted the MDS plot, it was first decided that only the granulocyte population would be included in the analysis. This was based on the analysis in Figure 5-2, where there is an indication that neutrophils are altered in early RA compared to health and pre-RA. However, this finding was in a small cohort of samples and it was not significant. Therefore to determine whether the granulocyte population were different in early RA compared to healthy donors, I gated just the granulocyte population using the markers CD11b and CD16. 15,000 granulocytes were analysed for 22 early RA donors and 12 healthy donors to determine and the MDS plot analysis showed that the granulocyte population was influencing the distinction between these two groups (Figure 5-9). Furthermore, the granulocyte population created a bipartite group within early RA patients, although it was indeterminate what could be driving this further division. The importance of granulocytes was further supported when all immune cell populations were included and the granulocyte population was removed as the distinction between health and early RA became blurred (Figure 5-10). This suggested that either phenotypically the granulocyte population were altered in early RA or were numerically altered in comparison to health. Although Diffcyt script was previously validated by Nowicka et al.,<sup>176</sup> it was important to determine that the script was compatible with the data generated in this study. It also provided confidence that number of immune cells were not influencing the MDS plot as 27,000 cells from each sample is a small number. A final analysis to ascertain that number of cells and cell types did not affect outcome of findings was performed by taking a pure population of 10,000 CD45<sup>+</sup> CD3<sup>+</sup> T cells from early RA and healthy donors and despite this increase in T cells and no other cell populations, this did not reveal further immune cell subsets or significant changes in the T cell population between early RA and health (data not shown). Therefore, differential abundance and expression testing was performed in the Diffcyt script across all immune cell populations to ensure subtle changes in immune cell populations were not missed.





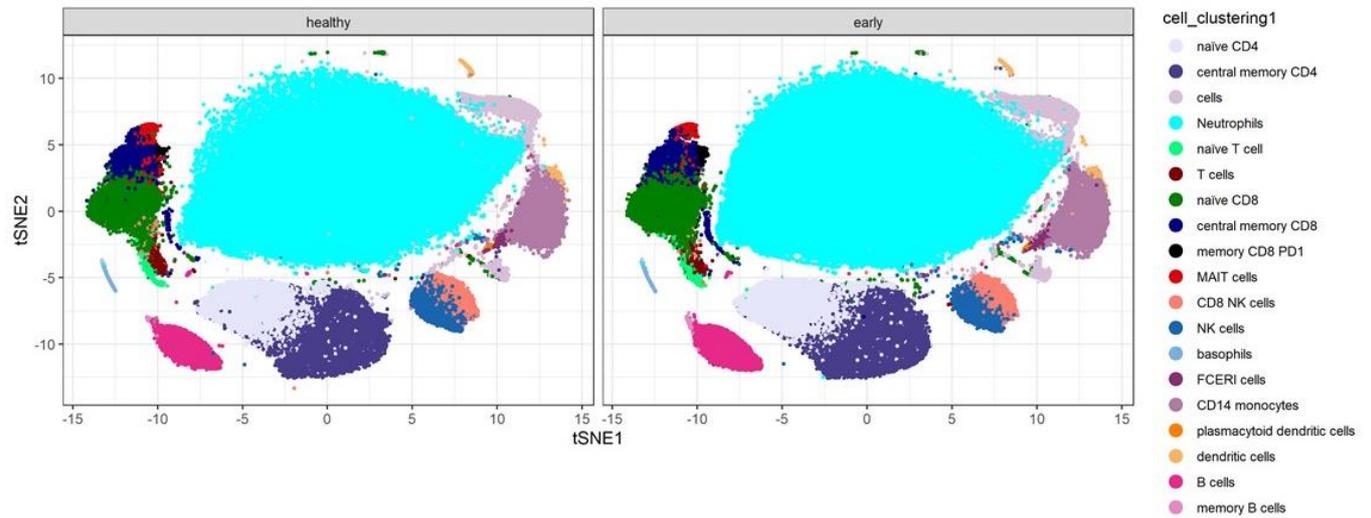
**Figure 5-10| Absence of granulocyte population results in a loss of distinction between early RA and health**

Inclusion of all other immune cell populations and removal of the granulocyte population no longer clearly distinguishes between health (n=11) and early RA (n=19) as represented by the multi dimensional scaling (MDS) plot.

**5.2.3.2 Differential abundance and expression testing reveals specific immune cell populations are dysregulated between early RA and health and span both adaptive and innate immune cell populations**

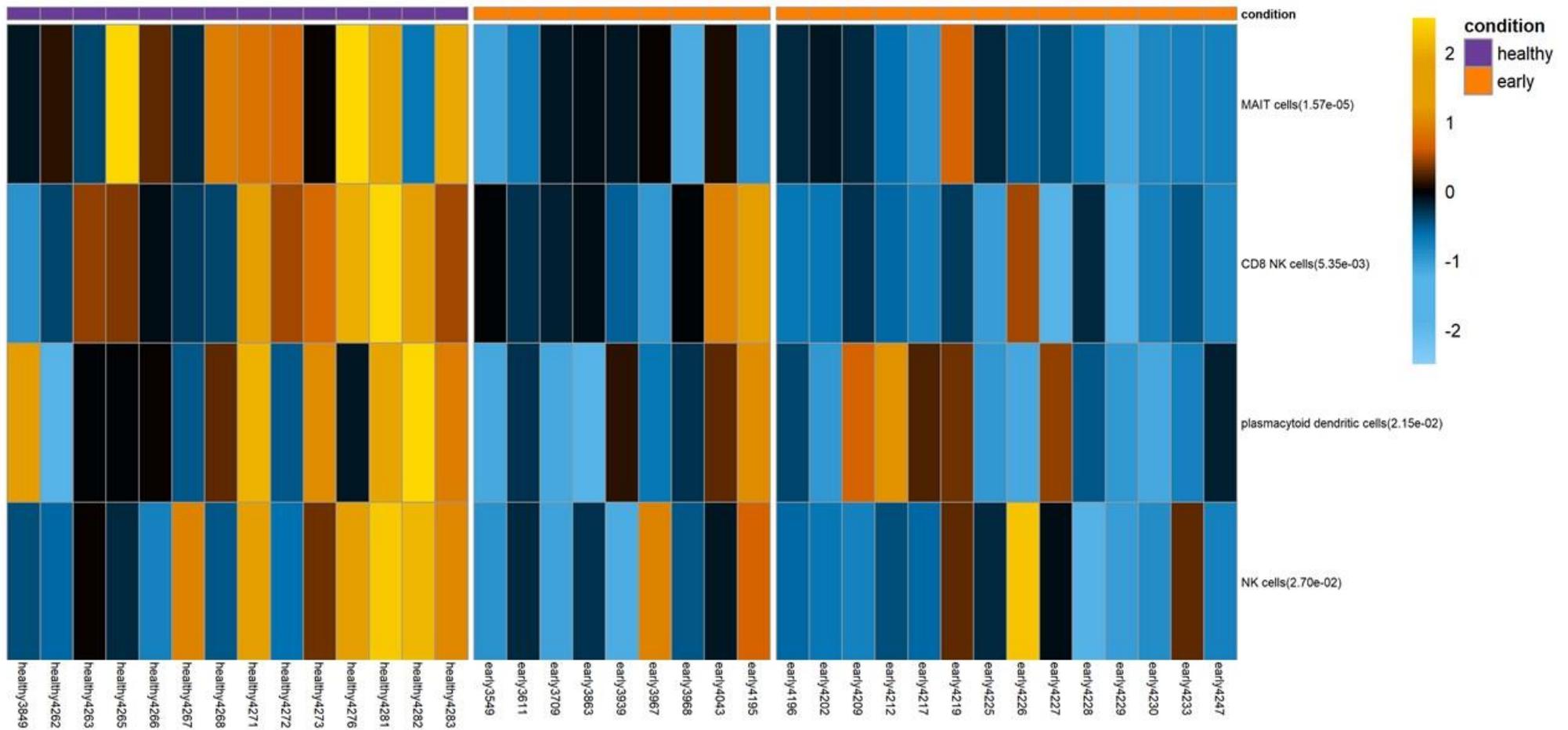
The key question to answer was which if any immune cell populations were influencing the distinction between early RA and health. Following FlowSOM clustering analysis (Figure 5-11), which identified 18 immune cell populations, differential abundance testing showed cell percentage decrease of MAIT cells ( $p=0.0000157$ ), CD8 NK cells ( $p=0.005$ ), pDCs ( $p=0.02$ ) and NK cells ( $0.03$ ) in early RA compared to healthy donors which was significant (Figure 5-12). The specific decrease of these cell populations suggested a migration of these myeloid cell populations from the blood to the joint and further still a selective recruitment of these populations to the joint microenvironment. Interestingly, the granulocyte population were not significantly expanded or decreased in percentage between early RA and health further suggesting that phenotypic differences were likely to drive the distinction between the two conditions. Figure 5-13 shows the boxplot distribution for percentage of cells in the neutrophil (health 42-82%, early RA 44-78%), plasmacytoid dendritic cells

(health 0.025-0.5%, early RA 0.02-0.13%), NK cells (health 1.2-6.5%, early RA 0.4-6.4%), CD8 NK cells (health 0.7-7.8%, early RA 0.2-3.2%) and MAIT cells (health 0.3-3.2%, early RA 0-0.7%). The decrease in cell populations in early RA did not further correlate with patient demographic factors including age, gender, serology, tender/swollen joint count or smoking history (full details in appendix) suggesting a common mechanism of cell migration to the joint occurring in early RA, a phenomenon that has been reported.



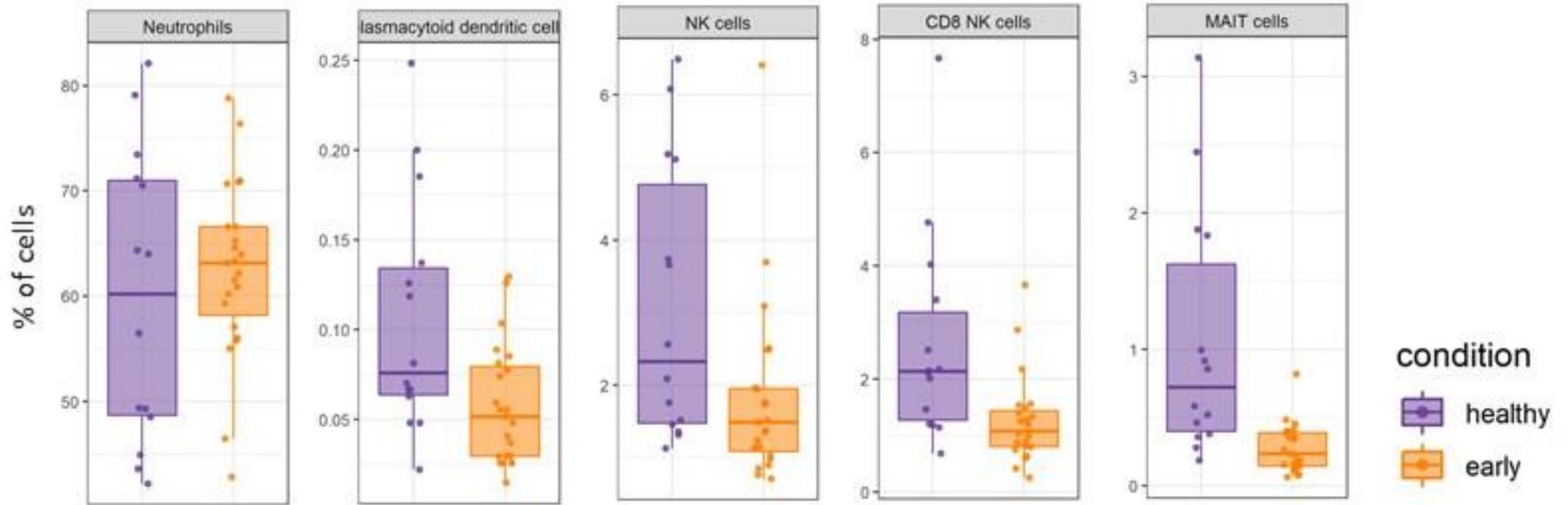
**Figure 5-11 | FlowSOM clustering identifies 18 immune cell populations in early RA and health whole blood profiles**

Immune cell populations in early RA and healthy donors are summarised by FlowSOM clustering.



**Figure 5-12 | Differential abundance heatmap represents significant changes in immune cell populations between early RA (n=23) and health (n=14)**

Differential abundance testing to identify significant changes in specific immune cell changes revealed a decreased percentage of MAIT cells ( $p=0.0000157$ ), CD8 NK cells ( $p=0.005$ ), plasmacytoid dendritic cells ( $p=0.02$ ) and NK cells ( $p=0.03$ ) in early RA compared to health. Adjusted p value is given. Significance was calculated to be  $\geq 0.05$ . Decrease in cell abundance is colour coded blue/black.



**Figure 5-13 | Differential abundance testing through the Diffcyt script shows percentage changes in specific cell populations between early RA (n=23) and health (n=14)**

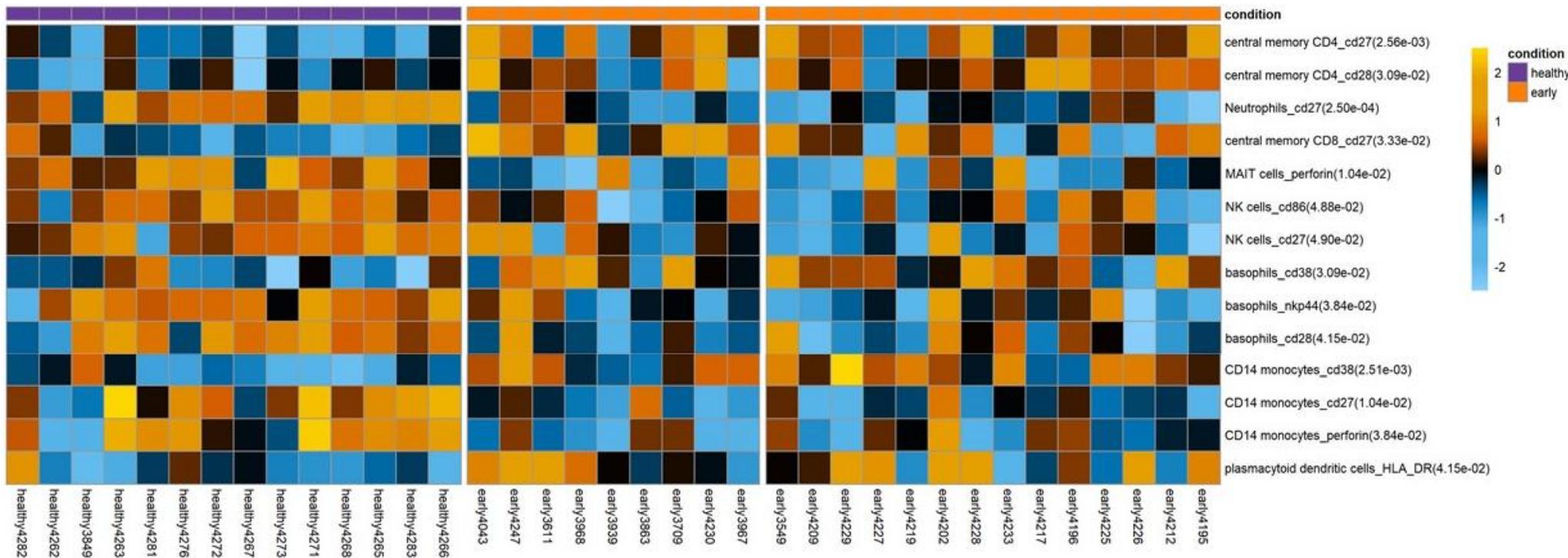
Plasmacytoid dendritic cells, natural killer cells, CD8 natural killer cells and MAIT cells all significantly decreased in early RA compared to health where significance was defined by a p value  $\geq 0.05$ . Difference in the neutrophil population was not significant in terms of percentage cells between early RA and health as observed by MDS .

Scales vary to represent percentage of cell population from 27,000 cells.

To further probe whether immunophenotypic changes occurred within the 18 identified immune populations, 15 functional markers as identified in the validation steps of this analysis, were examined by differential expression testing, stratified by the 18 cell populations. Differential expression testing detected that 8 immune cell populations had immunophenotypic changes. In total 14 immune cell phenotypes which had significantly changed between early RA and healthy donors represented in the heatmap (Figure 5-14) with significance p value reported next to the population. Table 5-1 shows the significant changes in median expression of functional markers in immune cell populations as identified by statistical testing in Diffcyt. Briefly, significantly increased median expression of activation markers in early RA were observed in CD27 and CD28 in memory CD4 T cells ( $p=0.003$  and  $p=0.03$  respectively), CD27 in memory CD8 T cells ( $p=0.03$ ), CD38 in basophils and CD14<sup>+</sup> monocytes ( $p=0.03$  and  $0.003$ ) and HLA-DR in pDCs ( $p=0.04$ ). Significantly decreased median expression of activation markers in early RA were observed in CD27 in neutrophils ( $p=0.0003$ ), perforin in MAIT cells ( $p=0.01$ ), CD86 and CD27 in NK cells ( $p=0.05$  for both), NKp44 and CD28 in basophils ( $p=0.04$  for both), CD27 and perforin in CD14<sup>+</sup> monocytes ( $p=0.01$  and  $p=0.04$  respectively).

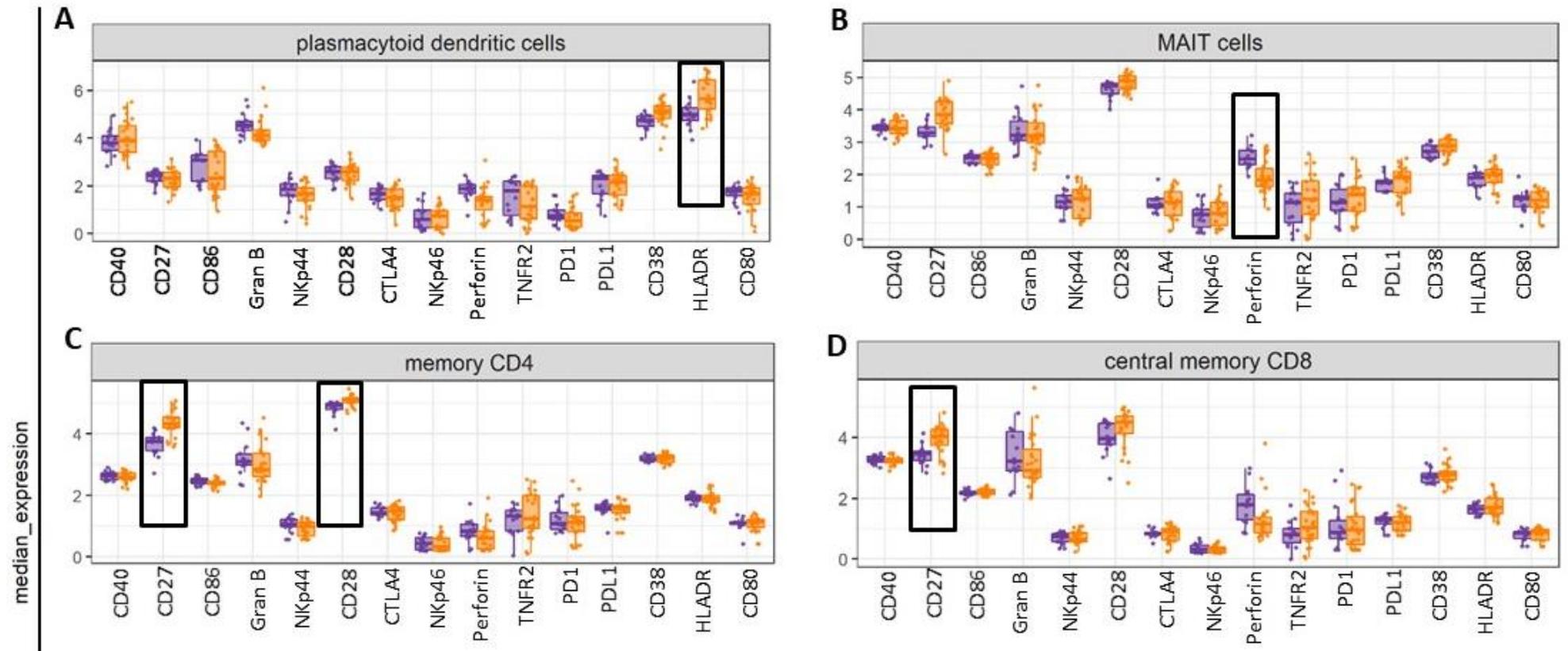
**Table 5-1 | Significant changes in functional protein marker median expression in immune cell subsets in early RA patients identified by differential expression testing in Diffcyt pipeline**

Immune cell subsets	Increased median expression of functional markers	Decreased median expression of functional markers
Basophils	CD38	NKp44, CD28
CD14 <sup>+</sup> monocytes	CD38	CD27, perforin
Plasmacytoid dendritic (pDCs) cells	HLA-DR	-
MAIT cells	-	Perforin
Memory CD4 T cells	CD27, CD28	-
Memory CD8 T cells	CD27	-
Neutrophils	-	CD27
Natural killer (NK) cells	-	CD27, CD86

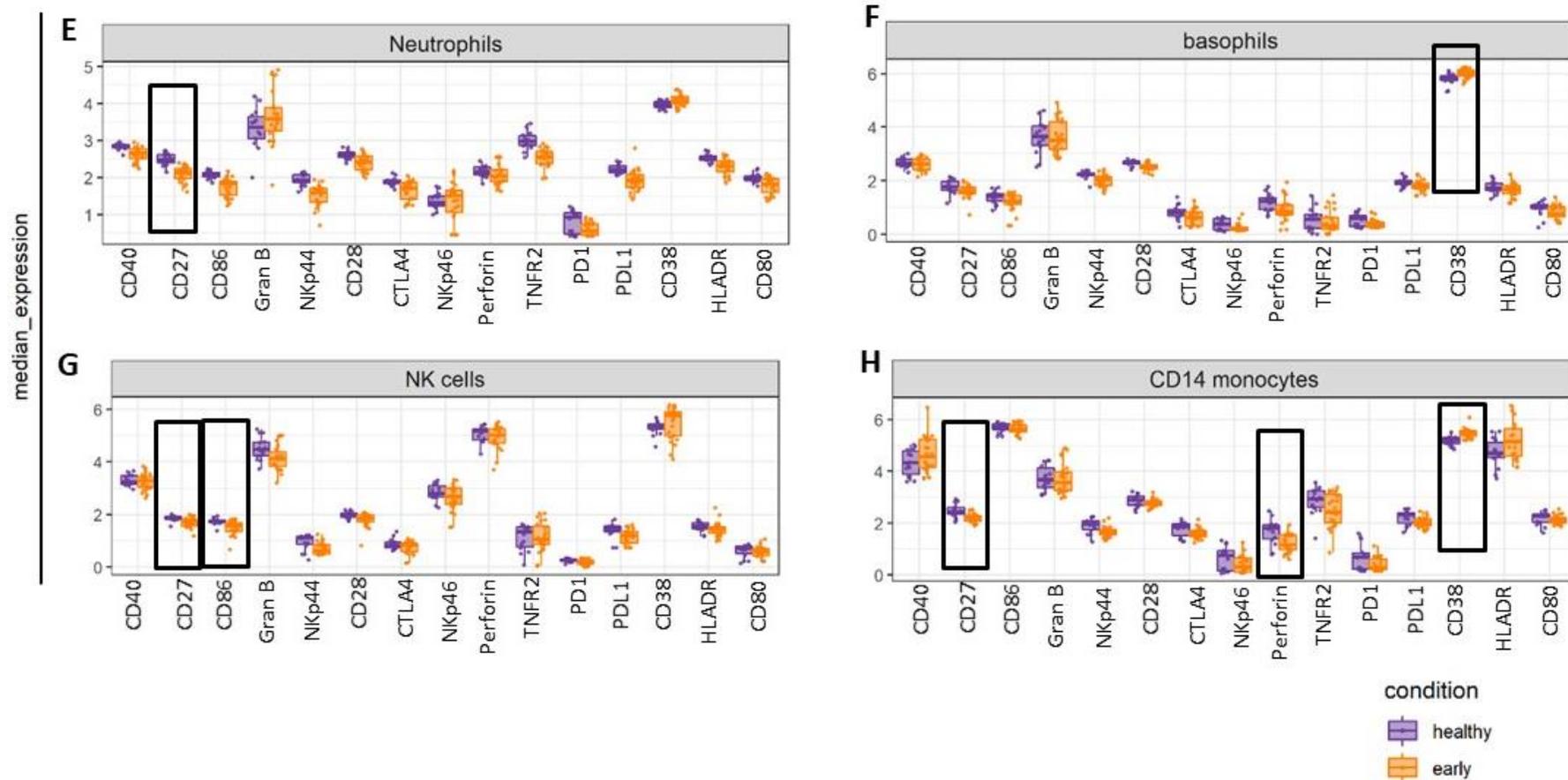


**Figure 5-14 | Differential expression heatmap represents 14 significant changes in functional activation markers across different immune cell populations between early RA (n=23) and health (n=14).**

Differential expression testing identifies significant changes in median expression of functional activation markers across immune cell populations in early RA compared with healthy donors. P value is shown in brackets next cell population. Significance was calculated to be  $\leq 0.05$ . Decrease in median expression of functional activation marker is represented by blue/black and expression of functional activation marker is represented by orange/yellow.



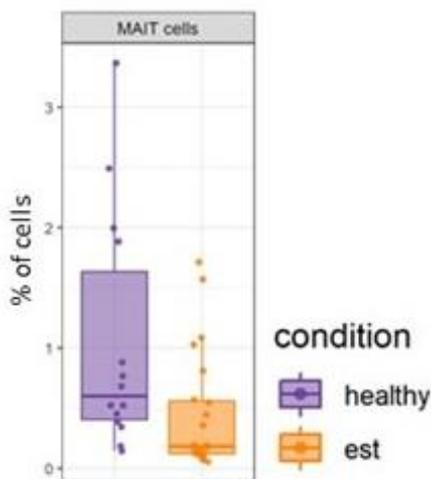
(Figure legend on next page)



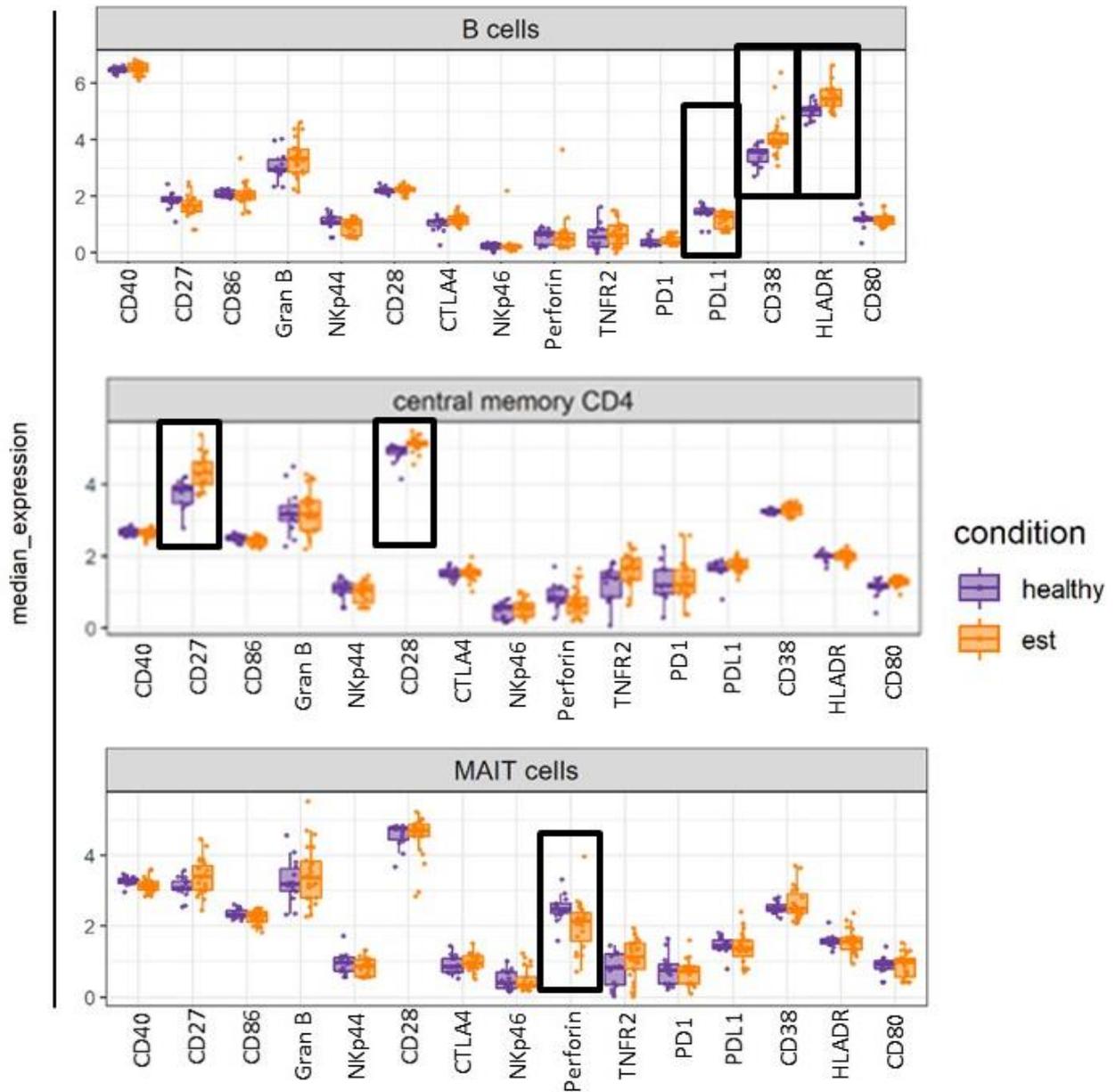
**Figure 5-15 | Differential expression testing through the Diffcyt script reveal significant changes in median expression of functional markers across different immune cell populations between early RA and health.**

**A** Plasmacytoid dendritic cells have an increased median expression of HLA-DR in early RA whereas **B** shows a decreased median expression of perforin in the MAIT cell population in early RA. **C** Median expression of CD27 and CD28 are increased in the memory CD4 T cell population of early RA and **D** CD27 is increased in early RA central memory CD8 T cells. **E** Median expression of CD27 is decreased in the early RA neutrophil population although in **F** CD38 is increased in the basophils of early RA patients. **G** Both CD27 and CD86 have a decreased expression in NK cells of early RA patients. **H** CD14<sup>+</sup> monocytes have an increased median expression of CD38 in early RA but decreased median expression of CD27 and perforin.

7 of the 15 functional protein markers included for analysis were found to be differentially expressed in early RA compared to health but did not correlate with patient demographic data including age, gender, serology, tender/swollen joint count or smoking history suggesting that the immunological differences observed between early RA and health were driven by an overall immune dysregulation in early RA. As demonstrated in the MDS plot analysis, no significant immunophenotypic differences between early RA and established RA could be detected (data not shown). Interestingly however, the MDS plot had shown that established RA patients were distinct compared to 13 healthy donors (Figure 5-5) and although not statistically significant, the percentage of MAIT cells were decreased in 22 established RA donors compared to health (Figure 5-16) suggesting that immune differences observed in early RA (untreated) are still detectable in patients who had received biologics. The established RA patients included in the analysis were not discriminated by amount/type of biologics the patients were receiving as the analysis aimed to capture an overall snapshot of the treated stage of RA. Differential expression analysis showed similar trends to those observed in early RA, including increased median expression of both CD27 and CD28 in memory CD4 T cells, decreased median expression of perforin in MAIT cells (Figure 5-17). In addition, it was observed in the B cell population, PDL1 median expression was decreased whereas in contrast, both CD38 and HLA-DR median expression was increased in established RA compared to health. These subtle changes in protein expression suggest that there are detectable differences at the protein level between established RA and health.



**Figure 5-16 | Percentage of MAIT cells decrease in established RA (n=23) (depicted in orange) compared to health (n=13).**



**Figure 5-17 | Differential expression testing comparing established RA with health identifies trends in functional marker expression in B cells, memory CD4 T cells and MAIT cells.**

Differential expression testing using the Diffcyt pipeline did not reveal significant differences between established RA (n = 23) and healthy (n = 13) however, differences could be detected within B, memory CD4 T and MAIT cell populations. Within the B cell population, PDL1 median expression was slightly decreased in established RA compared to health. In memory CD4 T cells, median expression of CD27 and CD28 were increased in established RA compared to health and in MAIT cells, perforin expression was slightly decreased in established RA compared to health.

## 5.3 Discussion

### 5.3.1 Strengths and limitations of study design and analysis through Diffcyt pipeline

A significant amount of time was invested in validating the Diffcyt analysis pipeline as this was both a new bioinformatics skill coupled with high-throughput data. The ability to use scripts to assist in the development of analysing cytometry data is becoming more prevalent with many options for automated becoming available for discovery analysis<sup>166,291–293</sup>. As complexity of datasets increase due to patient demographics, multi-parametric experiments, and emerging bioinformatics softwares to analyse data continue to evolve, a robust analysis approach is required. This is a complex decision however, it is worth dedicating careful thought as to what questions/hypothesis are being asked of the dataset. This should be carefully considered before commencement of sample collection, something which may in theory sound obvious and straight-forward but may encounter logistical challenges in practice. Sample collection is both a fundamental yet arduous component of translational medicine research and well formulated hypotheses can be marred by practical consequences. To a certain extent, a degree of flexibility is required which means the initial hypotheses formed may need to be adapted during the course of the study. In a heterogeneous disease such as RA, evolving hypotheses can quickly become complex and confusing in terms of which variables are being measured. For example when formulating the hypothesis about treatment, this can very quickly become complicated in how to distinguish patients such as by type/duration/amount of treatment. If this question was to be analysed broadly, then any patient receiving treatment can be included for analysis however, if the question becomes more nuanced such as analysing patients who have only been treated with one biologic such as a TNF inhibitor then the number of patients suitable for that analysis will decrease. This study is no exception to suffering from sample size limitation (even though this was a larger cohort compared to previous mass cytometry reports) and it is recognised findings from this research would need to be further investigated in a larger cohort for confirmation. Despite the small cohort, it was not thought that this negatively impacted on the data interpretation and instead provided insightful analysis of the phenotypic profiles in the conditions measured.

The Diffcyt pipeline provided a fantastic opportunity to explore cytometric data analysis through an automated approach and develop essential bioinformatics skills. UCB Pharma was instrumental in providing the bioinformatics training and at the time of analysis developing pipelines and scripts which could be used to explore this dataset. The scope of analysis in this study was wide however, due to time restraints it was practical to begin with broad questions such as differences between health and RA that could identify obvious changes between the conditions. The MDS analysis plot was especially helpful to determine whether immune similarities could be detected within a condition and whether this in turn would infer a difference between conditions. MDS analysis showed a strong staining similarity within healthy donors and a strong similarity within patients with early RA and these separated into two distinct groups based on condition. It was observed that those conditions that formed distinct groups by MDS analysis were a robust prediction for whether there was likely to be a statistically significant difference in immune cell populations for this analysis. MDS analysis that did not show distinct separation of conditions were accurate as downstream of the pipeline, no statistically significant changes were identified by differential testing. Whilst statistically relevant changes underpin the relevance of the findings, and provide a persuasive

argument and quantified justification to focus on certain immune cell populations, it is recognised that this study was not large enough to provide this justification. Instead the findings by the Diffcyt workflow serve as a proof-of-concept that automated bioinformatics pipelines can be successfully applied to analyse routine cytometry experiments and provide an additional option to manual gating.

There are several limitations to the Diffcyt pipeline which were encountered during this analysis. A difficult decision occurs at the very beginning of the analysis - to downsample or not to downsample? The decision to downsample the event number for each sample to 27,000 events may be viewed as restrictive. It was noted in the F1000 publication in which Nowicka et al. describe their pipeline, the reviewer highlights the concern of whether a highly variable number of cells from different samples may influence the final clustering result<sup>176</sup>. In response, Nowicka et al. acknowledge that downsampling itself can have two main negative impacts on the data analysis, firstly the loss of data that occurs as a result of downsampling and secondly the chances of detecting rare cell populations become compromised<sup>176</sup>. Undoubtedly both viewpoints are valid and therefore careful consideration was applied when deciding what the best approach would be for this analysis. What ultimately underpinned the decision was that the data generated had to be reliable and as the number of events from each sample fluctuated, it was decided that downsampling would be the optimum approach. An advantage of downsampling was that it demonstrated the Diffcyt workflow was sensitive enough to find differences despite the small event number. This should provide reassurance for future studies in particular where cell number may be limited to begin with such as biopsy tissue, meaningful immunophenotypic changes can be detected. An unexpected disadvantage of downsampling was that it did not make much difference to run time. For each run conducted through the Diffcyt pipeline it took approximately 10 hours. The rate limiting step was easy to identify which was the FlowSOM clustering process and the annotating of cell populations. The fact that population location on a FlowSOM plot change with each run is both frustrating and time-consuming and consumed a significant amount of time. It is understood that this step has to be treated carefully as further analysis is based on the accurate identification of these cell populations and that this process cannot be entirely automated and require a second step to manually check these populations. It is anticipated however, that improved clustering tools continue to be developed and aim to streamline this process such as Phenotyping by Accelerated Refined Community-partitioning which can be used for large-scale, high-dimensional single-cell data in excess of 1 million cells<sup>294</sup>.

Due to a labelling error it was not possible to analyse the marker 'IgD' as it was incorrectly labelled as '1gd'. Whilst this error would not make a difference in softwares such as FlowJo, in scripts such as Diffcyt this error interferes with how the Excel file is read and is not recognised as it begins with a numerical digit as opposed to a letter. Whilst this did not negatively impact on the analysis as IgD expression can be analysed manually, it does mean it had to be excluded for automated analysis.

The Diffcyt workflow was useful when faced with the overwhelming task of where to begin with multi-parametric mass cytometry data. There is no optimum approach for analysis as yet nor a prescribed workflow although papers are now emerging from experienced research groups about considerations to take into account when approaching cytometry analysis<sup>164</sup>. What is especially encouraging is that these papers recommend a similar approach to the one described in this study giving confidence in the approach taken to analyse mass cytometry data.

### 5.3.2 Innate cell populations abundance are significantly decreased in early RA peripheral blood compared to health

The Diffcyt analysis revealed that MAIT, CD8 NK cells, pDCs and NK cell populations were significantly decreased in early RA compared to health. Analysis did not reveal a significant change in abundance in B and T cell populations or myeloid populations for early RA. It is worth considering the abundance of each significant immune cell population in context to its relevance to RA.

#### 5.3.2.1 MAIT cells are decreased in early RA

MAIT cells are innate-like T cells that recognise bacterial pathogens presented by MHC class-1 related protein 1 (MR1) and were characterised as CD3+ CD8+ CD161+ V $\alpha$ 7.2+ cells in this study consistent with existing literature reporting changes in this population<sup>295–299</sup>. MAIT cells have been observed to have an altered phenotype reported by Koppejan et al. in 2019 in early untreated RA stage in blood and SF and after bacterial stimulation had a diminished response in comparison to spondyloarthritis MAIT cells measured by a decrease in CD25 and CD69 expression<sup>300</sup>. Koppejan et al. identified MAIT cells using an MR1 tetramer which they suggest is an accurate method to detect MAIT cells as CD161 expression can fluctuate on CD4 and CD8 T cells however, they did not observe a change in the frequency of MAIT cells in the blood. Numerous studies have reported the decrease in frequency of MAIT cells in RA and other pathologies providing further evidence to support the analysis in this work<sup>295,296,298,299,301</sup>. Very few studies have observed CD4+ MAIT cells as they are not an abundant subset within the blood with the majority of MAIT cells identified as CD8+ and approximately 10% of MAIT cells are double negative CD4-CD8-<sup>302</sup>.

It should be mentioned that there are very few publications at present looking at the role of MAIT cells in early RA with many existing reports conducted in patients with established RA symptoms or receiving treatment so it remains uncertain whether MAIT cells have a reduced frequency in response to treatment or as a direct result of pathology. Moreover, MAIT cells are abundant at mucosal sites and also in the liver and lungs but the majority of studies including this study have observed their role in the blood where they account for 5% of CD3+ T cells<sup>303</sup>. This could suggest why in nearly all pathologies, blood MAIT cells are decreased in frequency despite the heterogeneity across the diseases<sup>299</sup>. In RA, high levels of TNF and IL-1 $\beta$  in SF which is thought to upregulate adhesion molecules encouraging the binding of MAIT cells to endothelial cells via sialyl-Lewis X motif. It was also observed that sialyl-Lewis X motif is expressed on fewer circulating MAIT cells in RA compared to health suggesting a migration of MAIT cells to SF<sup>299</sup>. To date, MAIT cells have not been observed to correlate with disease activity in RA but have done so in ankylosing spondylitis<sup>301</sup>.

Several reports have suggested that MAIT cells may be chronically activated in RA by either MR1 or by cytokines including IL-12, IL-18, IL-6 or interferon- $\alpha$  which can lead to an exhausted state represented by PD-1 or IFN- $\gamma$  and TNF secretion<sup>304</sup>. Once activated, MAIT cells are capable of killing target cells through the Granzyme B/perforin pathway<sup>304</sup>. In this analysis, an exhausted phenotype was not observed when assessing PD-1 expression levels on MAIT cells however, perforin expression was significantly decreased on MAIT cells in early RA suggesting that their cytolytic function is compromised. This is intriguing as it has not been previously reported in RA that MAIT cells have a reduced perforin function but it could be speculated that their ability to eliminate bacterially infected cells becomes compromised thus indicating an exhausted phenotype as their cytotoxic function declines. Future work will be required to determine the functional role of MAIT cells in RA

which should also include analysis of mucosal tissue to determine their tissue-specific immunophenotype.

### 5.3.2.2 Early RA plasmacytoid dendritic cells do not display an activated phenotype

pDCs are a distinct lineage of bone-marrow derived cells that reside mainly in blood and lymphoid organs in steady state and are functionally important for their rapid production of type I interferon in response to viral infection<sup>305</sup>. pDCs were a distinct population of cells, separating from conventional dendritic cells and pDCs were phenotypically characterised as CD45+CD4+CD45RA+CD45RO+HLA-DR+CD40+CD38+FCεRI+. The phenotype would also include CD123 however, this marker was excluded from the analysis.

The Diffcyt analysis revealed that pDCs were significantly reduced in peripheral blood of early RA in comparison to health which has been previously reported<sup>306–312</sup>. Observation of a reduction in pDCs have been reported since 2004 by Van Krinks et al. in spondyloarthritis and RA peripheral blood and SF<sup>308</sup>. It was observed that there were low numbers of dendritic cells in peripheral blood but higher levels in SF in both pathologies but particular higher in spondyloarthritis<sup>308</sup>. This led to the authors to hypothesise that in RA, pDCs are trafficked in their immature form to synovial tissue as opposed to SF as the synovium provides the optimum environment for pDCs to mature through the expression of CCR7 which are present on mature pDCs<sup>308</sup>. To support the claim that pDCs are present in RA synovial tissue, Cavanagh et al. observed pDCs in synovial tissue and noted their higher expression of HLA-DR (which was also significantly higher on early RA pDCs in this study) and their location in perivascular regions of synovial tissues more specifically in the blood vessels<sup>306</sup>. In RA, pDCs are hypothesised to be activated by a viral pathogen such as influenza virus or CpG oligonucleotides lead to the production of cytokines including TNF, IFN- $\alpha/\beta$  which are capable of inducing DC differentiation and stimulation of activation function<sup>306</sup>. Subsequent studies have also characterised the pDC population in the sublining of inflamed synovial tissue and observe that pDC number are specifically and significantly higher in comparison to myeloid DCs and were prolific producers of IFN- $\alpha/\beta$ <sup>309</sup>. A proposed model for pDCs trafficking from the blood to the SF and finally to the synovium is thought to occur by CD83- and/or DC-LAMP- pDCs and mDCs which release proinflammatory cytokines which perpetuate the chronic inflammation<sup>309,312</sup>.

It appears that the role of pDCs can be perplexing to distinguish based on the environment in which they either originate or are studied in vitro. In 2018, Cooles et al. set out to investigate the frequency and phenotype of pDCs in early RA blood and identify whether immunology related genes were upregulated or downregulated using transcriptomics<sup>311</sup>. Their findings revealed that peripheral blood pDCs have a tolerogenic phenotype which is in contrast to its function

Lastly, mass cytometry analysis revealed pDC numbers were decreased in Primary Sjögren's blood but did not correlate with numbers in labial salivary gland tissue<sup>307</sup>. Despite the reduction, pDCs in blood produced comparable levels of IFN to that of tissue suggesting that pathology may result by activation through immune complexes, production of interferon and a subsequent induction of a negative feedback loop which disrupts pDCs homeostasis rather than as a result of tissue migration<sup>307</sup>. This somewhat conflicting evidence is also apparent in RA, where Cooles et al. reported that pDCs in early RA blood do not have an activated phenotype but perhaps instead demonstrate a tolerogenic phenotype<sup>311</sup>. The analysis presented here supports this hypothesis as I did not identify pDCs to have an increased expression of activation markers compared to health. Transcriptomic analysis by Cooles et al. showed in early RA there was a downregulation of *TNFRSF17* and upregulation of *PRDM1* and *CSF1R* which is considered to be immunoregulatory in function. *PRDM1*

(also known as BLIMP-1) is increased in human pDCs in response to IFN- $\alpha$  and in turn is thought to suppress production of proinflammatory cytokines leading to a tolerogenic function<sup>311</sup>. Both disease stage and microenvironment are considered to be central factors which determine the outcome of pDC function. This offers an explanation as to why the data is contradictory depending on the location in which pDCs are observed for example in SF, pDCs are capable of activating T cells and production of cytokines<sup>311</sup>.

In summary, pDCs have been identified as a significant population in this analysis despite the modest number of events (27,000) from each sample and pDCs being a non-abundant population. It is hoped that this finding will provide justification to further explore the dendritic cell population with particular focus on pDCs to definitively establish their role in the periphery and synovium.

### 5.3.2.3 NK and CD8 NK cells in early RA

Diffcyt analysis revealed two populations of NK cells which were significantly decreased in early RA blood compared to health which has previously been identified<sup>313,314</sup>. From this finding, it is hypothesised that MAIT cells, pDCs and NK cells have a common function in responding and eliminating bacterial/viral infected cells and secreting proinflammatory cytokines mainly IFN and TNF. The next observation is that all three cell populations are decreased in early RA compared to health suggesting that there is a common mechanism trafficking these cell types to the joint where they become influenced by the local microenvironment.

The two populations of NK cells detected by FlowSOM clustering showed NK cells and CD8 NK cells cluster separately but in close proximity to one another which indicated that phenotypically these were two different populations. The marker CD56 was excluded from analysis which would have been used in conjunction with CD16 to initially characterise this population. However, FlowSOM was capable of phenotyping NK cells using a combination of other markers to define NK cell subsets. NK cells in this analysis were defined as CD45+CD11b+CD16+CD45RO+CD11c+CD38+NKp46+ and CD8 NK cells were defined as CD45+CD3+CD8+CD11b+CD45RA+CD16+CD38+. It was uncertain whether CD8 NK cells. Without functional analysis of CD8 NK cells, it is difficult to state whether this cell population is similar to NKT cells although from just phenotypic analysis it does suggest that CD8 NK cells are similar in phenotype to these cells<sup>315</sup>. Another study showed that activated NK cells were observed to be decreased in RA peripheral blood but the frequency was not altered in resting NK cells however, the patient cohort did not look at classifying patients according to treatment naïve or treated patients<sup>316</sup>.

Studies using the CIA mouse have shown that NK cells are capable of causing joint destruction and are abundantly present in the RA joint<sup>317</sup>. In this analysis, both CD27 and CD86 expression levels were significantly decreased in early RA patients indicating an altered activated function. However, in the next chapter, I show that CD8 NK cells display a significantly altered phenotype which saw increased expression of inhibitory proteins CD80, NKp44 and PD-L1 and also suggestions of an activated phenotype including increased expression of CD27, CD28 and HLA-DR but an impairment in cytotoxic function demonstrated by a significant decrease in perforin. It is therefore hypothesised that the upregulation of inhibitory receptors results in deficient cytotoxic function and could be a possible mechanism which contributes to early RA pathogenesis. This altered phenotype in early RA has been observed in patients with seropositive RA where NK cells were decreased and had a reduced propensity for IFN- $\gamma$  production<sup>314</sup>. This decrease in numbers was noticeable only in seropositive patients and not in seronegative patients could be explained by apoptosis induced through FC $\gamma$ R triggered by IgG immune complexes<sup>314</sup>.

The role of NK cells in RA is not fully understood and furthermore it is still uncertain whether they actively contribute to pathogenesis or whether they are capable of modulating homeostasis<sup>318</sup>. A study using a CIA animal model, antibody-mediated NK cell depletion led to an aggravation of disease but conversely, another study showed depletion of NK cells suggested a resolving of CIA and prevented counterwork of bone loss upon NK cell depletion<sup>319,320</sup>. Data from this study shows CD27 and CD86 are decreased on NK cells, suggesting a deactivated phenotype whereas NK cells present in the inflamed synovium tend to have an activated phenotype<sup>317</sup>. Therefore it is considered that CD8 NK cells may be closer to the NK cells found in the inflamed synovium as they have an activated phenotype and upregulate inhibitory receptors suggesting a pathological involvement in RA. It is evident that there is not enough evidence as yet from this work or other studies as to the specific phenotype of the pathologic NK cells however, I predict that CD8 NK cells would produce more TNF and IFN proinflammatory cytokines and induce the maturation of DCs<sup>321</sup>.

Finally, NK cells are thought to possess the ability to regulate autoimmune disorders by killing autoreactive immune cells. Although this study did not reveal a reduction in perforin or granzyme B cytotoxic molecules, it has been shown by another study that cytotoxic activity decreases in RA compared to health<sup>322</sup> and another study showed that perforin-positive NK cells was reduced in patients with RA<sup>323</sup>. This presents an additional dimension to NK cells in terms of their role and whether their function is impaired as a consequence of pathology or whether their functional impairment precedes the onset of RA. Answers to these questions and verification of these findings will help to further elucidate the role of NK cells in RA.

#### **5.3.2.4 Neutrophils interact with innate immune cell populations in early RA**

Analysis in this study showed that neutrophil abundance did not significantly change between early RA and health and a striking phenotype could not be identified despite a significant decrease in CD27 in early RA peripheral blood. Despite the lack of specific phenotypes it is evident that neutrophils have an influential role on the immune system which goes beyond their ability to respond and eliminate pathogenic stimuli. Analysis depicted by the MDS plot was particularly informative as it suggested that functionally neutrophils were markedly distinct in early RA compared to health. The neutrophil population separated early RA from health and upon absence of neutrophils, this distinction between the two conditions became less prominent. This suggests that neutrophils are fundamentally altered in early RA compared to health supporting the concept of neutrophil heterogeneity and neutrophil plasticity<sup>324</sup>. Heterogeneity of neutrophil subsets have been well characterised in autoimmune diseases including SLE where low-density neutrophils were identified based on their high expression of CD15 and low expression of CD14 and their ability to synthesise type I interferons<sup>325</sup>. In addition, specific subsets of neutrophils have been characterised to have immunosuppressive functions where for example they can prevent the proliferation of T cells through macrophage-1 antigen and reactive oxygen species<sup>326</sup>.

It is the neutrophils ability to engage in cross-talk across the immunological spectrum that is especially fascinating and may provide an explanation for the data observed in this study. Evidence of this cross-talk is observed across numerous immune cell populations including DCs, NK cells and MAIT cells<sup>324</sup>. Neutrophils are capable of both positive and negative modulation of DC subsets including monocyte-derived subsets and pDCs<sup>327</sup>. Neutrophils can communicate with DCs through contact-dependent mechanisms or through the secretion of cytokines<sup>327</sup>. In addition to these interactions, neutrophil extracellular traps (NETs) have provided another option for communication between neutrophils and DCs and this has been studied in type I diabetes and SLE<sup>328,329</sup>. In a study

looking at the effects of NETs on patients with type 1 diabetes, it was observed that the composition of NETs was different compared to healthy donors<sup>328</sup>. In a mixed peripheral blood mononuclear cell culture, type I diabetes NETs caused a strong shift towards IFN- $\gamma$ -producing T lymphocytes<sup>328</sup>. Further in vitro investigation using monocyte-derived DC cultures, demonstrated that type I diabetes NETs induced cytokine production, phenotypic change and IFN- $\gamma$  production with RNA-sequencing further revealing downregulation of TGF $\beta$  and upregulation of IFN- $\alpha$ , thus skewing healthy monocyte-derived DC cultures towards a type I diabetes phenotype<sup>328</sup>. In a separate study, neutrophil apoptotic microparticles were shown to increase the expression of CD40, CD80, CD83 and CD86 and the production of proinflammatory cytokines including IL-6, TNF and IFN- $\alpha$  by blood derived pDCs and mDCs from patients with SLE<sup>329</sup>. Microparticles derived from SLE plasma were also capable of priming blood derived neutrophils for NETosis however, microparticles derived from RA did not show similar results<sup>329</sup>. Taken together, these studies demonstrate that neutrophils may have an immunomodulatory effect on DCs affecting their phenotype and effector functions and further investigation into this interaction in the RA joint may demonstrate a mechanism in which neutrophils perpetuate a proinflammatory microenvironment.

There is evidence to suggest that neutrophils and NK cells interact with one another with data suggesting that activated NK cells can increase the longevity of neutrophils but can conversely induce neutrophil apoptosis as evidenced by in vitro cocultures<sup>330</sup>. In contrast, neutrophils have been shown to immunomodulate NK cell survival and cytotoxic function in addition to being pivotal for NK cell development in human and mice<sup>331</sup>.

Lastly, in a recent study looking at the effect of neutrophils on MAIT cell activation and the effect of activated MAIT cells on neutrophils during bacterial infection<sup>332</sup>. It was observed that neutrophils suppress the activation of MAIT cells by a cell-contact and H<sub>2</sub>O<sub>2</sub> dependent mechanism and in turn highly activated MAIT cells were able to produce TNF that induced neutrophil apoptosis<sup>332</sup>. This study suggests that neutrophils and MAIT cells are capable of negatively regulating each other in order to control inflammation<sup>332</sup>.

It is apparent from the findings within this study and other studies, that neutrophils play an integral role in orchestrating inflammation and capable of working synergistically with other immune cell populations. During the early, untreated phase of RA, it is apparent that neutrophils could be regarded as master regulators of inflammation. It will be important to further investigate the relationship of neutrophils with other immune cell populations within the RA joint to define the mechanism/s perpetuating the underlying chronic inflammation.

### **5.3.3 Activation markers – a therapeutic target for RA?**

A recurring theme observed in differential expression analysis across immune cell populations was the fluctuating expression of CD27 and CD38 and to a lesser extent CD28. These markers are often associated with activation function of cells and as can be observed in this study are expressed on a broad selection of immune cell populations. On memory CD4 T cells, both CD27 and CD28 expression was elevated and CD27 expression was elevated on memory CD8 T cells in early RA patients. This is in contrast to the chronic activated phenotype that is observed in the joint where T cells start to lose their CD27 and CD28 expression over time thought to be as a result of persistent antigenic stimulation result in T cells characterised by the loss of CD27 and CD28<sup>333</sup>. However, in RA blood this phenomenon is reversed suggesting that during the early phase of RA, T cells are activated. Whilst knowledge surrounding the inflammatory cascade that occurs in established RA is well documented, the inflammatory events in early RA are less well understood. Thompson et al. assessed T cells in 25 patients with early RA and observed that CD8+ CD28- T cells percentage was

increased in patients who were also positive for cytomegalovirus and positively associated with CRP however, this was not observed in established RA patients<sup>334</sup>. Although this finding differs from that presented in this work, it is known that T cells transition gradually from CD28+ to a CD28- T cells phenotype where they become chronically activated<sup>335</sup>. CD27 expression was reduced on CD14+ monocytes in early RA compared to health and similarly perforin expression was also reduced again suggestive of a chronically activated phenotype. Lastly, CD38 expression is increased on basophils and CD14+ monocytes. CD38 is another marker associated with activation and involved in cell adhesion and signal transduction and associated with expression by plasma and B cells however, low expression levels have also been reported in additional cell types including lymphoid, myeloid and non-haematopoietic cells<sup>336</sup>. The results from this analysis coupled with existing reports have prompted an interest into whether activation markers would be an appropriate treatment target in RA due to their ubiquitous expression on a variety of cell populations and in particular during chronic inflammation<sup>336-339</sup>. Cole et al. in 2018 observed through both RNA-sequencing analysis of synovial biopsies obtained from different stages of RA disease progression, that CD38 was one of the genes significantly upregulated at all stages of RA including early RA compared to health and osteoarthritis<sup>337</sup>. Moreover, through in vitro cell culture they demonstrated that daratumumab, a monoclonal inhibitor of CD38, plasma cells were successfully depleted in PBMCs from patients with early RA in a dose dependent manner ex vivo. These findings encourage further clinical evaluation for daratumumab to be used as therapeutic target for RA. Existing clinical trial studies multiple myeloma have demonstrated that CD38 is a good immunotherapeutic target in these patients as both a monotherapy and in conjunction with chemotherapy with clinical trials currently underway<sup>338</sup>.

Existing treatments used in RA have also been analysed to see if they can target chronically activated T cells in established RA where patients tend to display an expanded CD4+ CD28- T cell phenotype which is discussed further in the chapter 6. It is worth briefly mentioning however, that TNF inhibition has been shown to limit the expansion of CD4+CD28- T cells in unstable angina<sup>340</sup>. In RA infliximab reinstated the expression of CD28 on T cells and methotrexate demonstrated a reduction in CD28- T cells suggesting that therapeutic targeting of activation markers may already be possible with existing treatment but not well understood mechanistically how this works<sup>341</sup>.

Finally, it will be important to comprehensively understand whether inhibiting activation markers expressed on a broad range of cells will be helpful without posing a risk to the patient. Would inhibition lead to the presence of malignant cells or a worsening of disease outcomes. How sustainable is it to target activation markers and to what degree? From this study, it would appear early inhibition of activation markers such as CD38, CD27 and CD28 might impede the progression of RA thus slowing joint damage. It seems logical that a broad target is initially used in early RA and it might be that existing treatments may already exist. Rather than inhibition would removing over-activated cells be more impactful and if so which therapies could induce specific apoptosis of these cells populations and which protein marker would be the best target? These concepts will need to be explored in greater detail however, provide an opportunity for novel therapeutic targeting in RA.

### **5.3.4 Significant changes in whole blood immune cell populations were not identified in established RA**

Diffcyt analysis did not reveal statistically significant changes when analysing patients with established RA despite the various comparisons that were performed. Initially, the analysis aimed to broadly group established RA irrespective of treatment, duration of disease or any other demographic or clinical factor to observe if there was an overall change in immune cell populations compared to early RA or health. MDS analysis did reveal that established RA patients did form a distinct group from health suggesting that underlying immune changes are driving this separation

but are not significant when comparing health to early RA. Early RA can be viewed as an authentic insight into chronic inflammatory events that occur prior to clinical intervention (i.e. treatment). A question to consider is which stage of RA is immunologically closest to health? This question is important as it questions how far from immune homeostasis each stage of RA represents and if we how best to intervene. The data from this analysis supports the view that early RA shows hallmarks of chronic activation as evidenced by the expression of CD27, CD28 and CD38 markers across numerous immune cell populations when compared to health. The fact that there is little difference between early RA and established RA when analysing blood immune cell populations suggests that immune dysregulation is comparable not distinct. Therefore to answer the question about which RA stage is further from health, it is difficult to confidently state which stage is furthest from health. And herein lies the problem – is blood a good immunological representative of RA?

When analysing the established RA cohort, I aimed to observe whether a specific immune cell signature could distinguish non-refractory patients from refractory patients using the current definition which suggests refractory patients have failed two or more mechanistically different biologics and non-refractory defined as patients receiving 1 or 2 mechanistically similar biologics <sup>290</sup>. However, this analysis did not show any specific immune cell populations that could predict or distinguish between non-refractory and refractory patients suggesting that resting blood is not a robust indication of discrepancies within the immune populations. Blood is often used in RA translational research studies as it is easier to obtain and because RA is a systemic disease, it is often thought that blood is a good proxy for reflecting the inflammatory events occurring at the joint. In reality it is recognised that any findings within the blood are an extrapolation of events in the joint and why there is currently no specific immune cell signature that can serve as a biomarker to define RA stage or treatment response. It may also go some way to explaining why treatment does not work in all patients as current treatments in RA focus on immune cells predominantly B and T cells or inhibition of cytokines and recently small inhibitors which aim to target signalling pathways upstream of cytokine production. It is apparent there are significant omissions in therapeutic targets including biologics which target innate immune cells and specifically cells that span both innate and adaptive immune functions including those of myeloid origin, NK cells and MAIT cells as identified in this study.

Another consideration to make is that immune dysregulation may not be the only factor driving disease and the case for analysing non-immune cell populations including fibroblasts becomes more apparent. This will require access to synovial tissue to understand their role and already specific fibroblast subsets have been associated with RA <sup>284–286,342</sup>. Whether fibroblasts can provide the distinction between refractory and non-refractory patients is an interesting hypothesis to pursue as it provides a logical alternative to understanding why certain patients do not respond to existing treatments. Clinical trials have already commenced to determine the optimum dose of Seliciclib which suppresses proliferation of synovial fibroblasts <sup>289</sup>.

In summary, to comprehensively interrogate the progression of patients with established RA, this study supports the current efforts directed towards synovial biopsy immunophenotyping to determine if there are specific immune signatures which associate with refractory patients and to better characterise stages of RA as evidenced by Pitzalis et al. when they reported specific RA clinical phenotypes (lympho-myeloid, diffuse-myeloid and pauci-immune) <sup>170</sup>. This classification does not include innate immune cell populations or immune cells such as NK and MAIT cells which span both innate and adaptive immunity. Blood is a useful biomarker for early RA and although logistically

easier to obtain, may not hold valuable information as RA progresses within the joint and hence efforts have returned to the synovial joint.

## Chapter 6 Immunophenotyping immune cell populations in paired RA synovial fluid and peripheral blood

### 6.1 Introduction

#### 6.1.1 Synovial fluid in joint pathology

The central site of RA pathology is at the synovial joint, where the focus of RA research has returned as a result of improved, minimally invasive techniques for sampling synovial biopsies in small and large joints guided by ultrasound<sup>58,343,344</sup>. However, despite improved techniques, acquiring synovial biopsies remains a surgical procedure which requires a clinician trained in obtaining high quality biopsies which capture the inflammation occurring at the joint<sup>344</sup>. Moreover, a dedicated biopsy unit needs to be organised to facilitate access to synovial biopsies for use in research which is not always logistically possible as was the case for this study. To address this, synovial fluid (SF) provides an opportunity to interrogate the immune microenvironment of the joint. Pathological joint fluid is readily obtainable in Rheumatology departments especially during arthrocentesis (also known as joint aspiration) from patients who have excess fluid in their joints. Furthermore, SF is in direct contact with the synovium and is secreted by a layer of connective tissue that lines the cavities of the joints.

In health, SF is essential for maintaining joint homeostasis and present in small volumes, approximately 1-4mL in the knee joint<sup>345</sup>. In the steady state, SF is typically acellular. The main functional properties of SF provide is to lubricate the joint by the secretion of large macromolecules, namely lubricin and hyaluronan<sup>346</sup>. These molecules give rise to the viscous consistency of SF and contribute to the clear/pale yellow appearance. SF is also important for maintaining metabolic functions which include the transport of nutrients, waste products, and other metabolites to and from synovial tissues<sup>346</sup>.

In joint pathology, SF is significantly altered. Basic observations of pathological SF show it is cloudy in appearance and either a more viscous or thinner texture depending on the consistency of hyaluronan<sup>347</sup>. Furthermore SF can be useful for identifying crystals such as monosodium urate monohydrate and calcium pyrophosphate dihydrate which can be useful for diagnosing gout or pseudogout<sup>348</sup>. The most striking observation is the increase in volume of SF within the joints of patients with inflammatory arthritides including RA and psoriatic arthritis. For the patient, excess SF can cause immense discomfort and pain, which is alleviated when aspirated. This increase in volume can be directly attributed to an increase in cellular infiltrate including lymphocytes, macrophages, dendritic cells and neutrophils. It is precisely this cellular infiltrate that is of research interest, as understanding whether there is a specific cell population which is implicated in recruiting cells to the joint and in turn whether this is responsible for perpetuating disease. In the next section, recent research using single cell technology has revealed specific immune cell populations identified in RA SF which could serve as future therapeutic targets.

#### 6.1.2 Immune cell populations detected in RA SF

##### 6.1.2.1 Neutrophils

Neutrophils are the most abundant cell populations found in RA SF <sup>349</sup>. In health, circulating neutrophils patrol the immune system, and upon injury or encountering a pathogen are capable of phagocytosis, and releasing cytotoxic granules to eliminate the pathogen <sup>350</sup>. It is this functional role that has characterised neutrophils as having a short life span as once the neutrophil has encountered and eliminated a pathogen, it undergoes apoptosis within 24 hours <sup>351</sup>. Initially it was considered that neutrophils in the joint were recruited for their cytotoxic ability and although they retain this function, neutrophils are observed to be altered in disease compared to health producing an array of proinflammatory cytokines and chemokines including GM-CSF, IL-6, TNF, IL-8 and IFN- $\gamma$  which contributes to their enhanced survival in the joint <sup>352</sup>. This proinflammatory environment primes other immune cells, thus perpetuating chronic inflammation in the joint. Immune complexes in the joint such as RF are observed to bind to FCyRs on neutrophils resulting in a large number of activated neutrophils detected in SF. The functional implications of neutrophils in RA has led to renewed interest and this is documented in detail in two excellent reviews <sup>353,354</sup>.

This renewed interest has encouraged researchers to interrogate neutrophils at the single cell level including a recent mass cytometry study where researchers immunophenotyped peripheral blood from 9 patients with RA and 5 healthy donors and identified a novel neutrophil subpopulation characterised by the phenotype CD11b<sup>low</sup> and CD16<sup>high</sup> which positively correlated with disease duration but not with disease severity measured by DAS28 <sup>355</sup>. Another recent study looked at transcriptomic profiling of a single cell neutrophil population isolated from paired RA blood and SF from three patients with active disease <sup>356</sup>. Comprehensive profiling of the neutrophil transcriptome showed 772 genes were significantly different between peripheral blood and SF and revealed pathways associated with antigen presentation, role of NFAT in the regulation of the immune response and acute phase response were all upregulated <sup>356</sup>. Chemokines IL-8, CXCL1 and CXCL2 were also upregulated in RA SF which are associated with recruitment of cells to the joint including T cells, monocytes, macrophages, dendritic cells and NK cells in addition to neutrophils. To study the functional role of neutrophils in vitro, the researchers isolated healthy neutrophils from peripheral blood and subsequently cultured with RA SF demonstrated that RA SF decreased the rate of neutrophil apoptosis <sup>356</sup>. This was similarly reported in 2006 where Raza et al. observed that early RA SF had lower levels of neutrophil apoptosis than patients with non-RA arthritis or resolved RA <sup>357</sup>. The presence of anti-apoptotic cytokines identified in synovial joint including IL-2, IL-4, IL-15, GM-CSF and G-CSF was proposed as the mechanism that prevented neutrophils and other leukocytes from undergoing apoptosis <sup>357</sup>. As this phenomenon was detected in patients diagnosed with early RA for  $\leq 3$  months, it suggests that this priming of the joint microenvironment contributes to persistent RA <sup>357</sup>.

### 6.1.2.2 CD4<sup>+</sup> T cell subsets

T cells have been extensively studied in the joint microenvironment and specific subpopulations have been identified in SF mainly by cytokine expression studies. It is hypothesised by many of these studies that specific cytokine expression profiles within a subpopulation of cells can accurately correlate and thus stratify arthritis. In 2017, Penatti et al. set out to distinguish between inflammatory osteoarthritis (OA) and RA by assessing CD4 T cell subsets in SF, peripheral blood and synovial membrane <sup>358</sup>. Assessment of RA SF showed an increase in the number of immune cells and a higher CD4:CD8 ratio compared to OA SF <sup>358</sup>. T regulatory type 1 and CD25<sup>+</sup> T regulatory cells were also observed to be enriched in RA SF. In addition, a high number of CCR6<sup>+</sup> IL-7R<sup>+</sup> Th17 cells and CXCR5<sup>+</sup> ICOS<sup>+</sup> T follicular helper (Tfh) cells were detected in many but not all RA SF. Tfh cells are a well characterised population of T helper B cells and support antibody production. During an

inflammatory response, expression of the chemokine receptor CXCR5, informs Tfh cells to migrate to the follicles within secondary lymphoid organs including the lymph node and spleen to help B cells<sup>359</sup>. This migration of Tfh cells is tightly regulated in health, however, in RA, this process becomes dysregulated as the anatomic distinction between the synovium and secondary lymphoid organs merges. The lymphoid aggregates that develop in the synovium are diverse in cellular composition, with histological analysis ranging from sparse lymphocytic clusters to fully developed organised follicles with germinal centres<sup>159</sup>. Analysis of the RA synovium reveals the presence of activated B cells which would logically suggest that Tfh cells would be detected in the synovium. However, mass cytometry analysis of synovial biopsy tissue from patients with RA discovered a pathologically expanded subset Tph cells with a phenotype of PD-1<sup>hi</sup> CXCR5<sup>-</sup> CD4<sup>+</sup><sup>157</sup>. Additional single cell analysis by flow cytometry with three paired RA SF and peripheral blood from seropositive patients revealed a large population of CD4<sup>+</sup> PD-1<sup>+</sup> T cells which co-expressed MHC II and accounted for 30% of CD4<sup>+</sup> T cells which was similar to that observed in the joint<sup>157</sup>. CD4<sup>+</sup> T cells with a similar phenotype to Tph cells have been previously reported to be enriched in RA synovium in comparison to OA but lacked detailed functional analysis<sup>283,360,361</sup>.

A defining marker of Tfh cells is that they are CXCR5<sup>+</sup>, a characteristic that is not observed on Tph cells. However, Tph cells appear to retain their function as T cells that help B cells as evidenced by the B cell enabling factors identified in SF. Tph cells isolated from RA SF showed a 100 fold increase in mRNA expression of IL-21 and 1000 fold increase of CXCL13<sup>157</sup>. IL-21 is a chemokine which promotes B cell proliferation in germinal centres and differentiation into plasma cells and CXCL13 is the chemokine ligand for the receptor CXCR5<sup>159</sup>. CXCL13 and CXCR5 work together to coordinate the organization of B cells within the follicles of lymphoid tissue. Tph cells sorted from RA SF provided further evidence that Tph cells were capable of secreting IL-21 and CXCL13 after in vitro activation. Another characteristic feature of Tfh cells is the upregulation of transcription factor Bcl6 which was observed to be downregulated in Tph cells derived from RA SF. Instead Blimp1 which is a known antagonist of Bcl6 and capable of directly inhibiting Bcl6 expression in B cells, was elevated in Tph cells further underscoring the differences between Tfh and Tph cells<sup>157,159,362</sup>. Rao et al. make a persuasive case to suggest that Tph cells are expanded in RA SF and tissue and thus have a direct association with RA prognosis<sup>157</sup>. To strengthen this observation, Tph cells increased in patients with seropositive RA who also had moderate to high disease activity and subsequently decreased upon a reduction in disease activity thus providing a future therapeutic target for RA<sup>157</sup>.

Better understanding of CD4<sup>+</sup> T cells and their subsets continues to be explored in RA and determine to what extent these cells are pathologically implicated with disease onset and progression. Whilst renewed interest has been piqued by the identification of Tph cells, other CD4<sup>+</sup> T cell subsets have been identified in RA SF including Tregs<sup>363-366</sup>. Immunophenotyping Tregs and having a clear distinction between the subsets has been extensively debated by established researchers in the field and in many of the flow cytometry immunophenotyping studies performed prior to 2010, Tregs have been characterised as CD4<sup>+</sup> CD25<sup>+</sup>. It has been observed that T regulatory cells are present at a significantly higher number in RA SF compared to peripheral blood obtained from the same patient, which at first appears logical given the inextricable link between Tregs and their role in maintaining tolerance to self antigens and thus suppressing induction and proliferation of effector T cells<sup>366,367</sup>. However, on closer inspection, elucidating the exact role of Tregs in the RA joint has proven to be ambiguous as despite their presence, inflammation continues to persist. Evidence showed that Tregs in the joint displayed an activated phenotype including upregulation of CD69, OX-40 and MHC II<sup>366</sup>. Although CD4<sup>+</sup>CD25<sup>+</sup> T cells from RA SF showed increased suppressive activity compared to CD4<sup>+</sup>CD25<sup>+</sup> Tregs from peripheral blood, this suppressive activity was not enough to subdue

activated responder T cells <sup>366</sup>. It is not well understood whether CD4<sup>+</sup> CD25<sup>+</sup> Treg cells with an immunosuppressive phenotype are lost during the progression of RA or whether their functional role becomes dysregulated in the joint microenvironment and thus induce/perpetuate progression of RA <sup>366,91,368</sup>. Whilst naturally occurring Tregs are unable to inhibit the proinflammatory cytokine production present in the synovium, an elegant study by Beavis et al. indicated that ectopic expression of FoxP3 in human effector T cells derived from the synovium, were capable of converting them to a Treg phenotype and attenuate their pathological effector functions providing an alternative therapeutic approach <sup>369</sup>.

An enriched population of CD4<sup>+</sup> HLA-DR<sup>+</sup>CD27<sup>-</sup> cytotoxic effector memory T cell subset were identified by mass cytometry and comprised 10% of the CD4<sup>+</sup> T cell population in the synovium <sup>261</sup>. Moreover, RA SF contained 5 fold higher frequencies of CD4<sup>+</sup>HLA-DR<sup>+</sup>CD27<sup>-</sup> T cells and gene expression analysis revealed an increase in CXCR3, CCR5, TBX21 and IFN- $\gamma$  suggestive of a Th1 skewed phenotype <sup>261</sup>. CD4<sup>+</sup>HLA-DR<sup>+</sup>CD27<sup>-</sup> T cells also displayed a cytolytic function observed both at the transcriptomic and intracellular flow cytometry protein analysis by the increased expression of perforin, granzymes A and B. This cell population was reported by the same group who reported the expansion of Tph cells in the RA synovium however, the CD4<sup>+</sup>HLA-DR<sup>+</sup>CD27<sup>-</sup> T cell subset are a distinct population. Similar to Tph cells, CD4<sup>+</sup>HLA-DR<sup>+</sup>CD27<sup>-</sup> T cells are not exhausted despite displaying features of chronic activation evidenced by loss of CD27 and upregulation of HLA-DR. Patients commencing a new treatment were observed to have a decreased frequency of CD4<sup>+</sup>HLA-DR<sup>+</sup>CD27<sup>-</sup> T cells and it is hypothesized that successful response to treatment reduces the expansion of the CD4<sup>+</sup>HLA-DR<sup>+</sup>CD27<sup>-</sup> phenotype <sup>261</sup>.

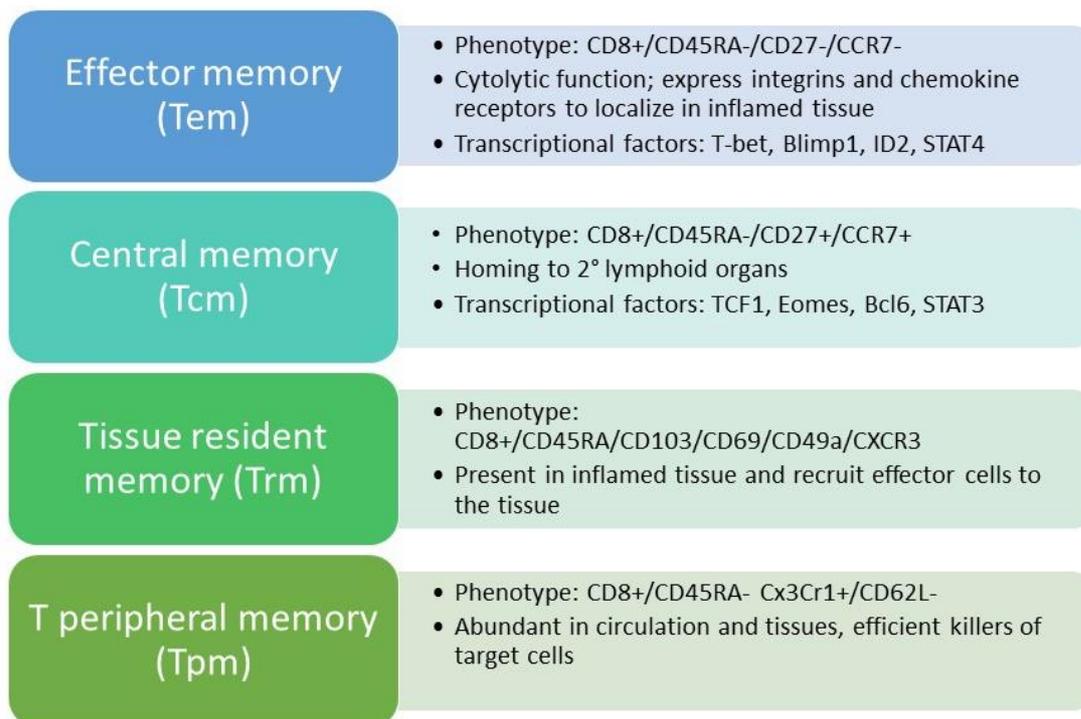
### 6.1.2.3 Memory CD8<sup>+</sup> T cell subsets

Whilst considerable attention has been directed towards the role of CD4<sup>+</sup> T cells due to implication of the HLA-DR locus observed from genetic studies, CD8<sup>+</sup> memory T cells are beginning to receive recognition for their role in pathogenesis of RA. The subsetting of memory T cells into their effector and central memory phenotypes and thus identifying their role in pathology is gaining momentum

<sup>370</sup>. The basic definition of a memory CD8 T cell is that it has already responded to cognate antigen and persists long term as part of the immune defence. Compared to naïve CD8 T cells of the same antigen specificity, memory CD8 T cells are greater in number, have higher proliferation capacity, populate peripheral organs, perform cytotoxic functions and secrete effector cytokines when re-encountering antigens and specific subsets have been functionally characterised by immunophenotyping and genetic studies <sup>370</sup>. Historically, observing memory CD8 T cells in humans has been performed by peripheral blood as opposed to tissue which presents a challenge when determining the age of memory CD8 T cells after antigen encounter. Furthermore, memory CD8 T cells have typically been analysed as a bulk population which can overlook subtle differences within the population. Figure 6-1 outlines the current defined subsets of memory CD8 T cells and their function informed by Martin and Badovinac's review focusing on heterogeneity of memory CD8 T cells <sup>370</sup>. This continues to be an active area of research as more immunophenotyping studies emerge refining the existing definitions.

Immunophenotyping RA SFMCs has also revealed effector CD8<sup>+</sup> T cells which resemble tissue resident memory phenotypes including CD69<sup>+</sup> CD103<sup>+</sup>CD45RA<sup>-</sup> CD8<sup>+</sup> T cells identified by flow cytometry <sup>371</sup>. These cells exhibited a cytotoxic phenotype and in RA SF, also expressed high levels of PD-1, CD101 and Eomes which are characteristic of a tissue resident memory phenotype which was

similarly observed in juvenile idiopathic arthritis <sup>371,372</sup>. Perforin expression in CD69<sup>+</sup> CD103<sup>+</sup>CD45RA<sup>-</sup> CD8<sup>+</sup> T cells was higher in ACPA positive patients compared to those who were ACPA negative and stimulation with the cytokine IL-15, induced perforin-mediated histone citrullination and encouraged neutrophil extracellular traps formation <sup>371</sup>. Intriguingly, Cho et al also identified an effector CD8<sup>+</sup> CD45RA<sup>-</sup> T cell population however, these cells showed reduced levels of perforin and granzyme B but did upregulate PD-1, CD80 and CD86 <sup>373</sup>. This CD8<sup>+</sup> CD45RA<sup>-</sup> T cell subset showed an activated phenotype in SF by the presence of CD27 and CD28 thought to be mediated by the cytokine milieu namely IL-15 and IL-21 <sup>373</sup>. The activated phenotype of CD8<sup>+</sup>CD45RO<sup>+</sup> is further confirmed in a separate study that compared CD8 T cell phenotypes between paired RA peripheral blood and synovial fluid by observing upregulation of CD69 and CXCR4, suggesting the latter is involved in homing T cells to sites of inflammation <sup>374</sup>.



**Figure 6-1 | Heterogeneity of memory CD8 T cells outlining their phenotypic characterisation, function and where known, transcription factor profile <sup>370</sup>**

#### 6.1.2.4 NK cells and rare immune cell populations detected in RA SF

Neutrophils and T cell subsets are the predominant immune cell populations found in RA SF but additional immune subsets have been reported. Paired analysis of peripheral blood and synovial fluid identified subsets of NK cells using the markers CD3<sup>-</sup>, CD56 and CD16 to define subsets. Synovial fluid analysis from inflammatory arthritis patients revealed approximately 16% of lymphocytes within the compartment were CD56<sup>bright</sup> NK cells and that an increased percentage of NK cells correlated with high DAS <sup>317,375</sup>. Additional analysis suggested that NK cells present in SF were recruited from peripheral blood as evidenced by the expression of chemokine receptors CCR5 and CXCR3<sup>375</sup>, and in another study CXCR3 and CCR1 were reported to be upregulated on SF NK cells <sup>317</sup>. Functional analysis of the CD56<sup>bright</sup> SF NK cell population demonstrated they had an enhanced capacity to produce IFN- $\gamma$  in response to cytokines IL-12 and IL-15, which are both present in increased quantities in the joint microenvironment suggesting an immunomodulatory effect on the

NK cell population. Furthermore, 26% of SF NK cells were capable of producing IFN $\gamma$  and this was especially pronounced in NK cells derived from patients with destructive RA in addition to increased secretion of TNF<sup>317</sup>. Lastly, NK cells were considered to be in an activated state demonstrated by the expression of CD69 and NKp46<sup>317</sup>.

Less abundant immune cell populations including pDCs and  $\gamma\delta$  T cells have also been identified in RA SF but reported to a lesser extent in the literature. In RA SF, two subsets of dendritic cells have been detected: myeloid and plasmacytoid. pDCs were distinguished from myeloid DCs by their high CD123 expression and lack of CD11c expression<sup>308</sup>. Although the pDC population were detected in RA SF, they were present at a higher frequency in spondyloarthritis SF. It was hypothesised that pDCs were retained and mature in the RA joint and thus not trafficking through to the fluid. pDCs detected in the SF have shown to be activated in the joint but were not good APCs in the joint microenvironment. IFN $\gamma$ , IL-10 and TNF production were observed when pDCs from RA SF were incubated with mixed lymphocyte reactions and thus contributing to an inflammatory environment<sup>306</sup>.

Limited literature about  $\gamma\delta$  T cells in RA SF exists, however, there are reports suggesting that this subpopulation is present at an increased frequency in RA SF compared to paired peripheral blood and healthy controls suggesting that either  $\gamma\delta$  T cells are recruited to the joint from the blood or that there is a suppression of  $\gamma\delta$  T cells in peripheral blood<sup>376,377</sup>.  $\gamma\delta$  T cells are functionally capable of secreting IL-2, IL-3, IL-4, IFN $\gamma$  and GM-CSF suggesting that they contribute to the pathogenic environment<sup>378,376</sup>. Whilst these findings were reported 30 years ago, recent research has focused on the increased IL-17 secretion in RA SF produced by  $\gamma\delta 17$  and Th17 cells and demonstrated when inhibited by ROR $\gamma$ t antagonist, they reduced joint inflammation<sup>379</sup>. A further study observed that  $\gamma\delta 17$  T cells were not as abundant as IFN $\gamma$  producing  $\gamma\delta$  T cells suggesting that these cells play a dominant role in disease pathogenesis<sup>380,381</sup>.

### 6.1.3 Aims

It is evident that extensive immunophenotyping studies have been conducted to elucidate the composition of cells in RA SF and furthermore to detect the function of these subsets. The limitations of these studies is that immunophenotyping panels have been designed to profile specific immune cell subsets as opposed to a global analysis of immune cell populations which can bias findings. Therefore the overall aim of this chapter was to broadly assess the immune cell subsets that are present in synovial fluid to determine whether recent novel subsets detected by mass cytometry could be observed in this modest sized cohort. Ultimately, consistency across immunophenotyping platforms will streamline biomarker discovery analysis and a consensus as to which immune cell populations in RA SF could be useful for assessing chronic inflammation, treatment response and/or disease prognosis would aid existing strategies and clinical management of RA.

The following aims in this chapter are as follows:

1. To compare the differences in immune cell populations in paired RA blood and SFMCs and determine to what extent microenvironment influences this composition.
2. To corroborate mass cytometry findings by Rao et al<sup>157</sup> and observe whether Tph cells are expanded in RA SF compared to paired peripheral whole blood.
3. Identify additional effector T cell subsets and differential expression of activation/inhibitory markers to determine the functional status within the synovial microenvironment.

## 6.2 Results

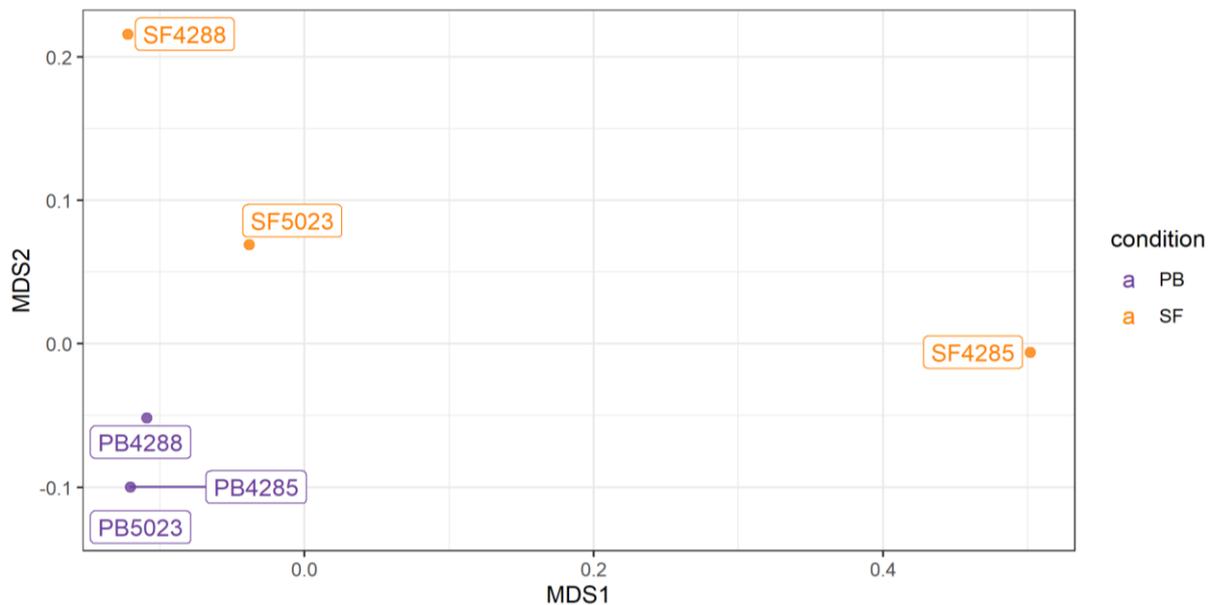
### 6.2.1 Comparison of immune cell populations present in paired whole blood and synovial fluid in RA

#### 6.2.1.1 MDS analysis shows global protein marker staining similarity is specific to biological microenvironment and not influenced by donor

Paired RA SF and peripheral whole blood were obtained from three patients presenting for arthrocentesis at Chapel Allerton Hospital. SF was obtained from the swollen knee joint for each patient.

The initial question for this analysis was assessing whether immune cell populations clustered according to the site from where they were obtained i.e. synovium and peripheral whole blood. Whilst the expectation was that different immune environments would influence which immune cell subsets were present and that these would also differ between the two compartments, it was still important to determine whether this assumption could be detected by the Diffcyt pipeline. Moreover, as these samples were paired, it was important to establish whether Diffcyt was not biased by donor label but driven by the underlying immunological differences.

27,000 events from each sample were included in the Diffcyt analysis workflow. To observe the global similarity across the protein marker staining expression, unsupervised MDS analysis (Figure 6-2) showed whole blood from all three patients clustered closely together towards the lower left quadrant of the plot indicating that the staining profiles are similar. Whilst the synovial fluid samples were not as close in proximity to one another on the MDS plot, it was observed that they are distinct from the peripheral blood profiles and are separated by the MDS 1 axis. MDS 1 refers to the first dimension of the scaling where the two conditions observed as to whether they can be separated. This initial analysis supports the expectation and well established concept that the rheumatoid microenvironment affects the immune cell landscape<sup>382-384</sup>. Although this is a well established phenomenon, it was encouraging to observe that even at the broadest level of analysis and with a modest sample size, this distinction is evident.



**Figure 6-2 | Multi-dimensional scaling (MDS) analysis shows the distinction of three paired RA synovial fluid (SF) and peripheral whole blood (PB).**

Paired samples of SF and PB from three patients with RA were analysed by the MDS plot. It can be observed that there is a clear distinction between the biological microenvironments, however, the SF between the three patients are broadly distributed compared to that observed for PB suggesting that the SF is more representative of the heterogeneity of RA compared to blood.

### 6.2.1.2 FlowSOM clustering shows the changes occurring at the immune cell level between paired RA peripheral whole blood and synovial fluid

The distinct clustering of peripheral whole blood and synovial fluid as evidenced by MDS analysis encouraged further analysis to determine which underlying immune cell populations were driving this distinction.

FlowSOM analysis shown in Figure 6-3 and heatmap (Figure 6-4) demonstrates the median expression of the lineage markers used to identify the main immune cell populations present in synovial fluid and peripheral whole blood (n=3 patients). Immune cells from each sample were pooled to represent each condition enabling a broad comparison between the two conditions. In total 15 immune cell populations were identified after heatmap analysis and each dot represents one cell on the two dimensional tSNE axes. As expected, neutrophils remain the largest population present in both peripheral blood and SF although further exploration of heterogeneity within this population was not possible due to limited neutrophil specific markers being included in the immunophenotyping panel. More immune cell populations were present in peripheral whole blood, including CD14<sup>+</sup> monocytes, B cells, memory B cells and naïve CD4<sup>+</sup> T cells whereas in SF, these populations were decreased in frequency or absent altogether. It is however evident, that memory CD8<sup>+</sup> T cells are present at a higher frequency in SF compared to peripheral whole blood.

The observation that naïve CD4 T cells are essentially absent in synovial fluid and that memory CD4 T cells form the most abundant T cell population suggests that this population is significantly expanded in SF compared to whole blood. Clustering analysis did not reveal specific subsets within the memory CD4 T population as described by Fonseka and Rao where CD4<sup>+</sup>CD27<sup>+</sup>HLA-DR<sup>+</sup> memory T

cells and CD4<sup>+</sup>PD-1<sup>+</sup>HLA-DR<sup>+</sup> Tph cells are two separate populations<sup>157,261</sup>. In this analysis, these two subsets were indistinguishable with the lineage markers included and in order to keep findings consistent with those in the previous chapter, it was decided not to redefine the lineage and functional markers. An advantage for this approach is that it is inclusive and not biased by the findings of previous researchers whilst simultaneously providing further evidence and strengthening the replicability of the data if similar immunophenotypes are identified.

The striking expansion of the memory CD4 T cell population in RA SF could have supported the decision to pursue analysis of the T cell population exclusively and remove additional populations thus increasing the number of CD45RO<sup>+</sup> CD4<sup>+</sup> T cells included in the Diffcyt pipeline. However, this would eliminate the prospect of observing functional differences across other immune cell populations. Moreover, should changes in the memory CD4 T cell subset be detected in a heterogeneous population of immune cells, this emphasises the importance of the cell population.

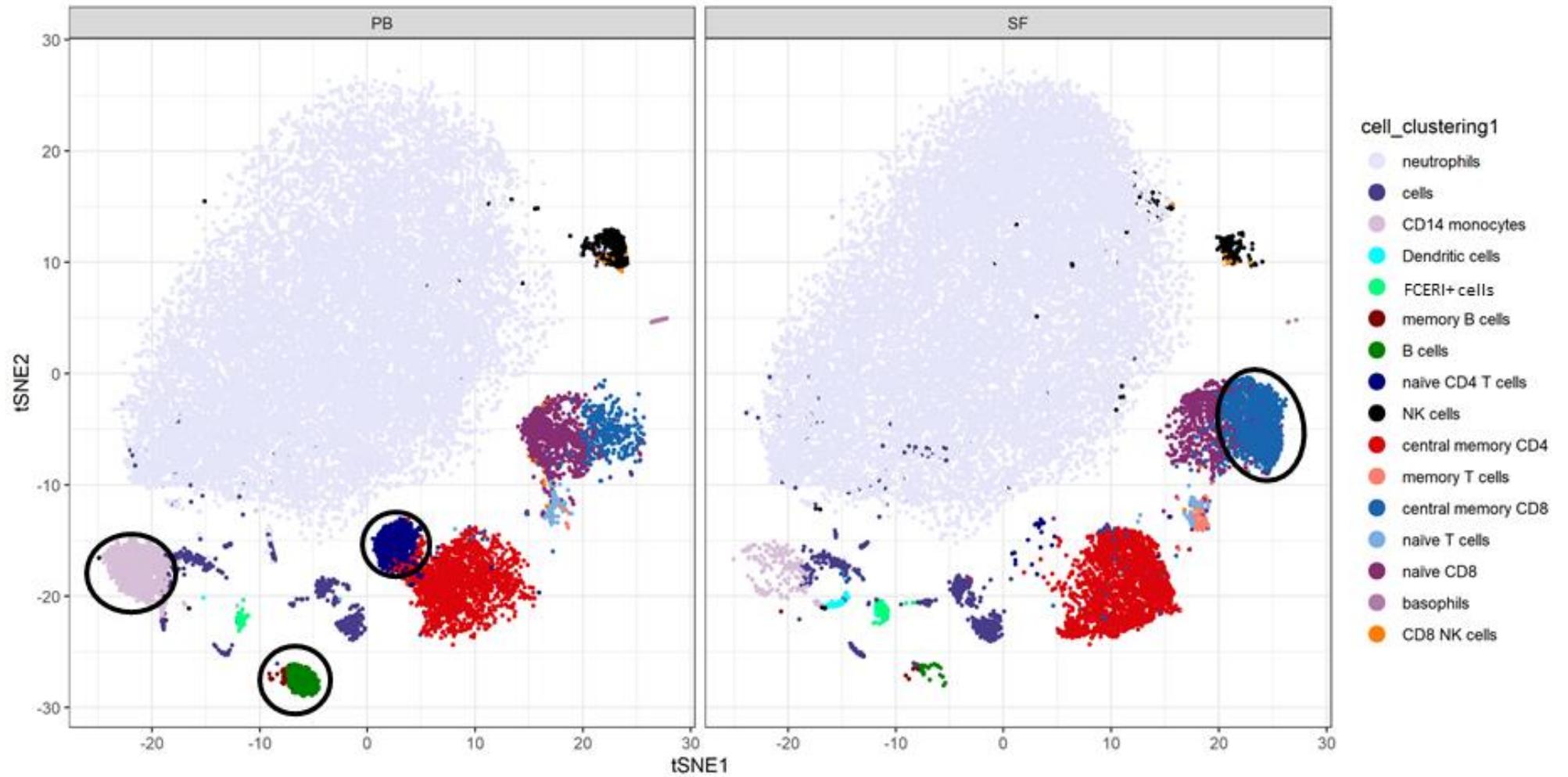
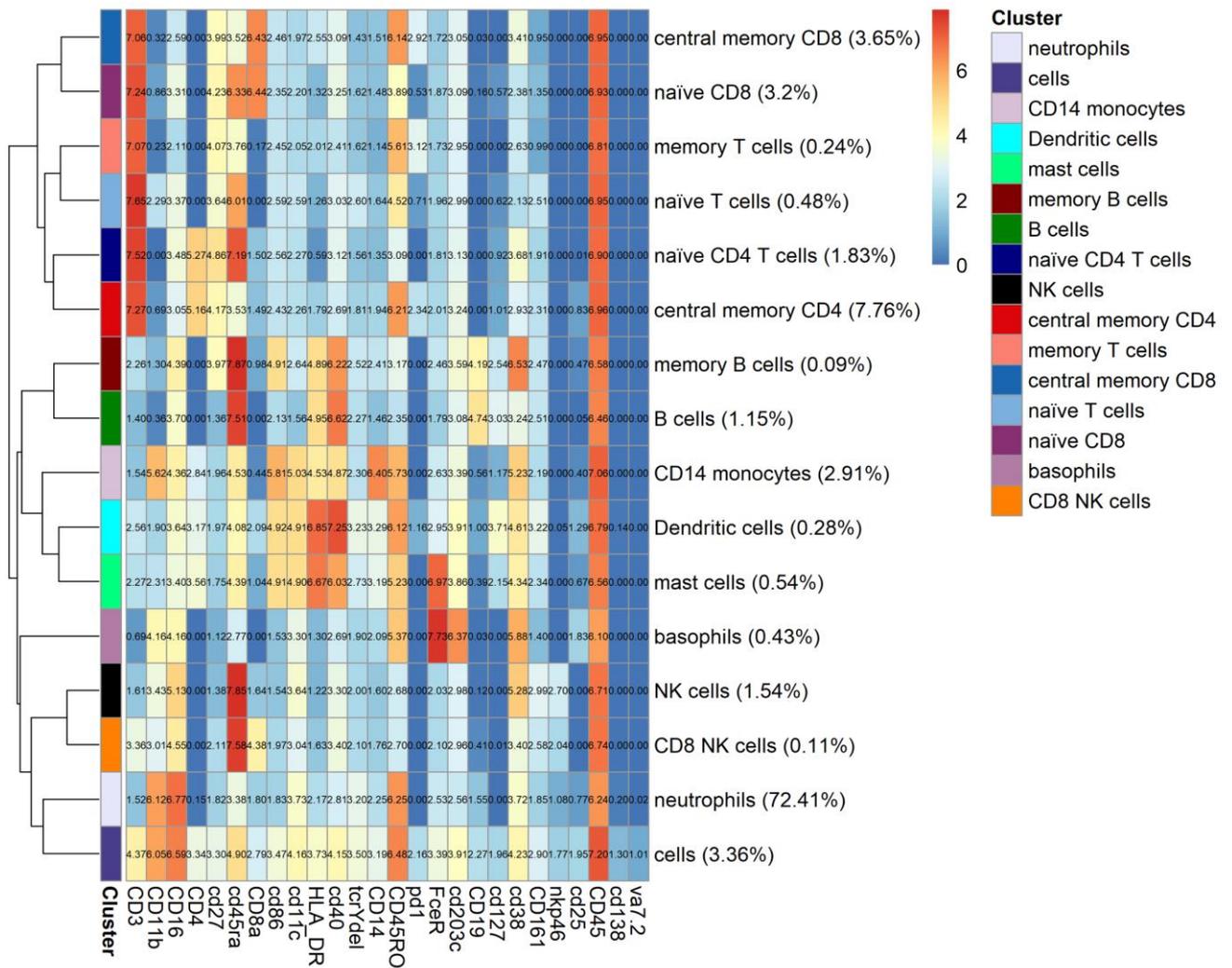


Figure 6-3 | FlowSOM clustering comparing paired RA peripheral whole blood (PB) and synovial fluid (SF) immune cell populations.

Clustering analysis representative of 3 patients shows differences in the immune cell populations indicated by the black outlined circles. In whole blood, CD14+ monocytes, B cells, memory B cells and naïve CD4 T cells are present whereas in synovial fluid, these populations are notably absent or present at a decreased frequency. In synovial fluid clustering analysis, the central memory CD8 T cells have an increased frequency in synovial fluid compared to blood.

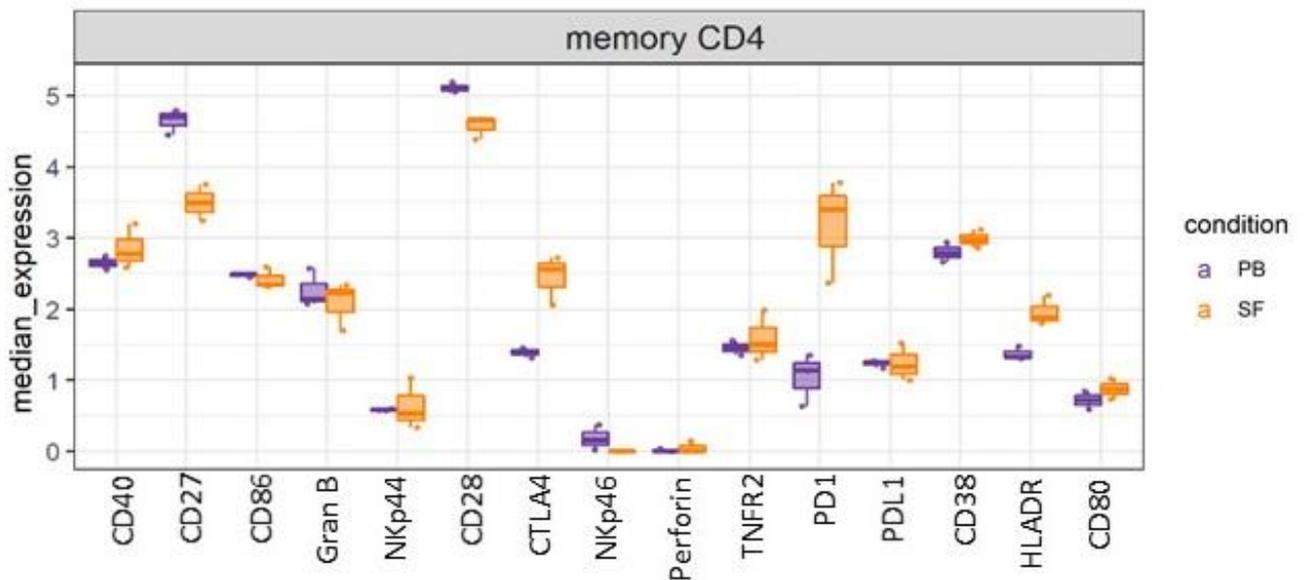


**Figure 6-4 | Heatmap showing the lineage markers used to define immune cell subsets**

Immune cell populations identified in peripheral whole blood and synovial fluid are shown with the percentage of cells in each cluster shown in brackets. Median expression intensity for lineage markers, coded by a colour scale blue = 0 (no expression) to red = 6 (high expression)

### 6.2.1.3 Chronic activation of T cell subsets in RA SF

Following high-dimensionality clustering, analysis of 15 functional protein markers stratified by immune cell populations revealed that within the memory CD4 T cell population, changes in functional marker median expression could be detected between matched peripheral whole blood and SF from patients with RA. This was calculated and presented as box plots depicting the minimum, interquartile range and maximum for each patient sample (Figure 6-5). A decrease in median expression in both CD27 and CD28 in RA SF suggests that these T cells in SF are chronically activated after continual exposure to autoantigens. Median expression changes were also observed in CTLA-4 and PD-1 checkpoint inhibitors where they were increased in SF compared to their matched peripheral whole blood. An increase in HLA-DR median expression in the memory CD4 population was also observed supporting previous reports of Tph cells and memory CD4 HLA-DR<sup>+</sup> CD27<sup>-</sup> T cells. From the FlowSOM clustering analysis it was not apparent that these two populations were separate.

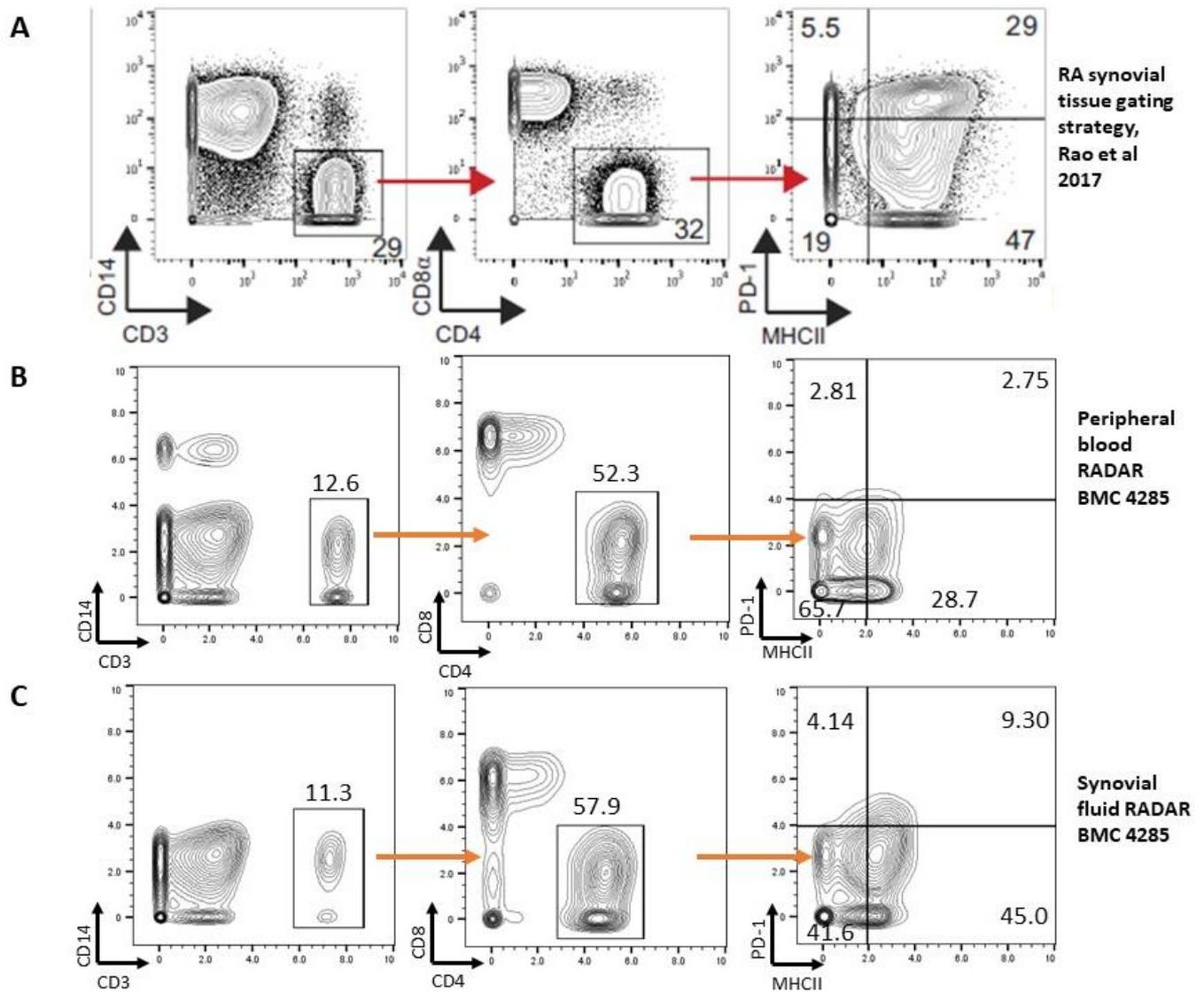


**Figure 6-5 | Box plot representation of median expression of functional protein markers within the memory CD4 T cell population reveal a distinct immune profile (PD-1<sup>+</sup> CTLA4<sup>+</sup> CD27<sup>-</sup> CD28<sup>-</sup> HA-DR<sup>+</sup>) in RA synovial fluid**

Median expression analysis of 15 functional protein markers CD40, CD27, CD86, Granzyme B (Gran B), NKp44, CD28, CTLA-4, NKp46, Perforin, TNFR2, PD-1, PD-L1, CD38, HLA-DR and CD80 in order from left to right for RA matched whole blood (PB) and synovial fluid (SF) n=3. A distinct profile which phenotypically resembles T peripheral helper cells previously reported to be implicated in RA could be detected by the changes in the median expression of CD27, CD28, CTLA-4, PD-1 and HLA-DR in SF. Median expression of CD27 decreased to 3.5 in RA SF whereas in matched whole blood median expression is 4.5. Similarly, CD28 median expression was decreased in RA SF with a median of 4.5 compared to 5 in in matched blood. Conversely, median expression of checkpoint inhibitor proteins CTLA-4 and PD-1 were markedly increased in RA SF (median expression 2.5 and 3.4 respectively) compared to matched peripheral whole blood. HLA-DR median expression levels were also slightly increased in RA SF at 2 whereas in blood median expression was 1.4.

To understand whether the Tph cells were indeed present in matched blood and SF, this was confirmed manually using FlowJo software and the gating strategy previously described by Rao et al<sup>157</sup>. In their paper, they describe the gating strategy for RA synovial tissue by mass cytometry and this gating strategy was adopted and applied to the samples. The gating strategy is shown for one patient using matched blood and SF and reveals that the Tph cell population are expanded in RA SF compared to blood providing further evidence that not only are the Tph cell population present in RA but are expanded in the joint microenvironment (represented by synovial tissue biopsy and SF) as shown in Figure 6-6. The gating strategy in Figure 6-6 B and C which correspond to blood and SF respectively, shows the sequential percentage of cells in each gate starting with 27,000 cells in the CD14/CD3 gate. The percentage of CD4<sup>+</sup> MHC I<sup>+</sup> (HLA-DR) PD-1<sup>+</sup> cells were 2.75% in whole blood however, in synovial fluid this population increased to 9.3%. This was still less than 29% which was identified in synovial tissue by Rao et al but in a stringent population of just 27,000 cells, the evidence still robustly demonstrates that Tph cells are detectable particularly in the joint.

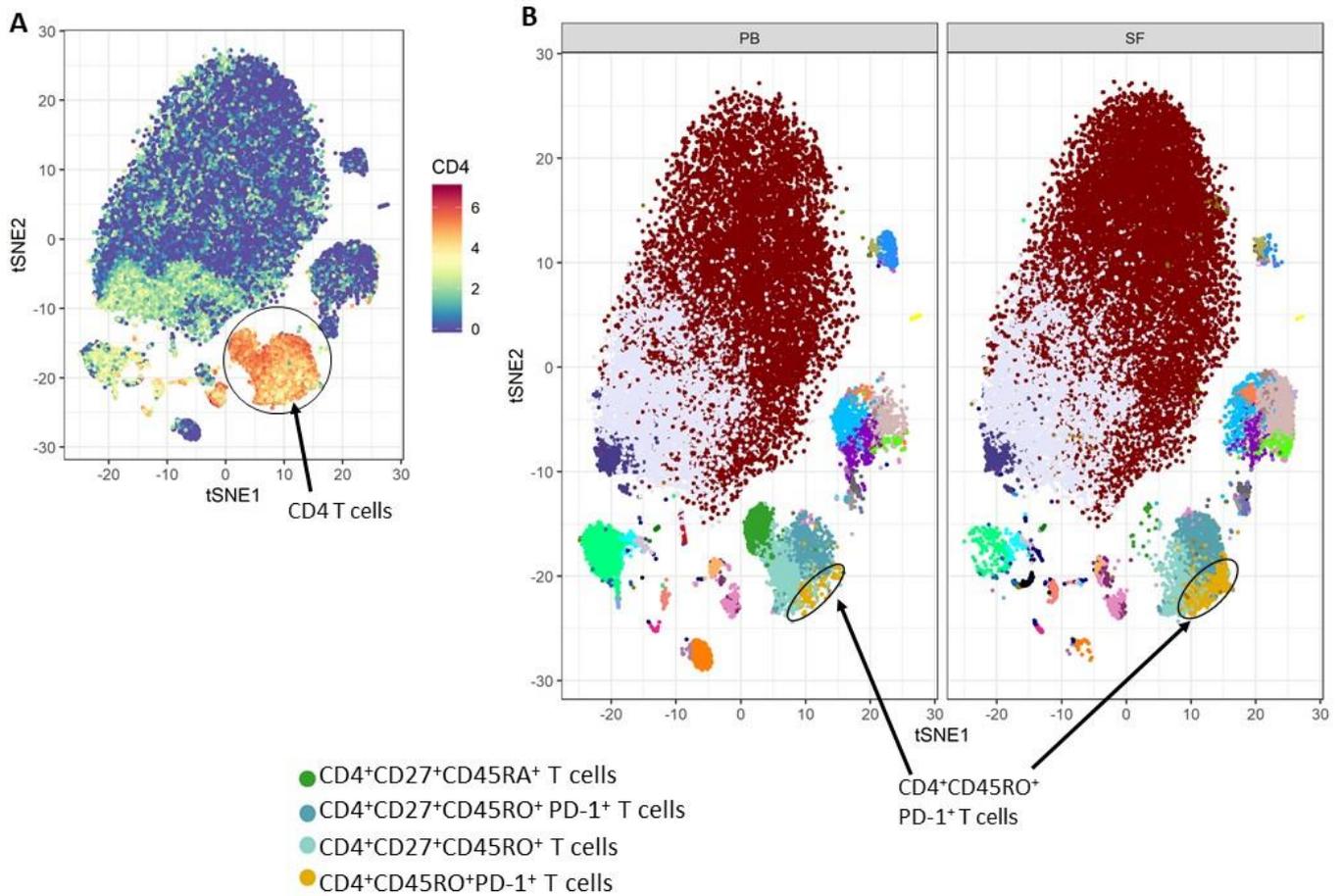
The initial analysis conducted by FlowSOM clustering algorithm in Figure 6-3 was performed to observe the global immune landscape between RA peripheral blood and SF and secondly to identify if any differences were detectable between the two biological compartments or whether the immune cell populations were similar. Figure 6-3 did reveal that there were differences in immune cell populations present in peripheral blood and SF. To further interrogate this finding, FlowSOM clustering was re-run (Figure 6-7A and B) with 27,000 cells from each sample (n= 3 matched RA SF and peripheral blood) with particular focus on observing immunophenotypic subsets within the CD4<sup>+</sup> T cell population (Figure 6-7A). Figure 6-7B shows four subsets identified in the CD4<sup>+</sup> T cell population which were immunophenotypically defined as: CD4<sup>+</sup>CD27<sup>+</sup>CD45RA<sup>+</sup> T cells, CD4<sup>+</sup>CD27<sup>+</sup>CD45RO<sup>+</sup>PD-1<sup>+</sup> T cells, CD4<sup>+</sup>CD27<sup>+</sup>CD45RO<sup>+</sup> T cells and CD4<sup>+</sup>CD45RO<sup>+</sup>PD-1<sup>+</sup> T cells and FlowSOM detected these 4 subsets individually prior to labelling. The expansion of Tph cells (CD4<sup>+</sup>CD45RO<sup>+</sup>PD-1<sup>+</sup> T cells) can be visually observed in RA SF compared to peripheral blood, which further confirms the presence and relevance of Tph cell in RA.



**Figure 6-6 | Gating strategy for identifying T peripheral helper cells by mass cytometry in RA synovial tissue and matched peripheral blood and synovial fluid.**

A Rao et al in 2017<sup>157</sup> identified the pathological expansion of T peripheral helper (Tph) cells in synovial tissue biopsies obtained from three patients by mass cytometry. The paper outlined the gating strategy applied to identify Tph cells by firstly gating on CD3+ CD14- T cells. A subsequent gate is drawn on the CD4+ CD8- T cell population and lastly a quadrant gate on a contour plot was drawn to identify the expansion of Tph cells. **B** The gating strategy outlined in A was applied to peripheral blood (RADAR BMC 4285) where the Tph cells are not expanded when compared to synovial tissue. **C** Synovial fluid taken from the same donor (RADAR BMC 4285) showed a similar Tph cell profile seen in synovial tissue suggesting the joint microenvironment is enriched for these cells. Percentages are shown of each sequential gate and in each quadrant the percentage is shown.

Note that the gating axes in A are on logarithmic scales and in B and C are shown on linear scales due to automated gating applied to the dataset which transforms the scale.

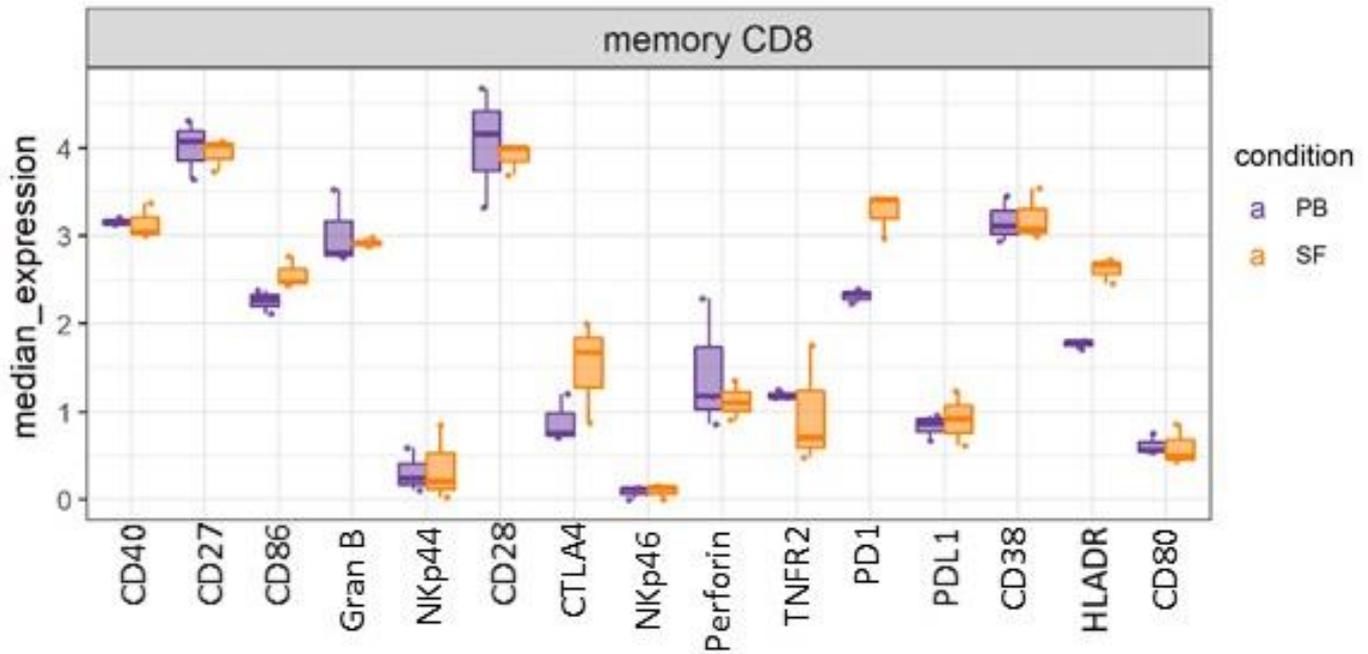


**Figure 6-7 | FlowSOM clustering analysis shows the expansion of CD4<sup>+</sup> CD45RO<sup>+</sup> PD-1<sup>+</sup> T cells in RA synovial fluid which immunophenotypically resemble T peripheral helper cells.**

**A** FlowSOM high-dimensional clustering shows the position of CD4<sup>+</sup> T cells (coloured red/yellow) on a dimensionality reduction t-SNE 1 by t-SNE 2 axes. Signal intensity is depicted by a colour scale with blue – red indicating no expression to high expression of specific marker. **B** shows the FlowSOM clustering analysis plot which summarises the cell populations from 3 patients where matched RA blood and synovial fluid was collected. 4 subsets within the CD4<sup>+</sup> T cell population in peripheral blood (PB) and synovial fluid (SF) was identified by FlowSOM including CD4<sup>+</sup>CD27<sup>+</sup>CD45RA<sup>+</sup> T cells which are present in PB but absent in SF. CD4<sup>+</sup>CD27<sup>+</sup>CD45RO<sup>+</sup>PD-1<sup>+</sup> and CD4<sup>+</sup>CD27<sup>+</sup>CD45RO<sup>+</sup> T cell subsets were also identified. CD4<sup>+</sup>CD45RO<sup>+</sup> PD-1<sup>+</sup> T cell subset are observed to be increased in SF compared to PB.

### 6.2.1.4 PD-1 expression is increased in memory CD8 T cells in RA SF compared to matched peripheral blood

To determine whether median expression changes in functional protein markers could be detected in other immune cell populations between matched blood and SF. Observation of the memory CD8 T cells (characterised as CD8<sup>+</sup> CD45RO<sup>+</sup> CD27<sup>+</sup>) showed an increased median expression in CD86, CTLA-4, PD-1 and HLA-DR in RA SF compared to blood (Figure 6-8).

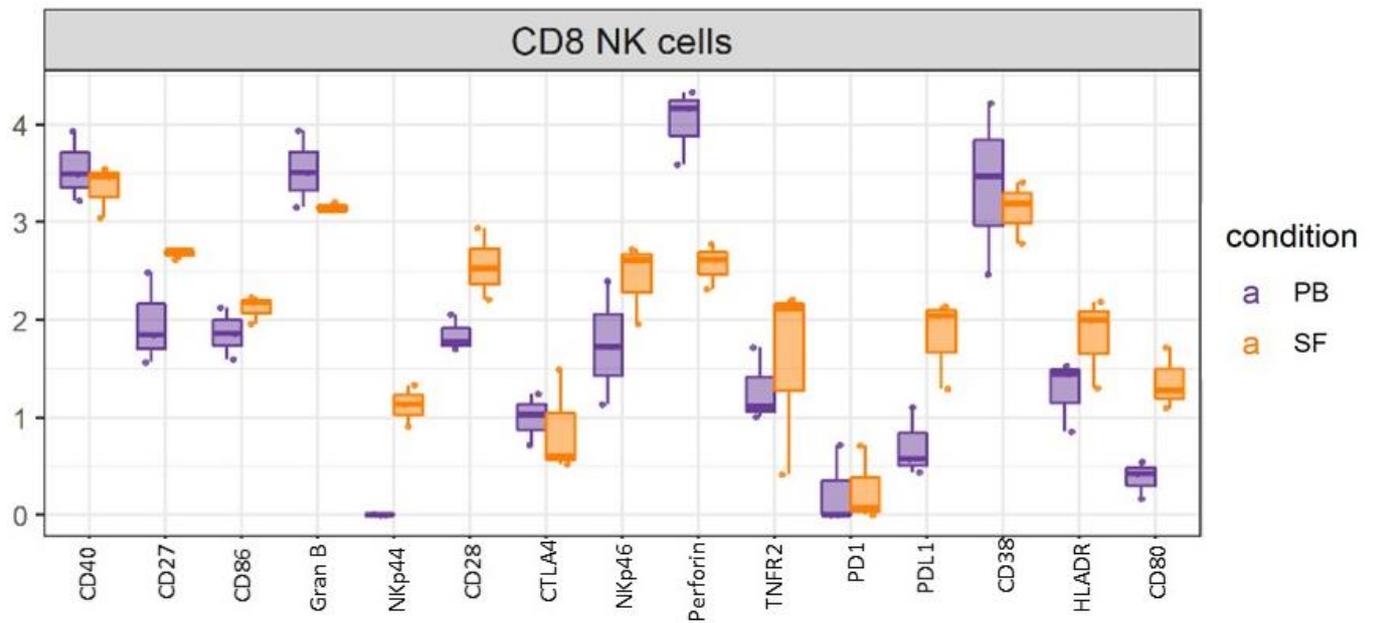


**Figure 6-8| Box plot representation of median expression of functional protein markers within the memory CD8 T cell population reveals higher median expression of checkpoint inhibitors CTLA-4 and PD-1 in RA synovial fluid compared to matched peripheral whole blood**

Median expression analysis of 15 functional protein markers CD40, CD27, CD86, Granzyme B (Gran B), NKp44, CD28, CTLA-4, NKp46, Perforin, TNFR2, PD-1, PD-L1, CD38, HLA-DR and CD80 in order from left to right for RA matched whole blood (PB) and synovial fluid (SF) n=3. A similar immunophenotypic profile that was observed in memory CD4 T cell population was similarly identified in this population including CTLA-4 which had a median expression of 1.8 compared to 0.8 in peripheral whole blood. PD-1 median expression was similarly increased in memory CD8 T cells in RA SF with a median expression of 3.4 and 2.4 in peripheral whole blood. HLA-DR was also showed an increase in median expression 2.6 in RA SF compared to 1.8 in peripheral whole blood. A slight increase in CD86 median expression can also be observed in RA SF.

### 6.2.1.5 CD8 NK cells display a chronically activated functional immunophenotype in RA SF compared to matched whole blood

Evidence of chronically activated CD8 NK cells in RA SF could be detected by the increase in median expression of CD27, NKp44, CD28, NKp46, PD-L1, HLA-DR and CD80 in RA SF compared to matched peripheral whole blood (Figure 6-9). Conversely, this activated immunophenotype suggests that CD8 NK cells may have a diminished ability to induce apoptosis in the joint demonstrated by the decrease in median expression of both Perforin and granzyme B.



**Figure 6-9 | Box plot representation of median expression of functional protein markers within the CD8 natural killer (NK) cells population reveals a chronically activated immune phenotype but an incapability of inducing apoptosis through Granzyme B and Perforin**

Median expression analysis of 15 functional protein markers CD40, CD27, CD86, Granzyme B (Gran B), NKp44, CD28, CTLA-4, NKp46, Perforin, TNFR2, PD-1, PD-L1, CD38, HLA-DR and CD80 in order from left to right for RA matched whole blood (PB) and synovial fluid (SF) n=3. Functional protein marker median expression was increased in CD27 (2.3), NKp44 (1.2), CD28 (2.5), NKp46 (2.6), PD-L1 (2.0), HLA-DR (2.0) and CD80 (1.25) in RA SF compared to matched peripheral whole blood. In contrast, Granzyme B (3.2) and Perforin (2.5) median expression decreased in RA SF compared to matched peripheral whole blood.

## 6.3 Discussion

### 6.3.1 Strengths and limitations of the study

It is recognised that this analysis was limited by the number of samples included. This can be directly attributed to the reduction of joint aspiration performed in clinic over the course of 18 months due to improved clinical management of RA. This can only be interpreted as positive progress in the clinical management of RA as a direct result of improved diagnosis and treatment and was thus considered fortunate to have been able to obtain 3 matched samples given this progress. This may then lead to the question of why pursue immunophenotyping within the blood and joint microenvironment if excessive inflammation is not apparent? But to an extent the question answers itself. The fact that large joint aspiration still occurs means that not all patients are prognostically identical. 40% of patients who fail initial biological treatment with a TNF inhibitor provides the biggest indication that subtle immune populations that do not constitute the majority of the immune cell landscape may contribute to treatment resistance and poor prognosis. The era of multi-parametric, high throughput technology such as mass cytometry allows for either novel immune cell populations or subtle immune signatures to be identified and furthermore confirm previous findings in the literature which were limited by the current technology. If findings from novel technologies support these previous reports it strengthens and re-emphasises the importance of these populations in RA.

Including all immune cells for analysis was deliberate to ensure that no major cell populations were overlooked. The panel design was designed in this approach so as to capture an overview and not to specifically interrogate a subset of immune cells. However, the panel design could be viewed as having a T cell focus especially with the inclusion of the checkpoint inhibitors. This was intentional as T cells are an abundant cell population and intrinsic to RA pathology and have been extensively explored by other research groups including the AMP<sup>157,161,261</sup>. These markers were included to serve as a validation control to observe whether the same could be identified in this analysis. The observation that chronically activated phenotypes were observed within the memory CD4 T cell population underscores the relevance of this population. Therapeutics targeting checkpoint inhibitors and particularly the effect of PD-1 and CTLA-4 inhibition in RA are underway to establish whether any adverse effects are apparent<sup>385-387</sup>.

It is acknowledged that a limitation of this analysis is the extent to which it was performed. It is anticipated that the approach described here and the findings which have been reported can be extended upon further either by using the existing approach or additional approaches which are now emerging such as Flowct<sup>165</sup>. Furthermore, comparison of these findings with SF obtained from other inflammatory arthritis in this analysis did not reveal any stark differences although it was noticed that PD-1 median expression levels were slightly higher in RA than other inflammatory arthritis (data not shown).

The findings in this study are preliminary and it is accepted that further mechanistic studies would be required to understand the deeper relevance of these populations. It is also clear that synovial tissue would provide an added insight with an aim to perform whole tissue analysis as opposed to using disaggregated tissue. However, it is not thought that a larger cohort of SF is required despite the small number in this study and this is because logistically this is difficult to scale up. Moreover, as the findings from this analysis corroborate existing findings in the literature, this increases confidence in the relevance of memory CD4 T cells and CD8+ NK cells as populations of interest in RA

pathology. An interesting question would be to assess whether acellular RA SF has immunomodulatory properties on healthy T cells in vitro. Is RA SF potent enough that loss of CD27 and CD28 expression can be detected in healthy T cells and are these cells chronically activated or anergic? Further mechanistic work should look at Tph cells in closer detail particularly in the context of treatment as this has not been fully established. Finally understanding the role of CD8+ NK cells in RA pathology by directly isolating them from patient blood/SF and characterising their functional role such as cytokine expression and their role in antigen presenting cells would be essential. Establishing the autoantigen/s that perpetuate chronic inflammation will be key and establishing how these antigens contribute to a dysregulated immune response continues to be extensively researched. Whilst this analysis cannot provide these answers, it is anticipated that this work will contribute to the existing efforts of pursuing functional characterisation of CD8+ NK cells and memory CD4 T cell populations especially in the joint and whether these populations and how these populations respond to existing treatment or whether they could be the target of future novel therapies.

### **6.3.2 Immune cell populations differ between donor matched RA whole blood and synovial fluid**

The main aim of this chapter was driven by the over-arching question ‘is there a difference between synovial fluid and whole blood at the immune cell level or a specific immunophenotype which can distinguish between the two biological microenvironments?’ This was robustly supported by the MDS analysis in Figure 6-2 which showed that RA SFMCs and whole blood from the same donor formed two distinct groups driven by the underlying immunological differences between the biological samples. The validity of this finding was strengthened by the fact that both blood and SF were matched from the same donor, processed in the laboratory simultaneously and therefore provided a control for donor variation. The confidence in this initial finding was reflected by FlowSOM high-dimensional clustering which identified 15 main lineage immune cell populations present across whole blood and SF.

### **6.3.3 Chronic pathologic activation within the memory CD4 T cell compartment in the RA SF**

The analysis conducted by the Diffcyt pipeline revealed that within the memory CD4 T cell population, a chronically activated functional phenotype could be observed. Memory CD4 T cells have been extensively reported to be enriched, both within RA blood and joint microenvironments<sup>157,160,261,388–393</sup>. Within RA, it is widely accepted that CD4 T cells are implicated in autoimmune disease as they recognise peptide antigens in the context of MHC II molecules and specific MHC alleles may increase predisposition to disease<sup>27</sup>. This is evident in RA based on the shared epitope hypothesis which is supported by the observation that HLA-DRB1 alleles are strongly associated with susceptibility to severe arthritis<sup>27,394</sup>. Indeed, a long held view is that as T cell stimulation is perpetuated by complex cytokine interactions, antigen stimulation and costimulatory signals, all of which are abundantly present in the joint microenvironment, this suggests that T cells migrating from the blood are influenced by the local environment within the joint<sup>395</sup>. As only 27,000 events from each sample were included for the Diffcyt workflow it was decided to look at the memory CD4 T cell population as a whole (CD4<sup>+</sup> CD45RO<sup>+</sup> CD45RA<sup>-</sup> T cells) and not split into subsets in order to observe the global expression of 15 functional markers as defined for this analysis. This was

considered to be an inclusive approach and reduce bias in the interpretation of the data that may be introduced through preferential analysis or existing findings in the literature. Although percentage abundance of memory CD4 T cells did not differ between whole blood and SF, further inspection of the functional markers defined within the Diffcyt workflow identified specific changes which suggest a chronically activated phenotype particularly visible in the joint.

The first of these was detected in the decreased expression of CD27 in RA SF. CD27 is a member of the TNFR superfamily and is involved in both cell activation and receptor mediated apoptosis<sup>396</sup>. CD27 is expressed on all naïve CD4 T cells and approximately 80% of memory CD4 T cells and can vary from high expression to dim expression intensity which can subsequently be used to determine subsets within lineage populations<sup>397</sup>. In vitro T cell activation studies have shown that memory CD4 CD45RO+ CD27- T cells are likely to emerge from CD27+ T cells as a result of prolonged stimulation which does not appear to be reversible<sup>398</sup>. The decrease in CD27 within SF and synovial tissue compartments has been reported by Kohem et al. in 1996 where CD45RO+CD45RA- CD4 T cells were expanded and within this population a further insight revealed this population was enriched for differentiated CD4+ CD45RB+ CD27- T cells, a mature T cell population thought to have formed as a direct result of prolonged antigenic stimulation<sup>398</sup>. This demonstrates that the enrichment of these cells in SF and synovial tissue are in part also due to these cells having an enhanced capacity to migrate from the vascular space into inflamed tissue<sup>398</sup>. Indeed, this analysis revealed by FlowSOM clustering that the presence of naïve CD4 T cells are diminished in SF compared to whole blood which further supports the hypothesis that the local joint environment influences the immunophenotype of T cells. Specifically which antigens are responsible for perpetuating this chronic activation of T cells remains unknown.

Since the findings of Kohem et al., many reports have detailed the expansion of CD45RO+ CD4 CD27- T cells in synovial fluid and whole blood from patients with RA<sup>397,399-402</sup>. CD27 expression has been observed in the CD4+ CD25+ population in SF obtained from patients with juvenile idiopathic arthritis<sup>401</sup>. The researchers used CD27 to distinguish between regulatory and effector T cells where CD4+CD25+CD27+ T cells expressed high levels of FoxP3 but did not produce IL-2, IFN $\gamma$  or TNF however, the converse was observed in CD4+CD25+CD27- where effector cytokines were produced and failed to prevent T cell proliferation<sup>401</sup>. In another study, gene expression profiling revealed that CD4+ CD27- T cells had an effector phenotype shown by the upregulation of transcription factors GATA-2 and T-bet and the chemokine receptor CX<sub>3</sub>CR1, chemokine RANTES and the activation marker HLA-DR<sup>397</sup>. The gene profile was markedly different compared to CD27+ T cells thus supporting a different functional role for CD27- T cells. Tak et al. observed the differentiation of memory CD4 T cells into CD45RA+ CD27+ T cells which subsequently change into a CD45RO+ CD27+ which in turn become CD45RO+ CD27- phenotype<sup>400</sup>. Within the synovial tissue compartment, infiltration of CD4+CD45RO+CD27+ T cells could be identified in the rheumatoid perivascular lymphocytic aggregates with a relative increase in the percentage of CD27- T cells in the diffuse lymphocytic infiltrate<sup>400</sup>. Finally, the mixed effects association of single cells algorithm revealed the pathologic expansion of the memory CD4+CD27- HLA-DR+ subset in RA peripheral blood by mass cytometry, again confirming the importance of this subset in RA<sup>261</sup>.

Further analysis within this work observed that aberrantly activated T cells were detected by a decrease in CD28 and an increase in CTLA-4, PD-1 and HLA-DR. CD4 CD28<sup>null</sup> T cells have received renewed interest due to their prevalence in autoimmune disease including multiple sclerosis and SLE<sup>403,404</sup>. Intriguingly, CD4 CD28<sup>null</sup> T cells were first characterised and observed to be expanded in RA peripheral blood and later found to infiltrate the joint<sup>391,392</sup>. Evidence suggests that CD4 CD28<sup>null</sup> T

cells have an immunophenotype indicative of cells having evaded peripheral tolerance which result in them having an autoreactive response to autoantigens<sup>392</sup>. Moreover, studies have shown that the presence of CD4 CD28<sup>null</sup> T cells in RA can lead to the development of extra-articular manifestations suggesting these cells are systemically implicated. A recent review by Bano et al. directly links the expansion of CD4 CD28<sup>null</sup> T cells based on prior exposure to cytomegalovirus<sup>333</sup>. This concept is further elaborated by the authors in the context of RA where they propose that cytomegalovirus-infected synovial fibroblasts present antigens indirectly to exosomes which are processed by professional and subsequently processed by antigen presenting cells which present to CD4 T cells via class II MHC. This interaction causes CD4 T cells to produce IFN $\gamma$  and TNF and results in CX3CR1 becoming upregulated on endothelial cells and synoviocytes which contributes to the recruitment of CD4 CD28<sup>null</sup> T cells to the inflamed synovium<sup>333</sup>. This is an interesting stance as it detracts from the identification of a specific autoantigen such as type II collagen and instead refers to either an inappropriate immune response to a viral pathogen (Bano et al. propose cytomegalovirus<sup>333</sup>) that results in the reprogramming of immune cells or insufficient recovery from a viral infection leading to persistent and chronic inflammation. It is likely that a combination of both processes have become dysregulated leading to the expansion of CD4 CD28<sup>null</sup> T cells. Evidence in support of the expansion of CD4 CD28<sup>null</sup> T cells in response to cytomegalovirus in RA has previously been reported including Thewissen et al. who identified CD4 CD28<sup>null</sup> T cells were reactive to cytomegalovirus<sup>403</sup>.

The loss of CD28 expression on T cells has been linked with ageing and immunosenescence<sup>333,405–407</sup>. A study comparing centenarians with younger individuals, reported a noticeable reduction in peripheral blood CD28+ T cells in centenarians and further in vitro analysis revealed that a loss of CD28 expression was also observed in T cells that had been in long term culture<sup>406</sup>. Until recently, immunosenescence was considered to be an irreversible process, that is once a cell reaches their maximum proliferative capacity this cannot be reset or reversed and this in part can be attributed to persistent chronic infection which contributes to accelerating the immune ageing of the cells<sup>408</sup>. This

The findings in this study also managed to identify the expansion of Tph cells in RA SF which was increased in abundance compared to peripheral blood, first reported in detail by Rao et al.<sup>157,158</sup> In addition, the analysis also showed an increase of PD-1 expression on memory CD8+ T cells in RA SF cells. These findings are intriguing as although the cell populations are different, a common feature is the expression of markers commonly associated with exhaustion including PD-1. However, expression of markers associated with exhaustion including PD-1 and CTLA-4 do not instantly identify these T cells as exhausted and within the RA joint, these cell populations are referred to as being chronically activated as opposed to showing functional exhaustion<sup>157,261</sup>. Whilst it is known that early senescence and exhaustion in T cells are both reversible, it is less certain whether this is the case after chronic, persistent infection. It will therefore be an important future research question to determine whether these populations can be therapeutically targeted leading to a potential reversal of these phenotypes and how this would affect the immune system long term. It is becoming apparent that the memory CD4 T cell compartment in RA are integral to pathology and understanding their role and attenuating their presence may offer a much needed therapeutic solution<sup>405</sup>.

#### **6.3.4 CD8+ NK cells – the bridge between innate and adaptive immunity in RA?**

Although this analysis is not the first to observe NK cells with T cell properties, interrogation of the literature suggests that it is the first time CD8+ NK cells have been immunophenotyped in detail by mass cytometry in RA. A limitation of current immunophenotyping studies is that despite the

increase of markers that can be interrogated and thus ability to acknowledge the heterogeneity of the immune landscape, reports often focus on one immune cell population or specific subset<sup>409</sup>. Examples of this can be observed in the initial reports that have been published regarding immune cell populations detected in RA by mass cytometry which has already been comprehensively detailed in earlier chapters<sup>157,161,261</sup>. This limitation arises due to the complexity and nature of immunology in general and maybe a flawed attitude towards the data we anticipate from novel technologies. Many of the questions and approaches suggest that it is a specific immune cell subset that will be responsible for perpetuating RA or that immune signatures will distinguish a subset of patients e.g. distinguishing early RA from refractory RA. Moreover, there is a tendency to view immune cells using the binary classification of innate and adaptive and there is still a leaning towards immunophenotyping adaptive immune cells<sup>157,161,261</sup>. Whilst one could support this bias towards B and T cells being the preferred cell populations to be studied in RA due to their inextricable association with autoantibody production, it is worth considering the wider implication of immune cell interaction and not overlooking innate cells or cells that do not fit this conventional classification. Whilst this creates a degree of complexity, it is necessary to view cell populations in context and not independently.

The CD8+ NK cell population once again showed relevance to RA pathology and particularly in the joint. In chapter 5, comparison of early RA peripheral blood revealed that both NK cells and CD8+ NK cells were significantly decreased compared to health suggesting that they were either migrating to the joint or undergoing apoptosis. Although modest in number, this analysis supports the theory that CD8+ NK cells are present in the joint and furthermore develop a different phenotypic profile compared to that in the blood. The occurrence of CD8+ NK cells reaffirms their importance in RA and the observation that they are distinct population from T cells and NK cells as depicted by the FlowSOM clustering provides encouragement for future investigation into this population of cells.

In this analysis, the CD8+ NK cell population appeared to have an activated phenotype in RA SF compared to blood, and this was evidenced by an increase in CD27, CD8+6, NKp44, CD28, NKp46, PDL1, HLA-DR and CD8+0 median expression. The expression of NKp44, PDL1 and CD8+0 suggest that CD8+ NK cells upregulate co-inhibitory receptors and develop an activated phenotype but have reduced cytotoxic potential as evidenced by the reduction of perforin and granzyme B expression. It is hypothesized that these cells are a hybrid of NK cells and T cells and have a heterogeneous phenotype<sup>410</sup>.

NK cells have a conflicting role in RA pathology and depending on the literature and interpretation of the data, NK cells could be interpreted as having a proinflammatory phenotype capable of mediating bone destruction and joint erosion and thus associated with a severely destructive RA progressing phenotype<sup>317</sup> or as cells that possess immunoregulatory functions by bridging the gap between T cell exhaustion and activation<sup>391,411,412</sup>. The role of NK cells in RA becomes significantly more complex when discerning the functional role of subsets of NK cells<sup>413</sup>. This study and the analysis presented is unable to comprehensively explain the role of CD8+ NK cells in RA, suffice to suggest that their role must be integral to RA pathology due to their significant reduction in early RA peripheral blood (observed in chapter 5) and their phenotype again being observed in RA SF. Analysis presented here cannot definitively confirm whether CD8+ NK cells are increased in SF, however, there it does suggest that there is an expansion of CD8+ NK cells in the joint<sup>391</sup>. It is also thought that CD8+ NK cells display an intriguing phenotype that should not be ignored. Despite the limitations of this analysis, CD8+ NK cells should receive further attention as it is hypothesized here that they may serve as an important mediator of T cell interaction<sup>391</sup>.

Warrington et al., in 2001 observed that a subset of CD4+CD28- T cells displayed NK cell features including CD158, p70, CD94, CD161, and CD8+ $\alpha$ <sup>391</sup>. Histology analysis suggested that NK receptor-expressing CD4 T cells were present in follicular microstructures typical of rheumatoid synovitis<sup>391</sup>. In this analysis, CD8+ NK cells clustered separately from the memory CD4 T cell population and it is likely that findings by Warrington et al., is restricted by the technology used (flow cytometry) and thus the number of parameters available for immunophenotyping. In addition, Warrington et al., used manual gating to analyze the data which could overlook this interpretation.

It is hypothesized that CD8+ NK cells can mediate interactions with chronically activated and exhausted immune cell populations, particularly maintaining homeostasis between CD4+ CD28- T cells and exhausted CD8+ T cells<sup>414</sup>. In the absence of cytokine data, it is uncertain whether CD8+ NK cells are proinflammatory but the phenotype that has emerged is an increase of activation and inhibitory receptors suggests that they are capable of activating immune cell populations by an escape mechanism i.e. not through the traditional CD28 activation pathway but rather through NK cell receptors which span both killer activation receptors, killer-cell immunoglobulin-like receptors and CD8+ MHC class I signaling<sup>391</sup>. It could also be speculated that CD8+ NK cells are derived from CD4+ CD28- T cells, whereby CD28 is downregulated over time in response to chronic antigenic stimulation and acquires NK cell receptors as a mechanism to switch off T cell activation<sup>391</sup>.

The wider implication of CD8+ NK cells in autoimmune disease can be observed in multiple sclerosis (MS), where CD8+ NK cells were hypothesized to have an immunoregulatory role which limited recurrent T-cell driven demyelination in relapsing/remitting multiple sclerosis<sup>412</sup>. The phenotype of CD8+ NK cells were CD3-CD8+CD56+ and RNA- sequencing on this subset showed that these cells were associated with reduced future relapse risk following an initial demyelinating event<sup>412</sup>. To understand the underlying biological mechanism driving this immunoregulatory role, in vitro analysis of CD8+ NK cells was undertaken. CD8+ NK cells were incubated with CD4 or CD8 T cells derived from patients with MS to assess whether they had a suppressive effect<sup>412</sup>. It was observed that CD8+ NK cells did have a significant suppressive effect on autologous CD4 T cells activation and proliferation but the same effect was not observed on CD8+ T cells<sup>412</sup>. Considering this work, it is apparent that CD8+ NK cells may possess a multi-functional role capable of both immunoregulatory and pathologic functions in RA.

Finally, it is uncertain at this stage to consider whether CD8+ NK cells would make a good therapeutic target let alone whether they could be therapeutically targeted. There is growing evidence that the CD8+ cell populations spanning across conventional and non-conventional subsets are becoming more prominent in autoimmune disease as immunophenotyping studies delve further into the populations present in any given biological sample<sup>409</sup>. In terms of RA, an autoimmune disease characterized by its dysregulated memory CD4+ T cell populations, researchers are looking to identify other immune cell populations responsible for perpetuating the chronically activated phenotype resulting in downregulation of CD27 and CD28. It is not clear what would happen if these T cell subsets were directly targeted or whether balancing the homeostasis between immunosenescence and chronic activation can be targeted long term and what these implications would be. Hypothetically, CD8+ NK cells could be a therapeutic target by depleting their numbers in the joint. It would be anticipated that by depleting CD8+ NK cells, this would indirectly disrupt T cell activation. Further work that can build on these findings is encouraged as it is thought individuals who possess CD8+ NK cells may respond differently to exogenous as well as self antigens, resulting in chronic inflammation within the synovium, that may be resistant to existing therapeutic options in RA.

## Chapter 7 General Discussion

### 7.1 Summary of findings

The main aim of this study was to understand whether mass cytometry could reveal specific changes within immune cell populations which corresponded with a specific 'stage' of RA or associated with specific biological compartments demonstrated by SF or peripheral blood. Alongside this main aim, it was important to ensure that validation of the methodology approach in this study was optimised, providing a template for future study designs which aims to take future researchers from conception of study design to data analysis and being able to draw informed conclusions. This study began in January 2017, where accessibility to a mass cytometer was not widely available to researchers. This is also reflected in this study, as it was done in collaboration with UCB Pharma as University of Leeds did not have a mass cytometry facility. Since the commencement of this study, data generated from single cell immunophenotyping by mass cytometry has increased exponentially and becoming a more familiar topic in Immunology. The cytometry field continues to strive for best practice in data generation from cytometry experiments and this study was particularly attentive to these expectations. Clone validation of the antibodies used in this study, technical processing of samples for acquisition by mass cytometry and validation of automated bioinformatics scripts including the Diffcyt script pipeline took a significant amount of time however, this was necessary to ensure confidence in the data.

Significant immune cell changes were detected in peripheral blood of patients with early RA when compared to healthy donors. Differential abundance testing showed MAIT cells, CD8 NK cells, pDCs and NK cells were reduced in early RA compared to health and further differential expression testing revealed that activation markers including CD27, CD28, CD38, HLA-DR were either significantly increased or decreased on both innate (basophils, CD14+ monocytes, pDCs and neutrophils) and lymphoid (memory CD4 and CD8 T cells and NK cells) populations in early RA. Interestingly, a specific early RA immune phenotype did not distinguish patients who had not received treatment compared to established RA (on treatment). It was evident that a specific immune signature did not further distinguish different demographic or clinical parameters within the established RA peripheral blood cohort. However, it was also detected by MDS analysis which observes global staining similarities between conditions that established RA patients (regardless of treatment, timepoint or clinical measurements) did separate from healthy donors which supported the established concept of atypical immune interactions in RA. However, it is obvious that progression of RA manifests in the joint and analysis of SF may reveal further specific phenotypes which contribute to the chronic inflammatory phenotype. Analysis of three matched RA peripheral blood and SF samples were analysed to reveal chronically activated T cells in the memory CD4 T population demonstrated by the increased median expression of checkpoint inhibitors CTLA-4, and PD-1, and an increase in HLA-DR in RA SF compared to blood. A decrease in median expression of CD27 and CD28 in RA SF further confirmed that memory CD4 T cells are aberrantly and chronically activated, an observation that has been previously reported in RA and in a wider context, a phenomenon that has been observed in other autoimmune diseases<sup>333,415</sup>. Furthermore, PD-1 expression was notably increased on memory CD8 T cells in RA SF. Lastly, the CD8 NK cell population in RA SF revealed a distinct phenotype which showed an increased median expression of CD27, NKp44, CD28, PD-L1, HLA-DR and CD80 and a decrease in perforin median expression. CD8 NK cells had a phenotype that was distinctly different in the SF joint in comparison to peripheral blood providing evidence that the local joint microenvironment influences phenotype.

Therefore using mass cytometry, a novel and not widely available single cell technology, coupled with current bioinformatics analysis, this study has demonstrated that it is possible to discover novel immunophenotypes and replicate findings which have been previously reported<sup>157,161,261</sup>.

## **7.2 Limitations of the study**

It is recognised there are limitations to this work and thus the scope of the findings. Firstly, despite the sample size being larger than the studies typically reported for mass cytometry, it is accepted that the findings reported here, would need to be corroborated in a larger cohort. Furthermore the panel design would have benefited by inclusion of cytokine and chemokine markers. However, as the study consisted of unstimulated peripheral blood, including a cytokine and chemokine panel would not have been useful. However, if time permitted, protocols exist for processing stimulated, cryogenically preserved PBMCs on a mass cytometer. This would require extensive validation as is the case for flow cytometry however, it would provide further insight into which immune cell populations for example produce the highest amount of TNF which could in turn improve specificity of TNF inhibitors as opposed to existing TNF inhibitors which broadly inhibit TNF. Another limitation is that the immune cell populations identified as being statistically significant in early RA were not able to be pursued in more detail to characterise their functional properties. In section 7.3.2, it is outlined how this could be potentially addressed in future work.

### **7.2.1 Critical appraisal of mass cytometry**

It is also acknowledged that new techniques such as mass cytometry are complex requiring comprehensive knowledge of both the technical and logistical aspects of cytometry. Coupled with complex bioinformatics analytical pipelines, this is a challenging study especially for those at a junior level (postgraduate) whose familiarity is not as extensive as a specialised cytometrist. Technical challenges in this study were encountered including batch effects as a result of variable staining but it is hoped that more barcodes for samples to be included in 1 batch will become available which will significantly reduce this inconsistency. In addition, two markers CD56 and CD123 were removed from the panel as they were inconsistently expressed and were removed from the analysis process.

In terms of the bioinformatics pipeline used to analyse the data from mass cytometry in this study, it is apparent that there were limitations. Channel labelling is more specific and each script needs to be carefully studied to avoid encountering script discrepancies with the data as evidence with the marker IgD which was mistakenly labelled as '1gd'. The Diffcyt script calculated cell abundance which translates as percentage of cells included in the analysis from each sample. A criticism of this approach is that there is no absolute count of each cell population however, this study did not have scope to give an absolute count as samples were barcoded. As a follow up experiment, it would be worth providing absolute counts of cell populations particularly those populations which are pathologically expanded such as Tph cells.

Considering some of the limitations encountered in this study, it is worth considering the merits of mass cytometry in the wider context of single cell technologies. Undoubtedly mass cytometry has pioneered the multi-parametric single cell technologies for protein analysis. Mass cytometry has pushed the boundaries of cytometry allowing for unparalleled deep immunophenotyping of biological analysis however, this has set precedent for existing technologies namely flow cytometry to develop better cytometers that can compete with mass cytometry. This trend has become a common sight at conferences and technology demonstrations where flow cytometers are demonstrated with increasing parameters due to improved and stable fluorescent dyes and more

channels. Compensation correction has also improved with many flow cytometers offering automated compensation software reducing time taken to construct marker panels and most importantly, reducing the amount of sample required for controls.

Therefore, in an era where technology is advancing at a considerable pace, is mass cytometry worth it? In short, no. If existing flow cytometers were still limited to 20 markers, then mass cytometry would have a significant advantage but as already mentioned, flow cytometers are offering 40 colour channels and parameters. The cost of running a mass cytometer is not practical for many research budgets. The added complication of needing to adapt a single room for the running of the mass cytometer requires institutions to make considerable adaptations to existing facilities with an experienced and dedicated cytometrist to ensure smooth operation of the facility. From first hand experience, this is by no means a trivial problem but requires careful consideration if this is a technology that researchers would like access to in their research. The reason for this hesitancy is once again because multi-parametric flow cytometers are available that only require bench top space and no significant alterations to the existing laboratory structure <sup>416</sup>. Furthermore flow cytometers can offer more than just identification of cells by immune protein markers. This is evidenced by technologies such as ImageStream which also provides morphometric and photometric features of the examined cells during acquisition <sup>417</sup>. The company 10X Genomics offers multiomics phenotyping which combines single cell RNA sequencing with single cell immunophenotyping <sup>418</sup>. Another limitation of mass cytometry is that it is slow during acquisition, with a maximum of 500 events/second and cells are completely obliterated meaning there is no scope for cell sorting. With flow cytometry, cell sorting can be an option albeit that it is usually performed by an experienced cytometrist. If considering logistics, then flow sorting itself is an unduly complex procedure given the time taken to obtain a specific immune cell subset from a heterogeneous population of cells. There are technical kits which magnetically separate specific immune cells and can be independently performed by the researcher with minimal experience however, this approach does not always yield 100% purity for downstream in vitro experiments. It is anticipated with the explosion and advancements of technology, fluorescence activated cell sorting technology will be simplified and robust, in turn expediting the process that can be performed by any researcher, thus emancipating the cytometrist.

In an era of technological advancement it is accepted that researchers are faced with a daunting but equally generous prospect of choice when it comes to choosing the right technology for their research question. Naturally, as technologies evolve, some older technologies may fall out of favour. A technology that offers robust, specific data on a multi-omic level at a relatively inexpensive cost which is technically simple to operate will be favoured over most other technologies. Mass cytometry has undoubtedly raised the bar for multiparametric cytometry however, technologists developing flow cytometers have risen to the challenge. Imaging mass cytometry has also overcome the limitations of immunohistochemistry however, again other vendors are competing and providing alternative approaches which offer equivalent if not better results. Should a research institution already have mass cytometry as part of their facility then the use of this technology should continue as naturally the technology will itself be under continuous development to surpass its current offerings. The limitations of mass cytometry I have raised in this chapter are aimed for researchers who do not currently have access to a mass cytometer and are considering using this approach for their own research and/or contemplating the purchase of a mass cytometer. The limitations are stated to provide an honest evaluation of mass cytometry and not to actively discourage future researchers from using the technology. In essence, it is hoped that this discussion will encourage researchers to carefully consider their options, and explore the range of choices on offer without

being solely influenced by the popularity or novelty of the technology. This study has aimed to be transparent and informative about mass cytometry highlighting both its merits and disadvantages throughout however, given the progress made in single cell technologies, it is fair to summarise that mass cytometry no longer stands alone in a rapidly advancing field of multi-parametric single cell protein cytometry<sup>419</sup>.

### 7.3 Future direction

This study aimed to open further opportunities for future research. I would like to propose a few suggestions which could guide future research for RA which have been influenced by the findings from this work. The suggestions proposed also take into consideration practical and logistical considerations of studies to provide novel and insightful data that would aim to answer unmet clinical needs that remain in RA.

#### 7.3.1 Imaging whole synovial biopsy tissue

Interrogation of the immune landscape within the joint of a well characterised patient cohort may yield insightful findings than a blood cohort. If synovial biopsy tissue were obtainable, it would be better to immunophenotype whole biopsy tissue as opposed to disaggregating the biopsy. Whilst synovial biopsy immunophenotyping has been extensively performed by several groups including the AMP group and Pitzalis et al.,<sup>167,420</sup> these have been performed in disaggregated tissue or by low resolution immunohistochemistry, leaving the opportunity for whole tissue deep immunophenotyping unexplored. Technologies now exist to accommodate whole tissue immunophenotyping including the CODEX platform and imaging mass cytometry<sup>174,421,422</sup> which allows the researcher to interrogate whether specific cell populations are infiltrating or tissue resident.

Isolation of specific cell populations and here a first suggestion would be to isolate Tph cells as evidence suggests that their expansion in the RA joint correlates specifically with RA pathology. In addition hypotheses that should be interrogated is the effect of treatment on immune cell populations with particular focus on the Tph cells in synovial tissue. This would embellish our understanding of the effects existing treatment including a combination of DMARDs and biologics has on the local tissue microenvironment. Questions aiming to specifically understand what constitutes refractory RA and whether patients who are deemed refractory (after failing two mechanistically different biologics) can be better defined through specific immune signatures.

The other aspect of characterising synovial biopsy is that it includes non-immune cell populations namely fibroblasts. Fibroblasts have received attention over the decades in particular from the research groups in Glasgow and Birmingham<sup>135,342,423–425</sup>. Whilst subsets of fibroblasts have been characterised through transcriptional and proteomic single cell studies, it would be intriguing to observe whether refractory RA could be explained by fibroblast pathotypes<sup>425</sup>.

Finally, synovial biopsy research continues to evolve and develop current efforts in new precision medicine approaches in RA<sup>426</sup>. The development of an organ-on-a-chip microfluidic culture device that simulates the microarchitecture and functions of living human organs offers hope to study whole organ functions. Fluidigm in collaboration with the FLAMIN-GO consortium have created synovia-on-a-chip that will aim to capture the complexities of the rheumatic joint with the aim of predicting treatment response and detecting biomarkers which ultimately would lead to improved

clinical management of RA <sup>426</sup>. This research will undoubtedly yield breakthrough findings and overcome current limitations with animal models and biopsy tissue.

### **7.3.2 Blood immunophenotyping early RA**

However, it is recognised that scaling up a synovial biopsy cohort is difficult in a short period of time ( $\leq 3$  years) and SF is not routinely available in the clinic which means that a blood is still a useful and logistically viable biological sample. As this study highlighted, there may be more insight into aberrant immune subsets to be gathered at the earliest opportunity. The earliest opportunity without the study becoming implemented as community testing is when the patient presents symptoms such as persistent joint pain to the general practitioner. A blood sample at this timepoint would be useful to probe the immune landscape and understand which patients go onto a confirmed diagnosis of RA and which patients either spontaneously resolve symptoms or develop non-RA inflammatory arthritis. The intention of this would be to develop a comprehensive insight into the immune dysregulation that is occurring. This study has already identified specific immune cell population changes which occur in early RA however, these patients already have a clinically confirmed diagnosis of RA. The discovery of specific changes in immune cell populations serves as a proof-of-concept that interrogating the RA immune environment is prudent.

Neutrophils are receiving renewed attention in terms of their functional contribution to RA pathology <sup>427–429</sup>. It would be recommended to include more neutrophil markers to identify whether there are subsets of neutrophils that reveal pathologic immune dysregulation including CD66b, CD44, CD15 and CD62L which would refine the neutrophil population and help to understand the role of neutrophils in RA and health. Neutrophils have often been ignored due to the general assumption that they are short-lived in vivo and are difficult to study in vitro. Moreover, neutrophils are often overlooked as they are in abundance and often masks less abundant cell populations. However, our understanding of neutrophils has advanced and they are considered to be one of the most important effector cells of the innate immune system <sup>430</sup>. From this work, neutrophils were seen to drive immunophenotypic distinction between early RA and health and it is thought that this distinction may be due to altered functional characteristics of neutrophils which could be further detected by in vitro assays for example which measure reactive oxygen species, neutrophil extracellular traps and apoptosis. This study did not show a significant change in the neutrophil phenotype within the joint although Wright et al. showed through transcriptomic analysis that neutrophils had an altered phenotype including raised reactive oxygen species and chemokine production <sup>356</sup>. It is worth pursuing these findings in a larger cohort of patients to ensure they are replicable and if it is not possible to obtain SF then blood neutrophils would still provide a significant insight into their pathologic role. Raised ESR and CRP are included in the RA diagnostic criteria although they are not specific to RA. Identifying neutrophil subsets that could correlate with inflammation may provide an added dimension with the intention to capture RA at its earliest opportunity.

### **7.3.3 Deep immunophenotyping mucosal tissue could provide new insight into pre-clinical RA**

It is also important to note that whilst immune dysregulation or a breakdown in tolerance are both concepts that are synonymous with RA, it is evident that this mechanism is only partially understood. It could be probable that before immune cells assume a permanent dysregulated phenotype, the cause of this imbalance could result from gut/oral microbiota dysbiosis, unresolved viral infection/pathology or other environmental factors such as dust inhalation or smoking <sup>1</sup>. What

links these concept is that they all indicate a mucosal origin, a concept which is receiving more traction, particularly for patients with seropositive RA <sup>431</sup>. A quick literature search suggests that RA mucosal tissue has not been deeply immunophenotyped by mass cytometry, offering both an insightful and novel approach in understanding overall immune dysregulation in RA. Nasal tissue biopsies have been successfully processed for acquisition by mass cytometry and offer an accessible tissue biopsy option <sup>432</sup>. A pneumococcal challenge study consisting of 20 patients observed B cells and MAIT cells were higher in non-colonized participants compared to colonized participants which correlated with blood MAIT cells which suggest protection against infection <sup>432</sup>. In this study, it was identified that MAIT cell abundance was decreased in RA compared to health however, additional analysis within RA SF did not show an increased number of MAIT cells within the joint microenvironment. This evidence therefore does not support the theory that MAIT cells from the blood are migrating to the joint and given that MAIT cells are native to mucosal sites, this future work would provide insightful data at the mucosal surface.

### 7.3.4 Mutual pathways between RA and Covid-19?

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) also known as Covid-19, is a virus that primarily affects the respiratory system which has disrupted the world and continues to pose a significant challenge to human health <sup>433,434</sup>. Earlier data focused on understanding which cells were implicated by the virus and it was observed that airway epithelial cells, alveolar epithelial cells, vascular endothelial cells and macrophages were targeted by the virus as they all express angiotensin converting enzyme 2, the receptor targeted by the virus to gain entry into the host cell <sup>435</sup>. It is apparent that Covid-19 has similarities with RA despite the two diseases differing in terms of their primary manifestation <sup>436,437</sup>. Treatments used in RA also being effective in patients with Covid-19 in particular IL-6 inhibitors and dexamethasone which have improved survival outcomes for patients who were severely ill with Covid-19 <sup>438,439</sup>.

Interrogating the immune landscape of Covid-19 has resulted in numerous publications however, there is still no specific biomarker or inflammatory endotype which is able to differentiate between mild and severe disease. It is possible to discern from the existing literature on Covid-19 that the innate immune system is dysregulated and within these innate immune cell populations, immune signatures are starting to emerge. Using mass cytometry, Chevrier et al. aimed to detect specific phenotypes in mild and severe Covid-19 compared to healthy donors <sup>440</sup>. Chevrier et al. observed that CD169+ monocytes associated with a IFN- $\gamma$ + MCP-2 signature which rapidly followed after symptom onset and a persistent inflammatory phenotype in patients with severe disease, dominated by high CCL3 and CCL4 abundance which correlated with the re-appearance of CD16+ monocytes <sup>440</sup>. Another study immunophenotyped Covid-19 post-mortem tissue from three patients by imaging mass cytometry revealed monocyte, macrophage and DC infiltration in the lung, kidney, intestine and liver <sup>441</sup>. Of note was CD11b+ macrophages and CD11c+ DCs infiltrated the liver, lungs and intestine and overproduced the immunosuppressive cytokine IL-10 <sup>441</sup>. Furthermore CD11b+ macrophages and CD11c+ DCs in the lungs and intestines of patients did not express HLA-DR but did produce higher expression of TNF in lungs and intestine <sup>441</sup>.

It is evident that the monocyte/macrophage axis is implicated in Covid-19 and this has led to further investigation into whether there is an overlap in the immune subsets that perpetuate both Covid-19 and RA. Single cell RNA sequencing of bronchoalveolar lavage fluid from 6 patients with severe Covid-19, discovered abnormally low numbers of resident alveolar macrophages although two specific macrophage subsets were increased immunophenotyped as FCN<sup>pos</sup> and FCN<sup>pos</sup>SPP1<sup>pos</sup> and

shared the same pathogenic mechanisms with pro-inflammatory macrophage clusters that drive synovitis characterized as CD48<sup>high</sup> S100A12<sup>pos</sup> and CD48<sup>pos</sup> SPP1<sup>pos</sup> <sup>442,443</sup>. Similar transcriptomic profiles were identified between healthy alveolar macrophages and healthy and remission synovial tissue lining layer macrophages providing additional evidence that mechanisms between Covid-19 and RA overlap.

Therefore, further exploration of innate immune cells and their role in Covid-19 pathology may also be useful in RA. Innate immune cell populations have often been overlooked which may explain why existing treatments in RA have limited efficacy. Questions that look to embellish our current understanding of innate immune cells and their implications in disease are desperately required and definitive answers about whether innate immune cells have functions that are assumed to traditionally be performed by adaptive immune cells need to be answered. It is not possible to provide a detailed discussion of current developments in Covid-19 as this is a rapidly evolving field. It is acknowledged that the unknown questions in Covid-19 strike a similar tone with RA. Questions such as being able to identify specific immune signatures that correlate with disease severity could apply to either disease and the effort to find specific, effective treatment remains a priority. The pathologic contribution of monocytes and macrophages is a central theme to both RA and Covid-19 pathologies. Research into both innate and adaptive immune cells will be necessary but this binary classification is reductionist when evidence exists there is considerable overlap between innate and adaptive functions. A specific question that should be pursued is better understanding of trained immunity, which is the concept of memory in innate cells. The concept of 'immune memory' is that upon a second encounter with a pathogen, certain immune cells, and traditionally cells classified as adaptive, can mount a faster and efficient response to eliminate the pathogen. In the context of Covid-19, if it could be identified which innate immune cells possess trained immunity, vaccines should be developed targeting these cells and discussed in two reviews <sup>444,445</sup>. Developing treatments in RA which target the innate immune cell population rather than their effector molecule production could improve specificity of treatment. Finally, if innate immune cells possess trained immunity, is this directed towards autoantigen recognition and thus contribute in chronic inflammation? There is insufficient evidence to draw any definitive conclusions at present and it is hoped that this discussion has identified some key areas which need further work.

## **7.4 Concluding remarks**

This study aimed to go beyond existing studies performed by mass cytometry and use the most current bioinformatics analysis pipelines to analyse the data. Furthermore the identification of specific immune cell populations in early RA peripheral blood and future efforts should use these findings for functional characterisation. This study also confirmed the chronically activated phenotype within the memory CD4 T cell population in RA SF and it is hoped that this work will stimulate further research using multi-parametric single cell technologies.

## Bibliography

1. Smolen, J. S. *et al.* Rheumatoid arthritis. *Nat. Rev. Dis. Primer* **4**, 18001 (2018).
2. Waaler, E. On the occurrence of a factor in human serum activating the specific agglutination of sheep blood corpuscles. 1939. *APMIS Acta Pathol. Microbiol. Immunol. Scand.* **115**, 422–438; discussion 439 (2007).
3. Song, Y. W. & Kang, E. H. Autoantibodies in rheumatoid arthritis: rheumatoid factors and anticitrullinated protein antibodies. *QJM Int. J. Med.* **103**, 139–146 (2010).
4. Ingegnoli, F., Castelli, R. & Gualtierotti, R. Rheumatoid Factors: Clinical Applications. *Dis. Markers* **35**, 727–734 (2013).
5. Schellekens, G. A. *et al.* The diagnostic properties of rheumatoid arthritis antibodies recognizing a cyclic citrullinated peptide. *Arthritis Rheum.* **43**, 155–163 (2000).
6. Burska, A. N. *et al.* Autoantibodies to Posttranslational Modifications in Rheumatoid Arthritis. *Mediators Inflamm.* **2014**, e492873 (2014).
7. Aletaha, D. *et al.* 2010 Rheumatoid arthritis classification criteria: an American College of Rheumatology/European League Against Rheumatism collaborative initiative. *Ann. Rheum. Dis.* **69**, 1580–1588 (2010).
8. Kurowska, W., Kuca-Warnawin, E. H., Radzikowska, A. & Maśliński, W. The role of anti-citrullinated protein antibodies (ACPA) in the pathogenesis of rheumatoid arthritis. *Cent.-Eur. J. Immunol.* **42**, 390–398 (2017).
9. Forslind, K. *et al.* Prediction of radiological outcome in early rheumatoid arthritis in clinical practice: role of antibodies to citrullinated peptides (anti-CCP). *Ann. Rheum. Dis.* **63**, 1090–1095 (2004).
10. Böhler, C., Radner, H., Smolen, J. S. & Aletaha, D. Serological changes in the course of traditional and biological disease modifying therapy of rheumatoid arthritis. *Ann. Rheum. Dis.* **72**, 241–244 (2013).
11. Aletaha, D. & Blüml, S. Therapeutic implications of autoantibodies in rheumatoid arthritis. *RMD Open* **2**, e000009 (2016).
12. Smith, M. D. The Normal Synovium. *Open Rheumatol. J.* **5**, 100–106 (2011).

13. Orr, C. *et al.* Synovial tissue research: a state-of-the-art review. *Nat. Rev. Rheumatol.* **13**, 463 (2017).
14. Ouboussad, L., Burska, A. N., Melville, A. & Buch, M. H. Synovial Tissue Heterogeneity in Rheumatoid Arthritis and Changes With Biologic and Targeted Synthetic Therapies to Inform Stratified Therapy. *Front. Med.* **6**, 45 (2019).
15. Myasoedova, E., Crowson, C. S., Kremers, H. M., Therneau, T. M. & Gabriel, S. E. Is the incidence of rheumatoid arthritis rising?: Results from Olmsted County, Minnesota, 1955–2007. *Arthritis Rheum.* **62**, 1576–1582 (2010).
16. Helmick, C. G. *et al.* Estimates of the prevalence of arthritis and other rheumatic conditions in the United States: Part I. *Arthritis Rheum.* **58**, 15–25 (2008).
17. Abhishek, A. *et al.* Rheumatoid arthritis is getting less frequent-results of a nationwide. *Arthritis Res.* **4**, S265-72 (2002).
18. Arnett, F. C. *et al.* The american rheumatism association 1987 revised criteria for the classification of rheumatoid arthritis. *Arthritis Rheum.* **31**, 315–324 (1988).
19. Gonzalez, A. *et al.* Mortality Trends in Rheumatoid Arthritis: The role of Rheumatoid Factor. *J. Rheumatol.* **35**, 1009–1014 (2008).
20. Nikiphorou, E. *et al.* Cardiovascular risk factors and outcomes in early rheumatoid arthritis: a population-based study. *Heart* **106**, 1566–1572 (2020).
21. Ziade, N. *et al.* Prevalence and pattern of comorbidities in chronic rheumatic and musculoskeletal diseases: the COMORD study. *Sci. Rep.* **10**, 7683 (2020).
22. Asquith, D. L., Miller, A. M., McInnes, I. B. & Liew, F. Y. Animal models of rheumatoid arthritis. *Eur. J. Immunol.* **39**, 2040–2044 (2009).
23. Lwin, M. N., Serhal, L., Holroyd, C. & Edwards, C. J. Rheumatoid Arthritis: The Impact of Mental Health on Disease: A Narrative Review. *Rheumatol. Ther.* **7**, 457–471 (2020).
24. Coffey, C. M., Crowson, C. S., Myasoedova, E., Matteson, E. L. & Davis, J. M. Evidence of Diagnostic and Treatment Delay in Seronegative Rheumatoid Arthritis: Missing the Window of Opportunity. *Mayo Clin. Proc.* **94**, 2241–2248 (2019).

25. MacGregor, A. J. *et al.* Characterizing the quantitative genetic contribution to rheumatoid arthritis using data from twins. *Arthritis Rheum.* **43**, 30–37 (2000).
26. Stastny, P. Mixed lymphocyte cultures in rheumatoid arthritis. *J. Clin. Invest.* **57**, 1148–1157 (1976).
27. Gregersen, P. K., Silver, J. & Winchester, R. J. The shared epitope hypothesis. an approach to understanding the molecular genetics of susceptibility to rheumatoid arthritis. *Arthritis Rheum.* **30**, 1205–1213 (1987).
28. Montcel, S. T. du *et al.* New classification of HLA–DRB1 alleles supports the shared epitope hypothesis of rheumatoid arthritis susceptibility. *Arthritis Rheum.* **52**, 1063–1068 (2005).
29. Suzuki, A. *et al.* Functional haplotypes of PADI4, encoding citrullinating enzyme peptidylarginine deiminase 4, are associated with rheumatoid arthritis. *Nat. Genet.* **34**, 395–402 (2003).
30. Barton, A. *et al.* A functional haplotype of the PADI4 gene associated with rheumatoid arthritis in a Japanese population is not associated in a United Kingdom population. *Arthritis Rheum.* **50**, 1117–1121 (2004).
31. Ikari, K. *et al.* Association between PADI4 and rheumatoid arthritis: A replication study. *Arthritis Rheum.* **52**, 3054–3057 (2005).
32. Begovich, A. B. *et al.* A Missense Single-Nucleotide Polymorphism in a Gene Encoding a Protein Tyrosine Phosphatase (PTPN22) Is Associated with Rheumatoid Arthritis. *Am. J. Hum. Genet.* **75**, 330–337 (2004).
33. Bottini, N. *et al.* A functional variant of lymphoid tyrosine phosphatase is associated with type I diabetes. *Nat. Genet.* **36**, 337–338 (2004).
34. Hinks, A. *et al.* Association between the PTPN22 gene and rheumatoid arthritis and juvenile idiopathic arthritis in a UK population: Further support that PTPN22 is an autoimmunity gene. *Arthritis Rheum.* **52**, 1694–1699 (2005).
35. Karlson, E. W. *et al.* Cumulative association of 22 genetic variants with seropositive rheumatoid arthritis risk. *Ann. Rheum. Dis.* **69**, 1077–1085 (2010).

36. Viatte, S. *et al.* Replication of Associations of Genetic Loci Outside the HLA Region With Susceptibility to Anti–Cyclic Citrullinated Peptide–Negative Rheumatoid Arthritis. *Arthritis Rheumatol.* **68**, 1603–1613 (2016).
37. Cribbs, A. P. *et al.* Treg Cell Function in Rheumatoid Arthritis Is Compromised by CTLA-4 Promoter Methylation Resulting in a Failure to Activate the Indoleamine 2,3-Dioxygenase Pathway: Defective Treg Cell Function in RA and the IDO Pathway. *Arthritis Rheumatol.* **66**, 2344–2354 (2014).
38. Zhang, Y. *et al.* Genome-wide DNA methylation analysis identifies hypomethylated genes regulated by FOXP3 in human regulatory T cells. *Blood* **122**, 2823–2836 (2013).
39. Mookherjee, N. & El-Gabalawy, H. S. High degree of correlation between whole blood and PBMC expression levels of miR-155 and miR-146a in healthy controls and rheumatoid arthritis patients. *J. Immunol. Methods* **400–401**, 106–110 (2013).
40. Stanczyk, J. *et al.* Altered expression of MicroRNA in synovial fibroblasts and synovial tissue in rheumatoid arthritis. *Arthritis Rheum.* **58**, 1001–1009 (2008).
41. Hutchinson, D. Heavy cigarette smoking is strongly associated with rheumatoid arthritis (RA), particularly in patients without a family history of RA. *Ann. Rheum. Dis.* **60**, 223–227 (2001).
42. Di Giuseppe, D., Discacciati, A., Orsini, N. & Wolk, A. Cigarette smoking and risk of rheumatoid arthritis: a dose-response meta-analysis. *Arthritis Res. Ther.* **16**, R61 (2014).
43. Padyukov, L., Silva, C., Stolt, P., Alfredsson, L. & Klareskog, L. A gene–environment interaction between smoking and shared epitope genes in HLA–DR provides a high risk of seropositive rheumatoid arthritis. *Arthritis Rheum.* **50**, 3085–3092 (2004).
44. Roelsgaard, I. K. *et al.* Smoking cessation is associated with lower disease activity and predicts cardiovascular risk reduction in rheumatoid arthritis patients. *Rheumatology* **59**, 1997–2004 (2020).
45. Stolt, P. *et al.* Silica exposure is associated with increased risk of developing rheumatoid arthritis: results from the Swedish EIRA study. *Ann. Rheum. Dis.* **64**, 582–586 (2005).

46. Too, C. L. *et al.* Occupational exposure to textile dust increases the risk of rheumatoid arthritis: results from a Malaysian population-based case–control study. *Ann. Rheum. Dis.* **75**, 997–1002 (2016).
47. Naranjo, A. *et al.* Smokers and non smokers with rheumatoid arthritis have similar clinical status: data from the multinational QUEST-RA database. *Clin. Exp. Rheumatol.* **28**, 820–827 (2010).
48. Straub, R. H. The Complex Role of Estrogens in Inflammation. *Endocr. Rev.* **28**, 521–574 (2007).
49. Kay, A. & Bach, F. SUBFERTILITY BEFORE AND AFTER THE DEVELOPMENT OF RHEUMATOID ARTHRITIS IN WOMEN. *Ann. Rheum. Dis.* **24**, 169–173 (1965).
50. Zhang, X. *et al.* The oral and gut microbiomes are perturbed in rheumatoid arthritis and partly normalized after treatment. *Nat. Med.* **21**, 895–905 (2015).
51. Wells, P. M., Williams, F. M. K., Matey-Hernandez, M. L., Menni, C. & Steves, C. J. ‘RA and the microbiome: do host genetic factors provide the link? *J. Autoimmun.* **99**, 104–115 (2019).
52. Otten, H. G. *et al.* IgA rheumatoid factor in mucosal fluids and serum of patients with rheumatoid arthritis: immunological aspects and clinical significance. *Clin. Exp. Immunol.* **90**, 256–259 (2008).
53. Scher, J. U. *et al.* Expansion of intestinal *Prevotella copri* correlates with enhanced susceptibility to arthritis. *eLife* **2**, e01202 (2013).
54. Ljung, L. & Rantapää-Dahlqvist, S. Abdominal obesity, gender and the risk of rheumatoid arthritis – a nested case–control study. *Arthritis Res. Ther.* **18**, 277 (2016).
55. Maxwell, J. R., Gowers, I. R., Moore, D. J. & Wilson, A. G. Alcohol consumption is inversely associated with risk and severity of rheumatoid arthritis. *Rheumatology* **49**, 2140–2146 (2010).
56. Firestein, G. S. & McInnes, I. B. Immunopathogenesis of Rheumatoid Arthritis. *Immunity* **46**, 183–196 (2017).
57. McInnes, I. B. & Schett, G. The Pathogenesis of Rheumatoid Arthritis. *N. Engl. J. Med.* **365**, 2205–2219 (2011).
58. Johnsson, H. & Najm, A. Synovial biopsies in clinical practice and research: current developments and perspectives. *Clin. Rheumatol.* (2020) doi:10.1007/s10067-020-05512-7.

59. Brennan, F. M. & McInnes, I. B. Evidence that cytokines play a role in rheumatoid arthritis. *J. Clin. Invest.* **118**, 3537–3545 (2008).
60. Brennan, F. M., Chantry, D., Jackson, A., Maini, R. & Feldmann, M. Inhibitory effect of TNF alpha antibodies on synovial cell interleukin-1 production in rheumatoid arthritis. *Lancet Lond. Engl.* **2**, 244–247 (1989).
61. Haworth, C. *et al.* Expression of granulocyte-macrophage colony-stimulating factor in rheumatoid arthritis: regulation by tumor necrosis factor-alpha. *Eur. J. Immunol.* **21**, 2575–2579 (1991).
62. Butler, D. M., Maini, R. N., Feldmann, M. & Brennan, F. M. Modulation of proinflammatory cytokine release in rheumatoid synovial membrane cell cultures. Comparison of monoclonal anti TNF-alpha antibody with the interleukin-1 receptor antagonist. *Eur. Cytokine Netw.* **6**, 225–230 (1995).
63. Feldmann, M. & Maini, R. N. TNF defined as a therapeutic target for rheumatoid arthritis and other autoimmune diseases. *Nat. Med.* **9**, 1245–1250 (2003).
64. Seyler, T. M. *et al.* BLYS and APRIL in rheumatoid arthritis. *J. Clin. Invest.* **115**, 3083–3092 (2005).
65. Shabgah, A. G., Shariati-Sarabi, Z., Tavakkol-Afshari, J. & Mohammadi, M. The role of BAFF and APRIL in rheumatoid arthritis. *J. Cell. Physiol.* **234**, 17050–17063 (2019).
66. Zhou, B. *et al.* Therapeutic effects of a novel BAFF blocker on arthritis. *Signal Transduct. Target. Ther.* **4**, 1–12 (2019).
67. Papadaki, M. *et al.* New Insights for RANKL as a Proinflammatory Modulator in Modeled Inflammatory Arthritis. *Front. Immunol.* **10**, (2019).
68. Takeuchi, T. *et al.* Effects of the anti-RANKL antibody denosumab on joint structural damage in patients with rheumatoid arthritis treated with conventional synthetic disease-modifying antirheumatic drugs (DESIRABLE study): a randomised, double-blind, placebo-controlled phase 3 trial. *Ann. Rheum. Dis.* **78**, 899–907 (2019).
69. Choy, E. H. *et al.* Translating IL-6 biology into effective treatments. *Nat. Rev. Rheumatol.* **16**, 335–345 (2020).

70. Guerne, P. A., Zuraw, B. L., Vaughan, J. H., Carson, D. A. & Lotz, M. Synovium as a source of interleukin 6 in vitro. Contribution to local and systemic manifestations of arthritis. *J. Clin. Invest.* **83**, 585–592 (1989).
71. Takagi, N. *et al.* Blockage of interleukin-6 receptor ameliorates joint disease in murine collagen-induced arthritis. *Arthritis Rheum.* **41**, 2117–2121 (1998).
72. Choy, E. & Rose-John, S. Interleukin-6 as a Multifunctional Regulator: Inflammation, Immune Response, and Fibrosis. *J. Scleroderma Relat. Disord.* **2**, S1–S5 (2017).
73. Favalli, E. G. Understanding the Role of Interleukin-6 (IL-6) in the Joint and Beyond: A Comprehensive Review of IL-6 Inhibition for the Management of Rheumatoid Arthritis. *Rheumatol. Ther.* **7**, 473–516 (2020).
74. Mateen, S., Zafar, A., Moin, S., Khan, A. Q. & Zubair, S. Understanding the role of cytokines in the pathogenesis of rheumatoid arthritis. *Clin. Chim. Acta* **455**, 161–171 (2016).
75. Yap, H.-Y. *et al.* Pathogenic Role of Immune Cells in Rheumatoid Arthritis: Implications in Clinical Treatment and Biomarker Development. *Cells* **7**, 161 (2018).
76. Volkov, M., Schie, K. A. van & Woude, D. van der. Autoantibodies and B Cells: The ABC of rheumatoid arthritis pathophysiology. *Immunol. Rev.* **294**, 148–163 (2020).
77. Giltiay, N. V., Chappell, C. P. & Clark, E. A. B-cell selection and the development of autoantibodies. *Arthritis Res. Ther.* **14 Suppl 4**, S1 (2012).
78. Samuels, J., Ng, Y.-S., Coupillaud, C., Paget, D. & Meffre, E. Impaired early B cell tolerance in patients with rheumatoid arthritis. *J. Exp. Med.* **201**, 1659–1667 (2005).
79. Menard, L. *et al.* The PTPN22 allele encoding an R620W variant interferes with the removal of developing autoreactive B cells in humans. *J. Clin. Invest.* **121**, 3635–3644 (2011).
80. Bugatti *et al.* - 2014 - B Cells in Rheumatoid Arthritis From Pathogenic P.pdf.
81. Humby, F. *et al.* Ectopic Lymphoid Structures Support Ongoing Production of Class-Switched Autoantibodies in Rheumatoid Synovium. *PLOS Med.* **6**, e1 (2009).
82. Bombardieri, M., Lewis, M. & Pitzalis, C. Ectopic lymphoid neogenesis in rheumatic autoimmune diseases. *Nat. Rev. Rheumatol.* **13**, 141–154 (2017).

83. Fox, D. A. The role of T cells in the immunopathogenesis of rheumatoid arthritis. New perspectives. *Arthritis Rheum.* **40**, 598–609 (1997).
84. Janossy, G. *et al.* RHEUMATOID ARTHRITIS: A DISEASE OF T-LYMPHOCYTE/MACROPHAGE IMMUNOREGULATION. *The Lancet* **318**, 839–842 (1981).
85. Chemin, K., Gerstner, C. & Malmström, V. Effector Functions of CD4+ T Cells at the Site of Local Autoimmune Inflammation—Lessons From Rheumatoid Arthritis. *Front. Immunol.* **10**, (2019).
86. Dolhain, R. J. E. M., Heiden, A. N. van der, Haar, N. T. ter, Breedveld, F. C. & Miltenburg, A. M. M. Shift toward T lymphocytes with a T helper 1 cytokine-secretion profile in the joints of patients with rheumatoid arthritis. *Arthritis Rheum.* **39**, 1961–1969 (1996).
87. Miossec, P. Local and systemic effects of IL-17 in joint inflammation: a historical perspective from discovery to targeting. *Cell. Mol. Immunol.* **18**, 860–865 (2021).
88. Aarvak, T., Chabaud, M., Miossec, P. & Natvig, J. B. IL-17 Is Produced by Some Proinflammatory Th1/Th0 Cells But Not by Th2 Cells. *J. Immunol.* **162**, 1246–1251 (1999).
89. Pelletier, M. *et al.* Evidence for a cross-talk between human neutrophils and Th17 cells. *Blood* **115**, 335–343 (2010).
90. Sakaguchi, S., Sakaguchi, N., Asano, M., Itoh, M. & Toda, M. Immunologic self-tolerance maintained by activated T cells expressing IL-2 receptor alpha-chains (CD25). Breakdown of a single mechanism of self-tolerance causes various autoimmune diseases. *J. Immunol. Baltim. Md 1950* **155**, 1151–1164 (1995).
91. Morgan, M. E. *et al.* CD25+ cell depletion hastens the onset of severe disease in collagen-induced arthritis. *Arthritis Rheum.* **48**, 1452–1460 (2003).
92. Morgan, M. E. *et al.* Effective treatment of collagen-induced arthritis by adoptive transfer of CD25+ regulatory T cells. *Arthritis Rheum.* **52**, 2212–2221 (2005).
93. Alunno, A. *et al.* Role of regulatory T cells in rheumatoid arthritis: facts and hypothesis. *Autoimmun. Highlights* **1**, 45–51 (2010).
94. Ziegler-Heitbrock, L. *et al.* Nomenclature of monocytes and dendritic cells in blood. *Blood* **116**, e74-80 (2010).

95. Cooper, D. L. *et al.* FcγRIIIa expression on monocytes in rheumatoid arthritis: role in immune-complex stimulated TNF production and non-response to methotrexate therapy. *PLoS One* **7**, e28918 (2012).
96. Chara, L. *et al.* The number of circulating monocytes as biomarkers of the clinical response to methotrexate in untreated patients with rheumatoid arthritis. *J. Transl. Med.* **13**, 2 (2015).
97. van Furth, R. *et al.* The mononuclear phagocyte system: a new classification of macrophages, monocytes, and their precursor cells. *Bull. World Health Organ.* **46**, 845–852 (1972).
98. Yona, S. *et al.* Fate Mapping Reveals Origins and Dynamics of Monocytes and Tissue Macrophages under Homeostasis. *Immunity* **38**, 79–91 (2013).
99. Sica, A. & Mantovani, A. Macrophage plasticity and polarization: in vivo veritas. *J. Clin. Invest.* **122**, 787–795 (2012).
100. Haringman, J. J. Synovial tissue macrophages: a sensitive biomarker for response to treatment in patients with rheumatoid arthritis. *Ann. Rheum. Dis.* **64**, 834–838 (2005).
101. Kinne, R. W., Bräuer, R., Stuhl Müller, B., Palombo-Kinne, E. & Burmester, G.-R. Macrophages in rheumatoid arthritis. *Arthritis Res.* **2**, 189–202 (2000).
102. Kennedy, A., Fearon, U., Veale, D. J. & Godson, C. Macrophages in Synovial Inflammation. *Front. Immunol.* **2**, (2011).
103. Edilova, M. I., Akram, A. & Abdul-Sater, A. A. Innate immunity drives pathogenesis of rheumatoid arthritis. *Biomed. J.* S2319417020300986 (2020) doi:10.1016/j.bj.2020.06.010.
104. Lanier, L. L. & Sun, J. C. Do the terms innate and adaptive immunity create conceptual barriers? *Nat. Rev. Immunol.* **9**, 302–303 (2009).
105. Boraschi, D. & Italiani, P. Innate Immune Memory: Time for Adopting a Correct Terminology. *Front. Immunol.* **9**, (2018).
106. Netea, M. G., Latz, E., Mills, K. H. G. & O’Neill, L. A. J. Innate immune memory: a paradigm shift in understanding host defense. *Nat. Immunol.* **16**, 675–679 (2015).
107. Paludan, S. R., Pradeu, T., Masters, S. L. & Mogensen, T. H. Constitutive immune mechanisms: mediators of host defence and immune regulation. *Nat. Rev. Immunol.* **21**, 137–150 (2021).

108. Nygaard, G. & Firestein, G. S. Restoring synovial homeostasis in rheumatoid arthritis by targeting fibroblast-like synoviocytes. *Nat. Rev. Rheumatol.* **16**, 316–333 (2020).
109. Hasegawa, T. *et al.* Identification of a novel arthritis-associated osteoclast precursor macrophage regulated by FoxM1. *Nat. Immunol.* **20**, 1631–1643 (2019).
110. Schett, G. Cells of the synovium in rheumatoid arthritis. Osteoclasts. *Arthritis Res. Ther.* **9**, 203 (2007).
111. Weinblatt, M. E. Methotrexate in Rheumatoid Arthritis: A Quarter Century of Development. *Trans. Am. Clin. Climatol. Assoc.* **124**, 16–25 (2013).
112. Wang, W., Zhou, H. & Liu, L. Side effects of methotrexate therapy for rheumatoid arthritis: A systematic review. *Eur. J. Med. Chem.* **158**, 502–516 (2018).
113. Upchurch, K. S. & Kay, J. Evolution of treatment for rheumatoid arthritis. *Rheumatology* **51**, vi28–vi36 (2012).
114. Köhler, B. M., Günther, J., Kaudewitz, D. & Lorenz, H.-M. Current Therapeutic Options in the Treatment of Rheumatoid Arthritis. *J. Clin. Med.* **8**, 938 (2019).
115. MA, X. & XU, S. TNF inhibitor therapy for rheumatoid arthritis. *Biomed. Rep.* **1**, 177–184 (2013).
116. Smolen, J. S. *et al.* Treating rheumatoid arthritis to target: recommendations of an international task force. *Ann. Rheum. Dis.* **69**, 631–637 (2010).
117. Pinals, R. S., Masi, A. T. & Larsen, R. A. Preliminary criteria for clinical remission in rheumatoid arthritis. *Arthritis Rheum.* **24**, 1308–1315 (1981).
118. 1\_economic\_burden\_of\_ra\_final\_30\_3\_10.pdf.
119. Nagy, G. & van Vollenhoven, R. F. Sustained biologic-free and drug-free remission in rheumatoid arthritis, where are we now? *Arthritis Res. Ther.* **17**, (2015).
120. Ajeganova, S. & Huizinga, T. Sustained remission in rheumatoid arthritis: latest evidence and clinical considerations. *Ther. Adv. Musculoskelet. Dis.* **9**, 249–262 (2017).
121. Baker, K. F. *et al.* Predicting drug-free remission in rheumatoid arthritis: A prospective interventional cohort study. *J. Autoimmun.* **105**, 102298 (2019).

122. Baker, K. F., Pratt, A. G., Thompson, B. & Isaacs, J. D. Let's not fool ourselves. In RA, the ACR/EULAR remission criteria are not perfect! *Ann. Rheum. Dis.* **76**, e12–e12 (2017).
123. Emery, P. The Dunlop-Dottridge Lecture: prognosis in inflammatory arthritis: the value of HLA genotyping and the oncological analogy. *J. Rheumatol.* **24**, 1436–1442 (1997).
124. Quinn, M. A. & Emery, P. Window of opportunity in early rheumatoid arthritis: Possibility of altering the disease process with early intervention. 5.
125. McGonagle, D. *et al.* The relationship between synovitis and bone changes in early untreated rheumatoid arthritis: A controlled magnetic resonance imaging study. *Arthritis Rheum.* **42**, 1706–1711 (1999).
126. Anderson, J. J., Wells, G., Verhoeven, A. C. & Felson, D. T. Factors predicting response to treatment in rheumatoid arthritis: The importance of disease duration. *Arthritis Rheum.* **43**, 22–29 (2000).
127. Firestein, G. S. Evolving concepts of rheumatoid arthritis. *Nature* **423**, 356–361 (2003).
128. Firestein, G. S. The disease formerly known as rheumatoid arthritis. *Arthritis Res. Ther.* **16**, 114 (2014).
129. Mankia, K. & Emery, P. Review: Preclinical Rheumatoid Arthritis: Progress Toward Prevention. *Arthritis Rheumatol.* **68**, 779–788 (2016).
130. Rakieh, C. *et al.* Predicting the development of clinical arthritis in anti-CCP positive individuals with non-specific musculoskeletal symptoms: a prospective observational cohort study. *Ann. Rheum. Dis.* **74**, 1659–1666 (2015).
131. Seyhan, A. A. *et al.* Novel biomarkers of a peripheral blood interferon signature associated with drug-naïve early arthritis patients distinguish persistent from self-limiting disease course. *Sci. Rep.* **10**, 8830 (2020).
132. Macías-Segura, N. *et al.* Transcriptional signature associated with early rheumatoid arthritis and healthy individuals at high risk to develop the disease. *PLOS ONE* **13**, e0194205 (2018).
133. Cooles, F. A. H. *et al.* The interferon gene signature is increased in patients with early treatment-naïve rheumatoid arthritis and predicts a poorer response to initial therapy. *J. Allergy Clin. Immunol.* **141**, 445–448.e4 (2018).

134. Lewis, M. J. *et al.* Molecular Portraits of Early Rheumatoid Arthritis Identify Clinical and Treatment Response Phenotypes. *Cell Rep.* **28**, 2455-2470.e5 (2019).
135. Raza, K. *et al.* 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–R795 (2005).
136. Ponchel, F. *et al.* T-cell subset abnormalities predict progression along the Inflammatory Arthritis disease continuum: implications for management. *Sci. Rep.* **10**, 3669 (2020).
137. Ridgley, L. A., Anderson, A. E. & Pratt, A. G. What are the dominant cytokines in early rheumatoid arthritis?: *Curr. Opin. Rheumatol.* **30**, 207–214 (2018).
138. Buch, M. H., Eyre, S. & McGonagle, D. Persistent inflammatory and non-inflammatory mechanisms in refractory rheumatoid arthritis. *Nat. Rev. Rheumatol.* **17**, 17–33 (2021).
139. Buch, M. H. Defining refractory rheumatoid arthritis. *Ann. Rheum. Dis.* **77**, 966–969 (2018).
140. Fitton, J. *et al.* FRI0090 SINGLE CENTRE COHORT OF REFRACTORY RHEUMATOID ARTHRITIS ALSO IDENTIFIES A RARE SUBGROUP OF MULTIPLE TARGETED THERAPY CLASS NON-RESPONSE. *Ann. Rheum. Dis.* **78**, 708–708 (2019).
141. Fitton, J., Melville, A. R., Emery, P., Nam, J. L. & Buch, M. H. Real-world single centre use of JAK inhibitors across the rheumatoid arthritis pathway. *Rheumatology* keaa858 (2020) doi:10.1093/rheumatology/keaa858.
142. Kearsley-Fleet, L. *et al.* Biologic refractory disease in rheumatoid arthritis: results from the British Society for Rheumatology Biologics Register for Rheumatoid Arthritis. *Ann. Rheum. Dis.* **77**, 1405–1412 (2018).
143. Melville, A. R., Kearsley-Fleet, L., Buch, M. H. & Hyrich, K. L. Understanding Refractory Rheumatoid Arthritis: Implications for a Therapeutic Approach. *Drugs* **80**, 849–857 (2020).
144. Evobrutinib fails to improve response rate in refractory RA patients. *rheumatology.medicinematters.com* <https://rheumatology.medicinematters.com/acr-2020/rheumatoid-arthritis-evobrutinib-refractory-ra-patients/18571522> (2020).
145. Misra, B. B., Langefeld, C., Olivier, M. & Cox, L. A. Integrated omics: tools, advances and future approaches. *J. Mol. Endocrinol.* **62**, R21–R45 (2019).

146. Cheung, P., Khatri, P., Utz, P. J. & Kuo, A. J. Single-cell technologies — studying rheumatic diseases one cell at a time. *Nat. Rev. Rheumatol.* **15**, 340–354 (2019).
147. McKinnon, K. M. Flow Cytometry: An Overview. *Curr. Protoc. Immunol.* **120**, 5.1.1-5.1.11 (2018).
148. Roederer, M. Compensation in flow cytometry. *Curr. Protoc. Cytom.* **Chapter 1**, Unit 1.14 (2002).
149. Feher, K., Kirsch, J., Radbruch, A., Chang, H.-D. & Kaiser, T. Cell population identification using fluorescence-minus-one controls with a one-class classifying algorithm. *Bioinformatics* **30**, 3372–3378 (2014).
150. Spitzer, M. H. & Nolan, G. P. Mass Cytometry: Single Cells, Many Features. *Cell* **165**, 780–791 (2016).
151. Bendall, S. C., Nolan, G. P., Roederer, M. & Chattopadhyay, P. K. A Deep Profiler’s Guide to Cytometry. *Trends Immunol.* **33**, 323–332 (2012).
152. Bandura, D. R. *et al.* Mass Cytometry: Technique for Real Time Single Cell Multitarget Immunoassay Based on Inductively Coupled Plasma Time-of-Flight Mass Spectrometry. *Anal. Chem.* **81**, 6813–6822 (2009).
153. Hartmann, F. J. & Bendall, S. C. Immune monitoring using mass cytometry and related high-dimensional imaging approaches. *Nat. Rev. Rheumatol.* **16**, 87–99 (2020).
154. Maecker, H. T. & Harari, A. Immune monitoring technology primer: flow and mass cytometry. *J. Immunother. Cancer* **3**, 44 (2015).
155. Olsen, L. R., Leipold, M. D., Pedersen, C. B. & Maecker, H. T. The anatomy of single cell mass cytometry data. *Cytometry A* **95**, 156–172 (2019).
156. Wilschefske, S. C. & Baxter, M. R. Inductively Coupled Plasma Mass Spectrometry: Introduction to Analytical Aspects. *Clin. Biochem. Rev.* **40**, 115–133 (2019).
157. Rao, D. A. *et al.* Pathologically expanded peripheral T helper cell subset drives B cells in rheumatoid arthritis. *Nature* **542**, 110–114 (2017).
158. Rao, D. A. The rise of peripheral T helper cells in autoimmune disease. *Nat. Rev. Rheumatol.* **15**, 453–454 (2019).

159. Rao, D. A. T Cells That Help B Cells in Chronically Inflamed Tissues. *Front. Immunol.* **9**, (2018).
160. Fonseka, C. Y., Rao, D. A. & Raychaudhuri, S. Leveraging blood and tissue CD4+ T cell heterogeneity at the single cell level to identify mechanisms of disease in rheumatoid arthritis. *Curr. Opin. Immunol.* **49**, 27–36 (2017).
161. Zhang, F. *et al.* Defining inflammatory cell states in rheumatoid arthritis joint synovial tissues by integrating single-cell transcriptomics and mass cytometry. *Nat. Immunol.* **20**, 928–942 (2019).
162. Mair, F. *et al.* The end of gating? An introduction to automated analysis of high dimensional cytometry data: Highlights. *Eur. J. Immunol.* **46**, 34–43 (2016).
163. Amir, E. D. *et al.* viSNE enables visualization of high dimensional single-cell data and reveals phenotypic heterogeneity of leukemia. *Nat. Biotechnol.* **31**, 545–552 (2013).
164. Rybakowska, P., Alarcón-Riquelme, M. E. & Marañón, C. Key steps and methods in the experimental design and data analysis of highly multi-parametric flow and mass cytometry. *Comput. Struct. Biotechnol. J.* **18**, 874–886 (2020).
165. Botta, C. *et al.* Flowct: A Semi-Automated Workflow for Deconvolution of Immunophenotypic Data and Objective Reporting on Large Datasets. *Blood* **134**, 4355–4355 (2019).
166. Weber, L. M., Nowicka, M., Sonesson, C. & Robinson, M. D. diffcyt: Differential discovery in high-dimensional cytometry via high-resolution clustering. *Commun. Biol.* **2**, 183 (2019).
167. Pitzalis, C., Kelly, S. & Humby, F. New learnings on the pathophysiology of RA from synovial biopsies. *Curr. Opin. Rheumatol.* **25**, 334–344 (2013).
168. Orange, D. E. *et al.* Identification of Three Rheumatoid Arthritis Disease Subtypes by Machine Learning Integration of Synovial Histologic Features and RNA Sequencing Data. *Arthritis Rheumatol.* **70**, 690–701 (2018).
169. Lewis, M. J. *et al.* Molecular Portraits of Early Rheumatoid Arthritis Identify Clinical and Treatment Response Phenotypes. *Cell Rep.* **28**, 2455–2470.e5 (2019).

170. Humby, F. *et al.* Synovial cellular and molecular signatures stratify clinical response to csDMARD therapy and predict radiographic progression in early rheumatoid arthritis patients. *Ann. Rheum. Dis.* **78**, 761–772 (2019).
171. Lliso-Ribera, G. *et al.* Synovial tissue signatures enhance clinical classification and prognostic/treatment response algorithms in early inflammatory arthritis and predict requirement for subsequent biological therapy: results from the pathobiology of early arthritis cohort (PEAC). *Ann. Rheum. Dis.* **78**, 1642–1652 (2019).
172. Dennis, G. *et al.* Synovial phenotypes in rheumatoid arthritis correlate with response to biologic therapeutics. *Arthritis Res. Ther.* **16**, R90 (2014).
173. Nerviani, A. *et al.* A Pauci-Immune Synovial Pathotype Predicts Inadequate Response to TNF $\alpha$ -Blockade in Rheumatoid Arthritis Patients. *Front. Immunol.* **11**, 845 (2020).
174. Goltsev, Y. *et al.* Deep Profiling of Mouse Splenic Architecture with CODEX Multiplexed Imaging. *Cell* **174**, 968-981.e15 (2018).
175. Gassen, S. V., Gaudilliere, B., Angst, M. S., Saeys, Y. & Aghaeepour, N. CytoNorm: A Normalization Algorithm for Cytometry Data. *Cytometry A* **n/a**, (2019).
176. Nowicka, M. *et al.* CyTOF workflow: differential discovery in high-throughput high-dimensional cytometry datasets. *F1000Research* **6**, 748 (2019).
177. Burel, J. G. *et al.* An integrated workflow to assess technical and biological variability of cell population frequencies in human peripheral blood by flow cytometry. *J. Immunol. Baltim. Md 1950* **198**, 1748–1758 (2017).
178. Alvarez, D. F., Helm, K., Degregori, J., Roederer, M. & Majka, S. Publishing flow cytometry data. *Am. J. Physiol. Lung Cell. Mol. Physiol.* **298**, L127-130 (2010).
179. *High-Dimensional Single Cell Analysis: Mass Cytometry, Multi-parametric Flow Cytometry and Bioinformatic Techniques.* (Springer-Verlag, 2014). doi:10.1007/978-3-642-54827-7.
180. McGuire, H. M. & Ashhurst, T. Mass Cytometry: Methods and Protocols. *Methods Mol. Biol.* (2019) doi:10.1007/978-1-4939-9454-0.

181. Nassar, A. F., Wisnewski, A. V. & Raddassi, K. Automation of sample preparation for mass cytometry barcoding in support of clinical research: protocol optimization. *Anal. Bioanal. Chem.* **409**, 2363–2372 (2017).
182. O’Gorman, W. E. *et al.* Mass cytometry identifies a distinct monocyte cytokine signature shared by clinically heterogeneous pediatric SLE patients. *J. Autoimmun.* **81**, 74–89 (2017).
183. Grandi, F. C. *et al.* Single-cell mass cytometry reveals cross-talk between inflammation-dampening and inflammation-amplifying cells in osteoarthritic cartilage. *Sci. Adv.* **6**, eaay5352 (2020).
184. Morbach, H., Eichhorn, E. M., Liese, J. G. & Girschick, H. J. Reference values for B cell subpopulations from infancy to adulthood. *Clin. Exp. Immunol.* **162**, 271–279 (2010).
185. Lens, S. M. A. *et al.* CD27–CD70 Interaction: Unravelling its Implication in Normal and Neoplastic B-Cell Growth. *Leuk. Lymphoma* **18**, 51–59 (1995).
186. Starzer, A. M. & Berghoff, A. S. New emerging targets in cancer immunotherapy: CD27 (TNFRSF7). *ESMO Open* **4**, (2020).
187. Granucci, F. & Zanoni, I. Role of CD14 in host protection against infections and in metabolism regulation. *Front. Cell. Infect. Microbiol.* **3**, (2013).
188. Yoon, B. R. *et al.* Functional Phenotype of Synovial Monocytes Modulating Inflammatory T-Cell Responses in Rheumatoid Arthritis (RA). *PLoS ONE* **9**, e109775 (2014).
189. Smiljanovic, B. *et al.* Monocyte alterations in rheumatoid arthritis are dominated by preterm release from bone marrow and prominent triggering in the joint. *Ann. Rheum. Dis.* **77**, 300–308 (2018).
190. Tsukamoto, M. *et al.* CD14<sup>bright</sup>CD16<sup>+</sup> intermediate monocytes are induced by interleukin-10 and positively correlate with disease activity in rheumatoid arthritis. *Arthritis Res. Ther.* **19**, 28 (2017).
191. Thomas Graham D. *et al.* Human Blood Monocyte Subsets. *Arterioscler. Thromb. Vasc. Biol.* **37**, 1548–1558 (2017).
192. Xiong, N. & Raulet, D. H. Development and selection of  $\gamma\delta$  T cells. *Immunol. Rev.* **215**, 15–31 (2007).

193. Gustafsson, K., Herrmann, T. & Dieli, F. Editorial: Understanding Gamma Delta T Cell Multifunctionality - Towards Immunotherapeutic Applications. *Front. Immunol.* **11**, (2020).
194. Collin, M., McGovern, N. & Haniffa, M. Human dendritic cell subsets. *Immunology* **140**, 22–30 (2013).
195. Shevach, E. M. CD4 + CD25 + suppressor T cells: more questions than answers. *Nat. Rev. Immunol.* **2**, 389–400 (2002).
196. Wing, J. B., Tanaka, A. & Sakaguchi, S. Human FOXP3+ Regulatory T Cell Heterogeneity and Function in Autoimmunity and Cancer. *Immunity* **50**, 302–316 (2019).
197. Kordasti, S. *et al.* Deep phenotyping of Tregs identifies an immune signature for idiopathic aplastic anemia and predicts response to treatment. *Blood* **128**, 1193–1205 (2016).
198. Yeap, W. H. *et al.* CD16 is indispensable for antibody-dependent cellular cytotoxicity by human monocytes. *Sci. Rep.* **6**, srep34310 (2016).
199. Saraiva, D. P., Jacinto, A., Borralho, P., Braga, S. & Cabral, M. G. HLA-DR in Cytotoxic T Lymphocytes Predicts Breast Cancer Patients' Response to Neoadjuvant Chemotherapy. *Front. Immunol.* **9**, (2018).
200. Distribution of antigen presenting cells in peripheral blood during chronic HIV infection. *figshare* /articles/figure/\_Distribution\_of\_antigen\_presenting\_cells\_in\_peripheral\_blood\_during\_chronic\_HIV\_infection\_/785642/1 (2015) doi:10.1371/journal.pone.0072789.g001.
201. Wajant, H. & Siegmund, D. TNFR1 and TNFR2 in the Control of the Life and Death Balance of Macrophages. *Front. Cell Dev. Biol.* **7**, (2019).
202. Knop, J. *et al.* TNFR2 induced priming of the inflammasome leads to a RIPK1-dependent cell death in the absence of XIAP. *Cell Death Dis.* **10**, 1–14 (2019).
203. Law, R. H. P. *et al.* The structural basis for membrane binding and pore formation by lymphocyte perforin. *Nature* **468**, 447–451 (2010).
204. Trapani, J. A. & Smyth, M. J. Functional significance of the perforin/granzyme cell death pathway. *Nat. Rev. Immunol.* **2**, 735–747 (2002).

205. Stamper, C. C. *et al.* Crystal structure of the B7-1/CTLA-4 complex that inhibits human immune responses. *Nature* **410**, 608–611 (2001).
206. Rowshanravan, B., Halliday, N. & Sansom, D. M. CTLA-4: a moving target in immunotherapy. *Blood* **131**, 58–67 (2018).
207. Robert, C. A decade of immune-checkpoint inhibitors in cancer therapy. *Nat. Commun.* **11**, 3801 (2020).
208. Bonelli, M. *et al.* Abatacept (CTLA-4lg) treatment reduces T cell apoptosis and regulatory T cell suppression in patients with rheumatoid arthritis. *Rheumatology* **55**, 710–720 (2016).
209. Treiner, E. *et al.* Selection of evolutionarily conserved mucosal-associated invariant T cells by MR. **422**, 8 (2003).
210. Gold, M. C. *et al.* Human Mucosal Associated Invariant T Cells Detect Bacterially Infected Cells. *PLOS Biol.* **8**, e1000407 (2010).
211. Dusseaux, M. *et al.* Human MAIT cells are xenobiotic-resistant, tissue-targeted, CD161hi IL-17-secreting T cells. *Blood* **117**, 1250–1259 (2011).
212. Sansom, D. M., Manzotti, C. N. & Zheng, Y. What’s the difference between CD80 and CD86? *Trends Immunol.* **24**, 313–318 (2003).
213. Alimonti, J. B., Ball, T. B. & Fowke, K. R. Mechanisms of CD4+ T lymphocyte cell death in human immunodeficiency virus infection and AIDS. *J. Gen. Virol.* **84**, 1649–1661 (2003).
214. MacGlashan, D. Expression of CD203c and CD63 in Human Basophils: Relationship to Differential Regulation of Piecemeal and Anaphylactic Degranulation Processes. *Clin. Exp. Allergy J. Br. Soc. Allergy Clin. Immunol.* **40**, 1365–1377 (2010).
215. Angelo, L. S. *et al.* Practical NK cell phenotyping and variability in healthy adults. *Immunol. Res.* **62**, 341–356 (2015).
216. Fergusson, J. R., Fleming, V. M. & Klenerman, P. CD161-Expressing Human T Cells. *Front. Immunol.* **2**, (2011).
217. Patel, S. P. & Kurzrock, R. PD-L1 Expression as a Predictive Biomarker in Cancer Immunotherapy. *Mol. Cancer Ther.* **14**, 847–856 (2015).

218. Loke, P. & Allison, J. P. PD-L1 and PD-L2 are differentially regulated by Th1 and Th2 cells. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 5336–5341 (2003).
219. Han, Y., Liu, D. & Li, L. PD-1/PD-L1 pathway: current researches in cancer. *Am. J. Cancer Res.* **10**, 727–742 (2020).
220. Karnell, J. L., Rieder, S. A., Ettinger, R. & Kolbeck, R. Targeting the CD40-CD40L pathway in autoimmune diseases: Humoral immunity and beyond. *Adv. Drug Deliv. Rev.* **141**, 92–103 (2019).
221. Stout, R. D. & Suttles, J. The many roles of CD40 in cell-mediated inflammatory responses. *Immunol. Today* **17**, 487–492 (1996).
222. Bots, M. & Medema, J. P. Granzymes at a glance. *J. Cell Sci.* **119**, 5011–5014 (2006).
223. Hiebert, P. R. & Granville, D. J. Granzyme B in injury, inflammation, and repair. **18**, 10 (2012).
224. Darrah, E. & Rosen, A. Granzyme B cleavage of autoantigens in autoimmunity. *Cell Death Differ.* **17**, 624–632 (2010).
225. Sharpe, A. H. & Pauken, K. E. The diverse functions of the PD1 inhibitory pathway. *Nat. Rev. Immunol.* **18**, 153–167 (2018).
226. Cai, J., Wang, D., Zhang, G. & Guo, X. The Role Of PD-1/PD-L1 Axis In Treg Development And Function: Implications For Cancer Immunotherapy. *OncoTargets Ther.* **12**, 8437–8445 (2019).
227. Pardoll, D. M. The blockade of immune checkpoints in cancer immunotherapy. *Nat. Rev. Cancer* **12**, 252–264 (2012).
228. Pauken, K. E. & Wherry, E. J. Overcoming T cell exhaustion in infection and cancer. *Trends Immunol.* **36**, 265–276 (2015).
229. Vitale, M. *et al.* NKp44, a Novel Triggering Surface Molecule Specifically Expressed by Activated Natural Killer Cells, Is Involved in Non-Major Histocompatibility Complex-restricted Tumor Cell Lysis. *J. Exp. Med.* **187**, 2065–2072 (1998).
230. Baychelier, F. *et al.* Identification of a cellular ligand for the natural cytotoxicity receptor NKp44. *Blood* **122**, 2935–2942 (2013).
231. Hofmeister, R. *et al.* Interleukin-7: physiological roles and mechanisms of action. *Cytokine Growth Factor Rev.* **10**, 41–60 (1999).

232. Barrow, A. D., Martin, C. J. & Colonna, M. The Natural Cytotoxicity Receptors in Health and Disease. *Front. Immunol.* **10**, (2019).
233. Malavasi, F. *et al.* Evolution and Function of the ADP Ribosyl Cyclase/CD38 Gene Family in Physiology and Pathology. *Physiol. Rev.* **88**, 841–886 (2008).
234. Hartman, W. R. *et al.* CD38 expression, function, and gene resequencing in a human lymphoblastoid cell line-based model system. *Leuk. Lymphoma* **51**, 1315–1325 (2010).
235. Todd, R. F. The continuing saga of complement receptor type 3 (CR3). *J. Clin. Invest.* **98**, 1–2 (1996).
236. Kawai, K. *et al.* CD11b-mediated migratory property of peripheral blood B cells. *J. Allergy Clin. Immunol.* **116**, 192–197 (2005).
237. Mair, F. Gate to the Future: Computational Analysis of Immunophenotyping Data. *Cytometry A* **95**, 147–149 (2019).
238. Kimball, A. K. *et al.* A Beginner's Guide to Analyzing and Visualizing Mass Cytometry Data. *J. Immunol.* **200**, 3–22 (2018).
239. Chen, T. J. & Kotecha, N. Cytobank: providing an analytics platform for community cytometry data analysis and collaboration. *Curr. Top. Microbiol. Immunol.* **377**, 127–157 (2014).
240. Schwartz, G. W. *et al.* TooManyCells identifies and visualizes relationships of single-cell clades. *Nat. Methods* **17**, 405–413 (2020).
241. McInnes, L., Healy, J. & Melville, J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. *ArXiv180203426 Cs Stat* (2020).
242. Becht, E. *et al.* Evaluation of UMAP as an alternative to t-SNE for single-cell data. <http://biorxiv.org/lookup/doi/10.1101/298430> (2018) doi:10.1101/298430.
243. Tomic, A. *et al.* SIMON, an Automated Machine Learning System, Reveals Immune Signatures of Influenza Vaccine Responses. *J. Immunol.* (2019) doi:10.4049/jimmunol.1900033.
244. Willis, L. M. *et al.* Tellurium-based mass cytometry barcode for live and fixed cells. *Cytometry A* **93**, 685–694 (2018).
245. Zunder, E. R. *et al.* Palladium-based Mass-Tag Cell Barcoding with a Doublet-Filtering Scheme and Single Cell Deconvolution Algorithm. *Nat. Protoc.* **10**, 316–333 (2015).

246. Bagwell, C. B. *et al.* Automated Data Cleanup for Mass Cytometry. *Cytometry A* **97**, 184–198 (2020).
247. Bruce Bagwell, C. High-Dimensional Modeling for Cytometry: Building Rock Solid Models Using GemStone™ and Verity Cen-se™ High-Definition t-SNE Mapping. *Methods Mol. Biol. Clifton NJ* **1678**, 11–36 (2018).
248. Bagwell, C. B. & Adams, E. G. Fluorescence spectral overlap compensation for any number of flow cytometry parameters. *Ann. N. Y. Acad. Sci.* **677**, 167–184 (1993).
249. Nettey, L., Giles, A. J. & Chattopadhyay, P. K. OMIP-050: A 28-color/30-parameter Fluorescence Flow Cytometry Panel to Enumerate and Characterize Cells Expressing a Wide Array of Immune Checkpoint Molecules. *Cytometry A* **93**, 1094–1096 (2018).
250. Takahashi, C. *et al.* Mass cytometry panel optimization through the designed distribution of signal interference. *Cytometry A* **91**, 39–47 (2017).
251. Chevrier, S. *et al.* Compensation of Signal Spillover in Suspension and Imaging Mass Cytometry. *Cell Syst.* **6**, 612–620.e5 (2018).
252. Chen, H. *et al.* Cytokit: A Bioconductor Package for an Integrated Mass Cytometry Data Analysis Pipeline. *PLOS Comput. Biol.* **12**, e1005112 (2016).
253. Chen, T. J. & Kotecha, N. Cytobank: providing an analytics platform for community cytometry data analysis and collaboration. *Curr. Top. Microbiol. Immunol.* **377**, 127–157 (2014).
254. Levine, J. H. *et al.* Data-Driven Phenotypic Dissection of AML Reveals Progenitor-like Cells that Correlate with Prognosis. *Cell* **162**, 184–197 (2015).
255. DiGiuseppe, J. A., Cardinali, J. L., Rezuke, W. N. & Pe'er, D. PhenoGraph and viSNE Facilitate the Identification of Abnormal T-Cell Populations in Routine Clinical Flow Cytometric Data. *Cytometry B Clin. Cytom.* **94**, 588–601 (2018).
256. Van Gassen, S. *et al.* FlowSOM: Using self-organizing maps for visualization and interpretation of cytometry data. *Cytom. Part J. Int. Soc. Anal. Cytol.* **87**, 636–645 (2015).
257. Cossarizza, A. *et al.* Guidelines for the use of flow cytometry and cell sorting in immunological studies. *Eur. J. Immunol.* **47**, 1584–1797 (2017).

258. Gassen, S. V., Gaudilliere, B., Angst, M. S., Saeys, Y. & Aghaeepour, N. CytoNorm: A Normalization Algorithm for Cytometry Data. *Cytometry A* **97**, 268–278 (2020).
259. Weber, L. M. & Robinson, M. D. Comparison of clustering methods for high-dimensional single-cell flow and mass cytometry data. *Cytom. Part J. Int. Soc. Anal. Cytol.* **89**, 1084–1096 (2016).
260. Nowicka et al. - 2019 - CyTOF workflow differential discovery in high-thr.pdf.
261. Fonseka, C. Y. *et al.* Mixed Effects Association of Single Cells Identifies an Expanded Th1-Skewed Cytotoxic Effector CD4+ T Cell Subset in Rheumatoid Arthritis. *bioRxiv* 172403 (2018) doi:10.1101/172403.
262. Li, Y. H. *et al.* Scalable multi-sample single-cell data analysis by Partition-Assisted Clustering and Multiple Alignments of Networks. *PLOS Comput. Biol.* **13**, e1005875 (2017).
263. Orlova, D. Y. *et al.* QFMatch: multidimensional flow and mass cytometry samples alignment. *Sci. Rep.* **8**, 3291 (2018).
264. Rybakowska, P., Alarcón-Riquelme, M. E. & Marañón, C. Key steps and methods in the experimental design and data analysis of highly multi-parametric flow and mass cytometry. *Comput. Struct. Biotechnol. J.* **18**, 874–886 (2020).
265. Bader, L. *et al.* Candidate Markers for Stratification and Classification in Rheumatoid Arthritis. *Front. Immunol.* **10**, (2019).
266. Nair, N. *et al.* Mass cytometry as a platform for the discovery of cellular biomarkers to guide effective rheumatic disease therapy. *Arthritis Res. Ther.* **17**, (2015).
267. Udalova, I. A., Mantovani, A. & Feldmann, M. Macrophage heterogeneity in the context of rheumatoid arthritis. *Nat. Rev. Rheumatol.* **12**, 472–485 (2016).
268. Ermann, J., Rao, D. A., Teslovich, N. C., Brenner, M. B. & Raychaudhuri, S. Immune cell profiling to guide therapeutic decisions in rheumatic diseases. *Nat. Rev. Rheumatol.* **11**, 541–551 (2015).
269. Dolgin, E. Massive NIH–industry project opens portals to target validation. *Nat. Rev. Drug Discov.* **18**, 240–242 (2019).

270. Bocharnikov, A. V. *et al.* PD-1<sup>hi</sup>CXCR5<sup>-</sup> T peripheral helper cells promote B cell responses in lupus via MAF and IL-21. *JCI Insight* **4**, (2019).
271. T, C., E, B. & El, C. Expansion of a Novel Subset of PD1+CXCR5-CD4+ T Peripheral Helper Cells in IgG4-Related Disease. *Clin. Transl. Gastroenterol.* **11**, e00111 (2020).
272. Kamekura, R. *et al.* Circulating PD-1+CXCR5-CD4+ T cells underlying the immunological mechanisms of IgG4-related disease. *Rheumatol. Adv. Pract.* **2**, (2018).
273. Liu, W., Zhou, X., Wang, A., Ma, J. & Bai, Y. Increased peripheral helper T cells type 17 subset correlates with the severity of psoriasis vulgaris. *Immunol. Lett.* **229**, 48–54 (2021).
274. Christophersen, A. *et al.* Distinct phenotype of CD4+ T cells driving celiac disease identified in multiple autoimmune conditions. *Nat. Med.* **25**, 734–737 (2019).
275. Smiljanovic, B. *et al.* Synovial tissue transcriptomes of long-standing rheumatoid arthritis are dominated by activated macrophages that reflect microbial stimulation. *Sci. Rep.* **10**, 7907 (2020).
276. Shanaj, Sara. & Donlin, L. T. Synovial Tissue: Cellular and Molecular Phenotyping. *Curr. Rheumatol. Rep.* **21**, 52 (2019).
277. Dakin, S. G. Synovial signatures signpost arthritis. *Sci. Transl. Med.* **11**, (2019).
278. Small, A. & Wechalekar, M. D. Synovial biopsies in inflammatory arthritis: precision medicine in rheumatoid arthritis. *Expert Rev. Mol. Diagn.* **20**, 315–325 (2020).
279. Carr, H. L., Turner, J. D., Major, T., Scheel-Toellner, D. & Filer, A. New Developments in Transcriptomic Analysis of Synovial Tissue. *Front. Med.* **7**, (2020).
280. Donlin, L. T. *et al.* Insights into rheumatic diseases from next-generation sequencing. *Nat. Rev. Rheumatol.* **15**, 327–339 (2019).
281. Rincon-Arevalo, H. *et al.* Deep Phenotyping of CD11c+ B Cells in Systemic Autoimmunity and Controls. *Front. Immunol.* **12**, (2021).
282. Synovial CD8 T Cells in Rheumatoid Arthritis Exhibit High Antigen-independent Cytokine Production and Low Cytotoxic Potential. *ACR Meeting Abstracts*  
<https://acrabstracts.org/abstract/synovial-cd8-t-cells-in-rheumatoid-arthritis-exhibit-high-antigen-independent-cytokine-production-and-low-cytotoxic-potential/>.

283. Raptopoulou, A. P. *et al.* The programmed death 1/programmed death ligand 1 inhibitory pathway is up-regulated in rheumatoid synovium and regulates peripheral T cell responses in human and murine arthritis. *Arthritis Rheum.* **62**, 1870–1880 (2010).
284. Mizoguchi, F. *et al.* Functionally distinct disease-associated fibroblast subsets in rheumatoid arthritis. *Nat. Commun.* **9**, 789 (2018).
285. Buckley, C. D. Why should rheumatologists care about fibroblasts?: Answering questions about tissue tropism and disease persistence. *Rheumatology* **56**, 863–864 (2017).
286. Croft, A. P. *et al.* Distinct fibroblast subsets drive inflammation and damage in arthritis. *Nature* **570**, 246–251 (2019).
287. Stephenson, W. *et al.* Single-cell RNA-seq of rheumatoid arthritis synovial tissue using low-cost microfluidic instrumentation. *Nat. Commun.* **9**, 791 (2018).
288. Orange, D. E. *et al.* RNA Identification of PRIME Cells Predicting Rheumatoid Arthritis Flares. *N. Engl. J. Med.* **383**, 218–228 (2020).
289. Pratt, A. G. *et al.* Targeting synovial fibroblast proliferation in rheumatoid arthritis (TRAFIC): an open-label, dose-finding, phase 1b trial. *Lancet Rheumatol.* **0**, (2021).
290. Buch, M. H. Defining refractory rheumatoid arthritis. *Ann. Rheum. Dis.* **77**, 966–969 (2018).
291. Aghaeepour, N. *et al.* Critical assessment of automated flow cytometry data analysis techniques. *Nat. Methods* **10**, 228–238 (2013).
292. Rahim, A. *et al.* High Throughput Automated Analysis of Big Flow Cytometry Data. *Methods San Diego Calif* **134–135**, 164–176 (2018).
293. Conrad, V. K. *et al.* Implementation and Validation of an Automated Flow Cytometry Analysis Pipeline for Human Immune Profiling. *Cytometry A* **95**, 183–191 (2019).
294. Stassen, S. V. *et al.* PARC: ultrafast and accurate clustering of phenotypic data of millions of single cells. **13**.
295. Li, Z. *et al.* The identification and functional analysis of CD8+PD-1+CD161+ T cells in hepatocellular carcinoma. *NPJ Precis. Oncol.* **4**, (2020).
296. Cho, Y.-N. *et al.* Mucosal-Associated Invariant T Cell Deficiency in Systemic Lupus Erythematosus. *J. Immunol.* **193**, 3891–3901 (2014).

297. Sugimoto, C. *et al.* Mucosal-Associated Invariant T Cell Is a Potential Marker to Distinguish Fibromyalgia Syndrome from Arthritis. *PLOS ONE* **10**, e0121124 (2015).
298. Toussiro, É., Laheurte, C., Gaugler, B., Gabriel, D. & Saas, P. Increased IL-22- and IL-17A-Producing Mucosal-Associated Invariant T Cells in the Peripheral Blood of Patients With Ankylosing Spondylitis. *Front. Immunol.* **9**, (2018).
299. Toubal, A., Nel, I., Lotersztajn, S. & Lehuen, A. Mucosal-associated invariant T cells and disease. *Nat. Rev. Immunol.* **19**, 643–657 (2019).
300. Koppejan, H. *et al.* Altered composition and phenotype of mucosal-associated invariant T cells in early untreated rheumatoid arthritis. *Arthritis Res. Ther.* **21**, (2019).
301. Gracey, E. *et al.* IL-7 primes IL-17 in mucosal-associated invariant T (MAIT) cells, which contribute to the Th17-axis in ankylosing spondylitis. *Ann. Rheum. Dis.* **75**, 2124–2132 (2016).
302. Dias, J. *et al.* The CD4–CD8– MAIT cell subpopulation is a functionally distinct subset developmentally related to the main CD8+ MAIT cell pool. *Proc. Natl. Acad. Sci.* **115**, E11513–E11522 (2018).
303. Chiba, A., Murayama, G. & Miyake, S. Mucosal-Associated Invariant T Cells in Autoimmune Diseases. *Front. Immunol.* **9**, (2018).
304. Keller, A. N., Corbett, A. J., Wubben, J. M., McCluskey, J. & Rossjohn, J. MAIT cells and MR1-antigen recognition. *Curr. Opin. Immunol.* **46**, 66–74 (2017).
305. Liu, Y.-J. IPC: professional type 1 interferon-producing cells and plasmacytoid dendritic cell precursors. *Annu. Rev. Immunol.* **23**, 275–306 (2005).
306. Cavanagh, L. L. *et al.* Rheumatoid arthritis synovium contains plasmacytoid dendritic cells. *Arthritis Res. Ther.* **7**, R230–R240 (2005).
307. Mingueneau, M. *et al.* Cytometry by time-of-flight immunophenotyping identifies a blood Sjögren’s signature correlating with disease activity and glandular inflammation. *J. Allergy Clin. Immunol.* **137**, 1809-1821.e12 (2016).
308. Van Krinks, C. H., Matyszak, M. K. & Gaston, J. S. H. Characterization of plasmacytoid dendritic cells in inflammatory arthritis synovial fluid. *Rheumatol. Oxf. Engl.* **43**, 453–460 (2004).

309. Lebre, M. C. *et al.* Rheumatoid Arthritis Synovium Contains Two Subsets of CD83–DC–LAMP–Dendritic Cells with Distinct Cytokine Profiles. *Am. J. Pathol.* **172**, 940–950 (2008).
310. Zhang, H. *et al.* A distinct subset of plasmacytoid dendritic cells induces activation and differentiation of B and T lymphocytes. *Proc. Natl. Acad. Sci.* **114**, 1988–1993 (2017).
311. Cooles, F. A. H. *et al.* Phenotypic and Transcriptomic Analysis of Peripheral Blood Plasmacytoid and Conventional Dendritic Cells in Early Drug Naïve Rheumatoid Arthritis. *Front. Immunol.* **9**, 755 (2018).
312. Jongbloed, S. L. *et al.* Enumeration and phenotypical analysis of distinct dendritic cell subsets in psoriatic arthritis and rheumatoid arthritis. *Arthritis Res. Ther.* **8**, R15 (2006).
313. Yanagihara, Y., Shiozawa, K., Takai, M., Kyogoku, M. & Shiozawa, S. Natural killer (NK) T cells are significantly decreased in the peripheral blood of patients with rheumatoid arthritis (RA). *Clin. Exp. Immunol.* **118**, 131–136 (1999).
314. Chalan, P., Bijzet, J., Kroesen, B.-J., Boots, A. M. H. & Brouwer, E. Altered Natural Killer Cell Subsets in Seropositive Arthralgia and Early Rheumatoid Arthritis Are Associated with Autoantibody Status. *J. Rheumatol.* **43**, 1008–1016 (2016).
315. Godfrey, D. I., Hammond, K. J. L., Poulton, L. D., Smyth, M. J. & Baxter, A. G. NKT cells: facts, functions and fallacies. *Immunol. Today* **21**, 573–583 (2000).
316. Elemam, N. M., Hachim, M. Y., Hannawi, S. & Maghazachi, A. A. Differentially Expressed Genes of Natural Killer Cells Can Distinguish Rheumatoid Arthritis Patients from Healthy Controls. *Genes* **11**, 492 (2020).
317. Yamin, R. *et al.* High percentages and activity of synovial fluid NK cells present in patients with advanced stage active Rheumatoid Arthritis. *Sci. Rep.* **9**, 1351 (2019).
318. Ahern, D. J. & Brennan, F. M. The role of Natural Killer cells in the pathogenesis of rheumatoid arthritis: Major contributors or essential homeostatic modulators? *Immunol. Lett.* **136**, 115–121 (2011).
319. Lo, C. K. C. *et al.* Natural killer cell degeneration exacerbates experimental arthritis in mice via enhanced interleukin-17 production. *Arthritis Rheum.* **58**, 2700–2711 (2008).

320. Söderström, K. *et al.* Natural killer cells trigger osteoclastogenesis and bone destruction in arthritis. *Proc. Natl. Acad. Sci.* **107**, 13028–13033 (2010).
321. Piccioli, D., Sbrana, S., Melandri, E. & Valiante, N. M. Contact-dependent stimulation and inhibition of dendritic cells by natural killer cells. *J. Exp. Med.* **195**, 335–341 (2002).
322. Aramaki, T. *et al.* A significantly impaired natural killer cell activity due to a low activity on a per-cell basis in rheumatoid arthritis. *Mod. Rheumatol.* **19**, 245–252 (2009).
323. Thanapati, S. *et al.* Impaired NK cell functionality and increased TNF- $\alpha$  production as biomarkers of chronic chikungunya arthritis and rheumatoid arthritis. *Hum. Immunol.* **78**, 370–374 (2017).
324. Scapini, P. & Cassatella, M. A. Social networking of human neutrophils within the immune system. *Blood* **124**, 710–719 (2014).
325. Carmona-Rivera, C. & Kaplan, M. J. Low density granulocytes: a distinct class of neutrophils in systemic autoimmunity. *Semin. Immunopathol.* **35**, 455–463 (2013).
326. Pillay, J. *et al.* A subset of neutrophils in human systemic inflammation inhibits T cell responses through Mac-1. *J. Clin. Invest.* **122**, 327–336 (2012).
327. Schuster, S., Hurrell, B. & Tacchini-Cottier, F. Crosstalk between neutrophils and dendritic cells: a context-dependent process. *J. Leukoc. Biol.* **94**, 671–675 (2013).
328. Parackova, Z. *et al.* Neutrophil Extracellular Trap Induced Dendritic Cell Activation Leads to Th1 Polarization in Type 1 Diabetes. *Front. Immunol.* **11**, (2020).
329. Dieker, J. *et al.* Circulating Apoptotic Microparticles in Systemic Lupus Erythematosus Patients Drive the Activation of Dendritic Cell Subsets and Prime Neutrophils for NETosis. *Arthritis Rheumatol. Hoboken NJ* **68**, 462–472 (2016).
330. Costantini, C. & Cassatella, M. A. The defensive alliance between neutrophils and NK cells as a novel arm of innate immunity. *J. Leukoc. Biol.* **89**, 221–233 (2011).
331. Jaeger, B. N. *et al.* Neutrophil depletion impairs natural killer cell maturation, function, and homeostasis. *J. Exp. Med.* **209**, 565–580 (2012).
332. Schneider, M. *et al.* Neutrophils Suppress Mucosal-Associated Invariant T Cells. *bioRxiv* 745414 (2019) doi:10.1101/745414.

333. Bano, A. *et al.* CD28null CD4 T-cell expansions in autoimmune disease suggest a link with cytomegalovirus infection. *F1000Research* **8**, 327 (2019).
334. Thompson, C., Davies, R., Williams, A., Jones, G. & Choy, E. H. S. CD28- Cells Are Increased in Early Rheumatoid Arthritis and Are Linked With Cytomegalovirus Status. *Front. Med.* **7**, 129 (2020).
335. Lee and Lee - 2016 - Unusual CD4 + CD28 – T Cells.pdf.
336. Benfaremo, D. & Gabrielli, A. Is There a Future for Anti-CD38 Antibody Therapy in Systemic Autoimmune Diseases? *Cells* **9**, 77 (2019).
337. Cole, S. *et al.* Integrative analysis reveals CD38 as a therapeutic target for plasma cell-rich pre-disease and established rheumatoid arthritis and systemic lupus erythematosus. *Arthritis Res. Ther.* **20**, 85 (2018).
338. Morandi, F. *et al.* CD38: A Target for Immunotherapeutic Approaches in Multiple Myeloma. *Front. Immunol.* **9**, (2018).
339. Burns, M. *et al.* Dysregulated CD38 Expression on Peripheral Blood Immune Cell Subsets in SLE. *Int. J. Mol. Sci.* **22**, 2424 (2021).
340. Rizzello, V. *et al.* Modulation of CD4+CD28null T Lymphocytes by Tumor Necrosis Factor- $\alpha$  Blockade in Patients With Unstable Angina. *Circulation* **113**, 2272–2277 (2006).
341. Maly, K. & Schirmer, M. The story of CD4+ CD28- T cells revisited: solved or still ongoing? *J. Immunol. Res.* **2015**, 348746 (2015).
342. Filer, A. *et al.* Identification of a transitional fibroblast function in very early rheumatoid arthritis. *Ann. Rheum. Dis.* **76**, 2105–2112 (2017).
343. Humby, F. *et al.* Use of Ultrasound-Guided Small Joint Biopsy to Evaluate the Histopathologic Response to Rheumatoid Arthritis Therapy: Recommendations for Application to Clinical Trials: EVALUATION OF RA THERAPY BY US-GUIDED BIOPSY OF SMALL JOINTS. *Arthritis Rheumatol.* **67**, 2601–2610 (2015).
344. Kelly, S. *et al.* 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**, 611–617 (2015).

345. Geborek, P., Saxne, T., Heinegård, D. & Wollheim, F. A. Measurement of synovial fluid volume using albumin dilution upon intraarticular saline injection. *J. Rheumatol.* **15**, 91–94 (1988).
346. Hui, A. Y., McCarty, W. J., Masuda, K., Firestein, G. S. & Sah, R. L. A systems biology approach to synovial joint lubrication in health, injury, and disease. *WIREs Syst. Biol. Med.* **4**, 15–37 (2012).
347. Tamer, T. M. Hyaluronan and synovial joint: function, distribution and healing. *Interdiscip. Toxicol.* **6**, 111–125 (2013).
348. Dieppe, P. & Swan, A. Identification of crystals in synovial fluid. *Ann. Rheum. Dis.* **58**, 261–263 (1999).
349. Edwards, S. W. & Hallett, M. B. Seeing the wood for the trees: the forgotten role of neutrophils in rheumatoid arthritis. *Immunol. Today* **18**, 320–324 (1997).
350. Kolaczowska, E. & Kubes, P. Neutrophil recruitment and function in health and inflammation. *Nat. Rev. Immunol.* **13**, 159–175 (2013).
351. McCracken, J. M. & Allen, L.-A. H. Regulation of Human Neutrophil Apoptosis and Lifespan in Health and Disease. *J. Cell Death* **7**, 15–23 (2014).
352. Moulding, D. A., Quayle, J. A., Hart, C. A. & Edwards, S. W. Mcl-1 Expression in Human Neutrophils: Regulation by Cytokines and Correlation With Cell Survival. *Blood* **92**, 2495–2502 (1998).
353. Wright, H. L., Moots, R. J. & Edwards, S. W. The multifactorial role of neutrophils in rheumatoid arthritis. *Nat. Rev. Rheumatol.* **10**, 593–601 (2014).
354. Wright, H. L., Moots, R. J., Bucknall, R. C. & Edwards, S. W. Neutrophil function in inflammation and inflammatory diseases. *Rheumatology* **49**, 1618–1631 (2010).
355. Leite Pereira, A. *et al.* Characterization of Phenotypes and Functional Activities of Leukocytes From Rheumatoid Arthritis Patients by Mass Cytometry. *Front. Immunol.* **10**, (2019).
356. Wright, H. L., Lyon, M., Chapman, E. A., Moots, R. J. & Edwards, S. W. Rheumatoid Arthritis Synovial Fluid Neutrophils Drive Inflammation Through Production of Chemokines, Reactive Oxygen Species, and Neutrophil Extracellular Traps. *Front. Immunol.* **11**, (2021).

357. Raza, K. *et al.* Synovial fluid leukocyte apoptosis is inhibited in patients with very early rheumatoid arthritis. *Arthritis Res. Ther.* **8**, R120 (2006).
358. Penatti, A. *et al.* Differences in serum and synovial CD4+ T cells and cytokine profiles to stratify patients with inflammatory osteoarthritis and rheumatoid arthritis. *Arthritis Res. Ther.* **19**, 103 (2017).
359. Campbell, D. J., Kim, C. H. & Butcher, E. C. Separable effector T cell populations specialized for B cell help or tissue inflammation. *Nat. Immunol.* **2**, 876–881 (2001).
360. Hatachi, S. *et al.* CD4+ PD-1+ T Cells Accumulate as Unique Anergic Cells in Rheumatoid Arthritis Synovial Fluid. *J. Rheumatol.* **10**.
361. Wan, B. *et al.* Aberrant Regulation of Synovial T Cell Activation by Soluble Costimulatory Molecules in Rheumatoid Arthritis. *J. Immunol.* **177**, 8844–8850 (2006).
362. Johnston, R. J. *et al.* Bcl6 and Blimp-1 are reciprocal and antagonistic regulators of T follicular helper cell differentiation. *Science* **325**, 1006–1010 (2009).
363. Lawson, C. A. *et al.* Early rheumatoid arthritis is associated with a deficit in the CD4 + CD25 high regulatory T cell population in peripheral blood. *Rheumatology* **45**, 1210–1217 (2006).
364. Cao, D. *et al.* Isolation and functional characterization of regulatory CD25brightCD4+ T cells from the target organ of patients with rheumatoid arthritis. *Eur. J. Immunol.* **33**, 215–223 (2003).
365. Yang, S. *et al.* Induced, but not natural, regulatory T cells retain phenotype and function following exposure to inflamed synovial fibroblasts. *Sci. Adv.* **6**, eabb0606 (2020).
366. Amelsfort, J. M. R. van, Jacobs, K. M. G., Bijlsma, J. W. J., Lafeber, F. P. J. G. & Taams, L. S. CD4+CD25+ regulatory T cells in rheumatoid arthritis: Differences in the presence, phenotype, and function between peripheral blood and synovial fluid. *Arthritis Rheum.* **50**, 2775–2785 (2004).
367. Lawson, C. A. *et al.* Early rheumatoid arthritis is associated with a deficit in the CD4 + CD25 high regulatory T cell population in peripheral blood. *Rheumatology* **45**, 1210–1217 (2006).

368. Wang, T. *et al.* Regulatory T cells in rheumatoid arthritis showed increased plasticity toward Th17 but retained suppressive function in peripheral blood. *Ann. Rheum. Dis.* **74**, 1293–1301 (2015).
369. Beavis, P. A. *et al.* Resistance to regulatory T cell-mediated suppression in rheumatoid arthritis can be bypassed by ectopic foxp3 expression in pathogenic synovial T cells. *Proc. Natl. Acad. Sci.* **108**, 16717–16722 (2011).
370. Martin, M. D. & Badovinac, V. P. Defining Memory CD8 T Cell. *Front. Immunol.* **9**, (2018).
371. Jung, J. H. *et al.* Synovial fluid CD69+CD8+ T cells with tissue-resident phenotype mediate perforin-dependent citrullination in rheumatoid arthritis. *Clin. Transl. Immunol.* **9**, e1140 (2020).
372. Petrelli, A. *et al.* PD-1<sup>+</sup>CD8<sup>+</sup> T cells are clonally expanding effectors in human chronic inflammation. *J. Clin. Invest.* **128**, 4669–4681 (2018).
373. Cho, B.-A. *et al.* Characterization of Effector Memory CD8+ T Cells in the Synovial Fluid of Rheumatoid Arthritis. *J. Clin. Immunol.* **32**, 709–720 (2012).
374. Carvalheiro, H., Duarte, C., Silva-Cardoso, S., Silva, J. A. P. da & Souto-Carneiro, M. M. CD8+ T Cell Profiles in Patients With Rheumatoid Arthritis and Their Relationship to Disease Activity. *Arthritis Rheumatol.* **67**, 363–371 (2015).
375. Dalbeth, N. & Callan, M. F. C. A subset of natural killer cells is greatly expanded within inflamed joints. *Arthritis Rheum.* **46**, 1763–1772 (2002).
376. Meliconi, R., Pitzalis, C., Kingsley, G. H. & Panayi, G. S.  $\gamma\delta$  T cells and their subpopulations in blood and synovial fluid from rheumatoid arthritis and spondyloarthritis. *Clin. Immunol. Immunopathol.* **59**, 165–172 (1991).
377. Liu, M.-F. *et al.* Distribution of Double-Negative (CD4 – CD8 – , DN) T Subsets in Blood and Synovial Fluid from Patients with Rheumatoid Arthritis. *Clin. Rheumatol.* **18**, 227–231 (1999).
378. Cron, R. Q. *et al.* Phenotypic and functional analysis of murine CD3+,CD4-,CD8- TCR-gamma delta-expressing peripheral T cells. *J. Immunol. Baltim. Md 1950* **142**, 3754–3762 (1989).
379. Xue, X. *et al.* Pharmacologic modulation of ROR $\gamma$ t translates to efficacy in preclinical and translational models of psoriasis and inflammatory arthritis. *Sci. Rep.* **6**, 37977 (2016).

380. Ito, Y. *et al.* Gamma/delta T cells are the predominant source of interleukin-17 in affected joints in collagen-induced arthritis, but not in rheumatoid arthritis. *Arthritis Rheum.* **60**, 2294–2303 (2009).
381. Pöllinger, B. *et al.* Th17 Cells, Not IL-17+  $\gamma\delta$  T Cells, Drive Arthritic Bone Destruction in Mice and Humans. *J. Immunol.* **186**, 2602–2612 (2011).
382. Aungier, S. R. *et al.* Targeting early changes in the synovial microenvironment: a new class of immunomodulatory therapy? *Ann. Rheum. Dis.* **78**, 186–191 (2019).
383. Buckley, C. D. Why do leucocytes accumulate within chronically inflamed joints? *Rheumatology* **42**, 1433–1444 (2003).
384. Canavan, M. *et al.* Rheumatoid arthritis synovial microenvironment induces metabolic and functional adaptations in dendritic cells. *Clin. Exp. Immunol.* **202**, 226–238 (2020).
385. Guo, Y. *et al.* Immune checkpoint inhibitor PD-1 pathway is down-regulated in synovium at various stages of rheumatoid arthritis disease progression. *PLOS ONE* **13**, e0192704 (2018).
386. Risk of Immunotherapy Related Toxicity in Patients with Rheumatoid Arthritis. *ACR Meeting Abstracts* <https://acrabstracts.org/abstract/risk-of-immunotherapy-related-toxicity-in-patients-with-rheumatoid-arthritis/>.
387. Steven, N. M. & Fisher, B. A. Management of rheumatic complications of immune checkpoint inhibitor therapy – an oncological perspective. *Rheumatology* **58**, vii29–vii39 (2019).
388. Zhang, F. *et al.* Defining Inflammatory Cell States in Rheumatoid Arthritis Joint Synovial Tissues by Integrating Single-cell Transcriptomics and Mass Cytometry. *bioRxiv* 351130 (2018).
389. Pitzalis, C., Kingsley, G., Murphy, J. & Panayi, G. Abnormal distribution of the helper-inducer and suppressor-inducer T-lymphocyte subsets in the rheumatoid joint. *Clin. Immunol. Immunopathol.* **45**, 252–258 (1987).
390. Cush, J. J. & Lipsky, P. E. Phenotypic analysis of synovial tissue and peripheral blood lymphocytes isolated from patients with rheumatoid arthritis. *Arthritis Rheum.* **31**, 1230–1238 (1988).

391. Warrington, K. J., Takemura, S., Goronzy, J. J. & Weyand, C. M. CD4<sup>+</sup>,CD28<sup>-</sup> T cells in rheumatoid arthritis patients combine features of the innate and adaptive immune systems. *Arthritis Rheum.* **44**, 13–20 (2001).
392. Schmidt, D., Goronzy, J. J. & Weyand, C. M. CD4<sup>+</sup> CD7<sup>-</sup> CD28<sup>-</sup> T cells are expanded in rheumatoid arthritis and are characterized by autoreactivity. *J. Clin. Invest.* **97**, 2027–2037 (1996).
393. Namekawa, T., Wagner, U. G., Goronzy, J. J. & Weyand, C. M. Functional subsets of CD4 T cells in rheumatoid synovitis. *Arthritis Rheum.* **41**, 2108–2116 (1998).
394. Stastny, P. Association of the B-Cell Alloantigen DRw4 with Rheumatoid Arthritis. *N. Engl. J. Med.* **298**, 869–871 (1978).
395. Takeshita, M. *et al.* Multi-dimensional analysis identified rheumatoid arthritis-driving pathway in human T cell. *Ann. Rheum. Dis.* **78**, 1346–1356 (2019).
396. Prasad, K. V. S. *et al.* CD27, a member of the tumor necrosis factor receptor family, induces apoptosis and binds to Siva, a proapoptotic protein. *Proc. Natl. Acad. Sci.* **94**, 6346–6351 (1997).
397. Schiödt, Å., Lindstedt, M., Johansson-Lindbom, B., Roggen, E. & Borrebaeck, C. A. K. CD27<sup>-</sup> CD4<sup>+</sup> memory T cells define a differentiated memory population at both the functional and transcriptional levels. *Immunology* **113**, 363–370 (2004).
398. Kohem, C. L. *et al.* Enrichment of differentiated CD45RB<sup>dim</sup>,CD27<sup>-</sup> memory T cells in the peripheral blood, synovial fluid, and synovial tissue of patients with rheumatoid arthritis. *Arthritis Rheum.* **39**, 844–854 (1996).
399. Gattorno, M. *et al.* Levels of soluble CD27 in sera and synovial fluid and its expression on memory T cells in patients with juvenile idiopathic arthritides. *Clin. Exp. Rheumatol.* **20**, 863–866 (2002).
400. Tak, P. P. *et al.* Expression of the Activation Antigen CD27 in Rheumatoid Arthritis. *Clin. Immunol. Immunopathol.* **80**, 129–138 (1996).
401. Ruprecht, C. R. *et al.* Coexpression of CD25 and CD27 identifies FoxP3<sup>+</sup> regulatory T cells in inflamed synovia. *J. Exp. Med.* **201**, 1793–1803 (2005).

402. Davis, L. S., Cush, J. J., Schulze-Koops, H. & Lipsky, P. E. Rheumatoid synovial CD4+ T cells exhibit a reduced capacity to differentiate into IL-4-producing T-helper-2 effector cells. **3**, 11.
403. Thewissen, M. *et al.* CD4+CD28null T Cells in Autoimmune Disease: Pathogenic Features and Decreased Susceptibility to Immunoregulation. *J. Immunol. Baltim. Md 1950* **179**, 6514–23 (2007).
404. Kosmaczewska, A. *et al.* CD4+CD28null T cells are expanded in moderately active systemic lupus erythematosus and secrete pro-inflammatory interferon gamma, depending on the Disease Activity Index. *Lupus* **29**, 705–714 (2020).
405. Akbar, A. N. & Henson, S. M. Are senescence and exhaustion intertwined or unrelated processes that compromise immunity? *Nat. Rev. Immunol.* **11**, 289–295 (2011).
406. Effros, R. B. *et al.* Decline in CD28+ T cells in centenarians and in long-term T cell cultures: A possible cause for both in vivo and in vitro immunosenescence. *Exp. Gerontol.* **29**, 601–609 (1994).
407. Vallejo, A. N., Nestel, A. R., Schirmer, M., Weyand, C. M. & Goronzy, J. J. Aging-related Deficiency of CD28 Expression in CD4+ T Cells Is Associated with the Loss of Gene-specific Nuclear Factor Binding Activity. *J. Biol. Chem.* **273**, 8119–8129 (1998).
408. Akbar, A. N., Henson, S. M. & Lanna, A. Senescence of T Lymphocytes: Implications for Enhancing Human Immunity. *Trends Immunol.* **37**, 866–876 (2016).
409. Kurioka, A., Klenerman, P. & Willberg, C. B. Innate-like CD8+ T-cells and NK cells: converging functions and phenotypes. *Immunology* **154**, 547–556 (2018).
410. Bix, M. & Locksley, R. M. Natural T cells. Cells that co-express NKRP-1 and TCR. *J. Immunol.* **155**, 1020–1022 (1995).
411. Waggoner, S. N., Cornberg, M., Selin, L. K. & Welsh, R. M. Natural killer cells act as rheostats modulating antiviral T cells. *Nature* **481**, 394–398 (2012).
412. McKinney, E. F. *et al.* A CD8 + NK cell transcriptomic signature associated with clinical outcome in relapsing remitting multiple sclerosis. *Nat. Commun.* **12**, 635 (2021).

413. de Matos, C. T. *et al.* Activating and inhibitory receptors on synovial fluid natural killer cells of arthritis patients: role of CD94/NKG2A in control of cytokine secretion. *Immunology* **122**, 291–301 (2007).
414. Pallmer, K. *et al.* NK cells negatively regulate CD8 T cells via natural cytotoxicity receptor (NCR) 1 during LCMV infection. *PLOS Pathog.* **15**, e1007725 (2019).
415. Goronzy, J. J. & Weyand, C. M. T-cell co-stimulatory pathways in autoimmunity. *Arthritis Res. Ther.* **10**, S3 (2008).
416. O'Donnell, E. A., Ernst, D. N. & Hingorani, R. Multiparameter Flow Cytometry: Advances in High Resolution Analysis. *Immune Netw.* **13**, 43–54 (2013).
417. Holzner, G. *et al.* Ultra High-Throughput Multiparametric Imaging Flow Cytometry: Towards Diffraction-Limited Sub-Cellular Detection. *bioRxiv* 695361 (2019) doi:10.1101/695361.
418. See, P., Lum, J., Chen, J. & Ginhoux, F. A Single-Cell Sequencing Guide for Immunologists. *Front. Immunol.* **9**, (2018).
419. Lee, J., Hyeon, D. Y. & Hwang, D. Single-cell multiomics: technologies and data analysis methods. *Exp. Mol. Med.* **52**, 1428–1442 (2020).
420. Accelerating Medicines Partnership Rheumatoid Arthritis and Systemic Lupus Erythematosus (AMP RA/SLE) Consortium *et al.* Defining inflammatory cell states in rheumatoid arthritis joint synovial tissues by integrating single-cell transcriptomics and mass cytometry. *Nat. Immunol.* (2019) doi:10.1038/s41590-019-0378-1.
421. Giesen, C. *et al.* Highly multiplexed imaging of tumor tissues with subcellular resolution by mass cytometry. *Nat. Methods* **11**, 417–422 (2014).
422. Baharlou, H., Canete, N. P., Cunningham, A. L., Harman, A. N. & Patrick, E. Mass Cytometry Imaging for the Study of Human Diseases—Applications and Data Analysis Strategies. *Front. Immunol.* **10**, (2019).
423. Croft, A. P. *et al.* Distinct fibroblast subsets drive inflammation and damage in arthritis. *Nature* **570**, 246–251 (2019).
424. Kurowska-Stolarska, M. *et al.* MicroRNA-155 as a proinflammatory regulator in clinical and experimental arthritis. *Proc. Natl. Acad. Sci.* **108**, 11193–11198 (2011).

425. Buckley, C. D., Ospelt, C., Gay, S. & Midwood, K. S. Location, location, location: how the tissue microenvironment affects inflammation in RA. *Nat. Rev. Rheumatol.* **17**, 195–212 (2021).
426. Imaging Mass Cytometry Utilized by Global Consortium Focused on New Precision Medicine Approaches for Treatment of Rheumatoid Arthritis. *BioSpace*  
<https://www.biospace.com/article/imaging-mass-cytometry-utilized-by-global-consortium-focused-on-new-precision-medicine-approaches-for-treatment-of-rheumatoid-arthritis-/>.
427. Wright, H. L., Moots, R. J., Bucknall, R. C. & Edwards, S. W. Neutrophil function in inflammation and inflammatory diseases. *Rheumatology* **49**, 1618–1631 (2010).
428. Zhang, L., Yuan, Y., Xu, Q., Jiang, Z. & Chu, C.-Q. Contribution of neutrophils in the pathogenesis of rheumatoid arthritis. *J. Biomed. Res.* **34**, 86–93 (2020).
429. O’Neil, L. J. & Kaplan, M. J. Neutrophils in Rheumatoid Arthritis: Breaking Immune Tolerance and Fueling Disease. *Trends Mol. Med.* **25**, 215–227 (2019).
430. Hellebrekers, P., Vrisekoop, N. & Koenderman, L. Neutrophil phenotypes in health and disease. *Eur. J. Clin. Invest.* **48**, e12943 (2018).
431. Holers, V. M. *et al.* Rheumatoid arthritis and the mucosal origins hypothesis: protection turns to destruction. *Nat. Rev. Rheumatol.* **14**, 542 (2018).
432. Jochems, S. P. *et al.* Defining mucosal immunity using mass cytometry following experimental human pneumococcal challenge. *bioRxiv* 546929 (2019) doi:10.1101/546929.
433. Gupta, S. D. Coronavirus Pandemic: A Serious Threat to Humanity. *J. Health Manag.* **22**, 1–2 (2020).
434. Hu, B., Guo, H., Zhou, P. & Shi, Z.-L. Characteristics of SARS-CoV-2 and COVID-19. *Nat. Rev. Microbiol.* **19**, 141–154 (2021).
435. Tay, M. Z., Poh, C. M., Rénia, L., MacAry, P. A. & Ng, L. F. P. The trinity of COVID-19: immunity, inflammation and intervention. *Nat. Rev. Immunol.* **20**, 363–374 (2020).
436. Favalli, E. G. *et al.* COVID-19 infection and rheumatoid arthritis: Faraway, so close! *Autoimmun. Rev.* **19**, 102523 (2020).
437. Baimukhamedov, C., Barskova, T. & Matucci-Cerinic, M. Arthritis after SARS-CoV-2 infection. *Lancet Rheumatol.* **3**, e324–e325 (2021).

438. Abani, O. *et al.* Tocilizumab in patients admitted to hospital with COVID-19 (RECOVERY): a randomised, controlled, open-label, platform trial. *The Lancet* **397**, 1637–1645 (2021).
439. Mahase, E. Covid-19: Low dose steroid cuts death in ventilated patients by one third, trial finds. *BMJ* **369**, m2422 (2020).
440. Chevrier, S. *et al.* A distinct innate immune signature marks progression from mild to severe COVID-19. *Cell Rep. Med.* **2**, 100166 (2021).
441. Wang, C. *et al.* Imaging Mass Cytometric Analysis of Postmortem Tissues Reveals Dysregulated Immune Cell and Cytokine Responses in Multiple Organs of COVID-19 Patients. *Front. Microbiol.* **11**, (2020).
442. MacDonald, L. *et al.* COVID-19 and Rheumatoid Arthritis share myeloid pathogenic and resolving pathways. *bioRxiv* 2020.07.26.221572 (2020) doi:10.1101/2020.07.26.221572.
443. Alivernini, S. *et al.* Distinct synovial tissue macrophage subsets regulate inflammation and remission in rheumatoid arthritis. *Nat. Med.* **26**, 1295–1306 (2020).
444. Mantovani, A. & Netea, M. G. Trained Innate Immunity, Epigenetics, and Covid-19. *N. Engl. J. Med.* **383**, 1078–1080 (2020).
445. O’Neill, L. A. J. & Netea, M. G. BCG-induced trained immunity: can it offer protection against COVID-19? *Nat. Rev. Immunol.* **20**, 335–337 (2020).

## Appendix A

### **PATIENT INFORMATION SHEET & CONSENT FORM**

Investigator: Professor Paul Emery

Study Title: RADAR: Rheumatoid Arthritis Disease Research

Protocol Number: RR09/9134

Study Sponsor: University of Leeds

Subject No.: \_\_\_\_\_ Initials: \_\_\_\_\_ Date of Birth: \_\_\_\_\_

---

You are being invited to take part in a research study. Before you decide, it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully and discuss it with friends, relatives and your doctor if you wish. Ask us if there is anything that is not clear or if you would like more information. Take time to decide whether or not you wish to take part.

You can find independent information on participating in clinical trials on the following web site: <http://www.clinicaltrialsportal.co.uk/medical-volunteers-patients.html>. If you do not have access to the internet, or prefer a hard copy of the information, please ask the study doctor or nurse, who will be happy to provide you with one.

#### **1. What is the purpose of the study?**

Inflammatory arthritis is a common condition and we would like to find out more about the very early stages. We know that early treatment of inflammatory arthritis gives better results than if treatment is delayed. Unfortunately, it can be very difficult to catch patients in the early stage of the disease as often symptoms can be mild and can be atypical, such as initially just single joint involvement.

The University of Leeds aims to develop a comprehensive research program to investigate the clinical characteristics of inflammatory arthritis patients attending the Early Arthritis Clinic. This would be of great benefit in providing further insights into the disease in order to assist with future treatments.

The reason behind studying people at an early stage, when symptoms are new and may be associated with only small amounts of joint swelling, is that such individuals hold the key to understanding what makes some people develop a persistent arthritis and others go into remission. By identifying the first changes of inflammatory arthritis using the most sensitive techniques available, our knowledge will be significantly advanced. This ultimately will help improve early diagnosis and tailor treatment to the individual. We therefore want to collect data on all patients with possible or probable or definite early inflammatory arthritis over time, producing a database from which we can address these important research questions.

## **2. Why have I been chosen?**

You have been invited to participate in this study because you have been identified as having possible or probable or definite inflammatory arthritis. We want to assess people like you over time to learn more about inflammatory arthritis.

## **3. Do I have to take part?**

It is up to you to decide whether or not to take part in the study. If you decide to take part you will be given this information sheet to keep and be asked to sign a consent form. If you decide to take part you are still free to withdraw from the study at any time and without giving a reason. This will not affect the standard of care you receive.

## **4. What will happen to me if I take part?**

Your arthritis will be treated according to our usual treatment guidelines. This is NOT a study looking at new or experimental treatments and there will be little change in the care that you receive from your rheumatologist. However, your answers and findings from the routine questions asked and examinations that occur within your normal consultation with your rheumatologist will be recorded into a secure database. Results from any investigations you undergo will be recorded. Retrospective data from your clinical notes may also be recorded.

Only your initials and your unique study number and no other personal identifiers will be recorded on the database. Access to the database will be protected by password and restricted to only those involved in the study.

Patients will be asked to return for outpatient visits approximately every 3 to 6 months and then annually thereafter, unless there is a clinical need to see you more or less often. This is the same number of visits as would happen in normal clinical care.

Your doctor or nurse will carry out the following procedures which are over and above the standard clinical care you would receive if you were not taking part in the study:

- If you decide to take part, you will have the opportunity to discuss in detail what participation means and you will then need to sign the consent form as part of your initial visit for the study.
- You may also be asked to take part in one or more optional sub-studies. These include additional procedures which are over and above standard care. Each sub-study has an additional patient information sheet which outlines the procedures. Your Doctor or Nurse will discuss these with you.

At your study visits, we will capture clinical information that is part of routine care. As part of your normal hospital visit your Doctor and Nurse will:

- Ask questions about your medical history and medications you have been taking.
- Conduct routine physical examinations including blood pressure, pulse rate, height and weight.
- Take blood samples for routine arthritis tests, including fasting samples if applicable.
- Assessments will be completed regarding your current symptoms, general wellbeing and ability to carry out everyday tasks.
- X-rays of your hands and/or feet will be conducted at your initial visit and annually. An x-ray is a commonly used diagnostic procedure. The small dose of radiation you receive will be the same as any standard x-ray and should not be a significant health risk.
- Ultrasound (US) of your hands, feet and other joints may be conducted when clinically relevant to your care. Ultrasound is done frequently on patients in our early arthritis clinic and is standard practice in our clinic. An ultrasound scan involves placing special jelly on your joints (in this case over your hand and knee) and then a scanner, which is in the shape of a flat probe, is run over this will last approximately 30 minutes. This should not be painful or uncomfortable in any way.
- Annual questions about your cardiovascular health:
  - The amount of exercise you do normally
  - Your dietary intake (including how much fruit and vegetables you eat)
  - Any family history of premature cardiovascular disease
  - Chest pains (if you have them)

## **5. What do I have to do?**

There should be no reason to change your current way of life if you participate in this study. As with normal clinical appointments, if you decide to take part in the study you should:

- Tell the Rheumatology doctor about any illness that you currently have or have had in the past.

- Tell your Rheumatology doctor about any other medications you are taking, as well as those supplied by your doctor.
- Keep the appointments for your study visits.
- Tell your Rheumatology doctor about any changes in your health that occurs during the study

## **6. What if I do not wish to take part?**

All studies are always completely voluntary. If you do not wish to take part this will not affect the treatment or care you receive.

## **7. What are the possible disadvantages and risks of taking part?**

Your care will not be greatly affected by consenting to this study. You will not require any further procedures to those that you would already undergo routinely as part of your care within the rheumatology department.

## **8. What are the possible benefits of taking part?**

It cannot be guaranteed that you will gain personal benefit from this study: however, beneficial information may be acquired for patients who develop rheumatoid arthritis and may help us to treat these future patients better.

## **9. What if new information becomes available?**

Sometimes during the course of a research project, new information becomes available. If this happens, your Rheumatology doctor will tell you about it and discuss with you whether you want to continue in the study. If you decide to withdraw, your Rheumatology doctor will make arrangements for your care to continue. If you decide to continue in the study, you will be asked to sign an updated consent form after reading a new information sheet.

Also, on receiving new information, your Rheumatology doctor might consider it to be in your best interests to withdraw you from the study. He/she will explain the reasons and arrange for your care to continue.

## **10. What happens when the research study stops?**

Once the study is over, your research doctor will decide whether you need to be continued to be monitored in the Rheumatology Unit or whether you can be followed by your GP.

## **11. Other information**

All your written and computer records will be kept strictly confidential at all times. Data Protection Act regulations have been complied with to ensure confidentiality.

## **12. What will happen to the results of the research study?**

The results from the study will be compiled on a database. These results will be analysed at various stages by Statisticians to look into markers of diagnosis, prognosis and treatment response. These results may be presented at Rheumatology meetings and published in the medical literature. All data will be fully anonymised.

## **13. Who has reviewed the study?**

The Leeds (West) Research Ethics Committee has reviewed this study.

## **14. Contact for further information**

In the event of study related questions or problems, please contact the following telephone number:

Chapel Allerton Hospital

David Pickles (Lead Research Nurse)

During working hours

- Research room phone: 0113 3924729 After hours
- Ward C2 phone: 0113 3924202

**Finally, thank you for taking the time to read the information and considering whether to take part in this study.**

**CONSENT FORM -**

Title of Project: Rheumatoid Arthritis Disease Research (RADAR)

SUBJECT INITIALS 

--	--	--

 SUBJECT NO. 

--

 SUBJECT DOB 

--	--	--

**Name of Researcher:**.....

Please initial line

1. I confirm that I have read and understand the information sheet dated 19.01.15 (version 8.0) for the above study, and have had the opportunity to ask questions. I understand that my participation is voluntary and that I am free to withdraw at any time without my medical care or legal rights being affected. I agree to take part in the study. \_\_\_\_\_
2. I understand that my medical records may be looked at by authorised individuals from the Sponsor for the study, the UK Regulatory Authority or the Independent Ethics Committee in order to check that the study is being carried out correctly. I give permission, provided that strict confidentiality is maintained, for these bodies to have access to my medical records for the above study and any further research that may be conducted in relation to it. I also give permission for a copy of my consent form to be sent to the Sponsor for the study. \_\_\_\_\_
3. I understand that even if I withdraw from the clinical study, the data and samples collected from me will be used in analysing the results of the study, unless I specifically withdraw consent for the laboratory study. I understand that my identity will remain anonymous. \_\_\_\_\_
4. I consent to the storage including electronic and personal information for the purposes of this study. I understand that any information that could identify me will be kept strictly confidential and that no personal information will be included in the study report or other publication. \_\_\_\_\_
5. I agree for any previous samples collected on me during previous ethically approved studies to be used for the purposes of the current study. \_\_\_\_\_
6. I agree for the results from this study to be submitted to the Rheumatology department BioBank and for this to be updated from my medical and electronic records in order to support additional research activities. \_\_\_\_\_
7. I consent for my GP to be informed of my participation in this study \_\_\_\_\_

\_\_\_\_\_  
 Name of Patient

\_\_\_\_\_  
 Date

\_\_\_\_\_  
 Signature



To understand whether a person will develop inflammatory arthritis and, if they do, how severe it will be and how they will respond to treatment are very important questions. To try and answer these questions involves the analysis of blood and urine tests of people with possible, probable or definite inflammatory arthritis.

Biological tests (blood and urine):

Many autoimmune, inflammatory and degenerative diseases are associated with the presence of specific changes in an individual's immune or tissue repair systems, which may be a reflection of their genetic makeup. This can lead to alterations in the different components of the immune and tissue repair systems or in the proteins that are present in the blood or urine and produced by the tissues. We feel these changes may be important for the development of either the disease itself, specific disease features or complications or that they may even help us predict response to different treatments. We would like to perform some further research to gain a better understanding of the biology of your cells, proteins and genes to work out how they may contribute to these various diseases.

Often research studies involve working with other national researchers in this field. Therefore, we would also like to store some of the components of your blood, such as the proteins and a sample of your DNA and urine. These samples will only be used in future studies that continue with this agreed line of research. In order to give us a permanent source of specific genes and proteins we or third parties in the UK, in Europe, Switzerland and/or USA may also like to make some cell-lines from specific proteins or cells from your blood in the laboratory.

## **2. Why have I been chosen?**

You have been invited to participate in this sub-study because you have been identified as having symptoms that can be associated with inflammatory arthritis. We want to assess people like you over time to learn more about inflammatory arthritis.

## **3. Do I have to take part?**

It is up to you to decide whether or not to take part in the study. If you decide to take part you will be given this information sheet to keep and be asked to sign a consent form. If you decide to take part you are still free to withdraw from the study at any time and without giving a reason. This will not affect the standard of care you receive.

## **4. What will happen to me if I take part?**

If you decide to take part you will be asked to sign an informed consent sheet, and you will be given a copy of the information sheet to keep. In addition to the standard clinical care described in the main study information sheet, as part of the sub-study, we would like to ask you to donate a urine sample and an additional blood sample (up to a maximum of 5 tablespoons) at each study visit which will be taken at the same time as your routine clinic

blood tests. Sometimes, if a joint is inflamed and fluid is present, joint fluid may also be collected as part of your study visits. The samples will be stored solely for future laboratory based research. We might also ask your permission to obtain these samples at extra times to coincide with changes in your disease.

If you have donated any samples for one of our previous studies, we would now like to use those samples in this study.

## **5. What do I have to do?**

There should be no reason to change your current way of life if you participate in this study. If you decide to take part in the study you will be asked to:

- Tell the Rheumatology doctor about any illness that you currently have or have had in the past.
- Tell your Rheumatology doctor about any other medications you are taking, as well as those supplied by your doctor.
- Keep the appointments for your study visits.
- Tell your Rheumatology doctor about any changes in your health that occurs during the study

## **6. What will happen to my samples?**

We will remove your personal details from all research samples after separation into their constituent parts. However, it will be possible to link the clinical and laboratory databases through a unique laboratory code to enable use to study long-term disease outcomes and response to future therapies you may receive.

Your samples will be used in various research projects. Some of your samples and associated data may be transferred abroad, under confidentiality agreements, to other public or private research groups of which the results may be patented, published or used for commercial purposes.

In the context of the present study, we will examine the cells in the blood and to study certain proteins involved in the control and activation of your immune or tissue repair systems. Your DNA will be isolated from your blood and will be used to help us find out which genes are important in inflammatory arthritis and in predicting the long-term disease outcome or response to treatment. Other components of your blood will be isolated, frozen and stored for subsequent studies.

In genetic studies we compare how often the gene(s) of interest are found in people with a disease to individuals that don't have it ("controls"). We would therefore like to store some of your DNA and other biological materials to form part of our "disease DNA bank" that we can use in our current and future studies. This is purely for research purposes and you will not be told the results of the tests on your samples. The information may be shared with other research groups conducting similar investigations. This is because large numbers of individuals are required to undertake such studies and they now need to be undertaken at

the National or International level. Insurance companies, however, may ask you whether you have previously had genetic tests. Should this situation arise, we advise you to answer “no” in your insurance policy application form, as the tests carried out have no relevance to insurance.

If consent has been given, any samples that are not used at the end of this study will be transferred to a Research Tissue Bank and will be used in future studies within this research area.

## **7. What if I do not wish to take part?**

All studies are always completely voluntary. You do not have to consent to having blood taken for this purpose. If you do not wish to have some or all of these tests, it will not affect your standard of care or your ability to enter the main study.

If you decide to take part, you are still free to withdraw at any time and without giving a reason. This will not affect the standard of care you receive.

You should be aware that if in the future you experience the loss of capacity (i.e. ability to agree to continue to take part in the study), the research team would retain tissue and personal data collected and continue to use it confidentially. This could include further research after the current project has ended.

## **8. What are the possible disadvantages and risks of taking part?**

There are no additional health risks associated with donating specimens for research purposes if they are taken as part of a normal diagnostic procedure. If we are taking a blood sample at a different time from your routine tests, the only risks would be minor bruising. If you are a patient and anything in the procedure for obtaining your specimens were to go wrong, the normal complaint mechanisms of the NHS are open to you.

## **9. What are the possible benefits of taking part?**

There is good evidence that treating patients with inflammatory arthritis is much better if it is done as early as possible. As we are seeing you regularly in clinic we will be able to commence treatment at an early stage if this is needed.

It cannot be guaranteed that you will gain any other personal benefit from this study. However, beneficial information may be acquired for patients who develop inflammatory arthritis and may help us to treat these future patients more effectively.

## **10. What if new information becomes available?**

Sometimes during the course of a research project, new information becomes available. If this happens, your Rheumatology doctor will tell you about it and discuss with you whether you want to continue in the study. If you decide to withdraw, your Rheumatology doctor will make arrangements for your care to continue. If you decide to continue in the study, you will be asked to sign an updated consent form after reading a new information sheet.

Also, on receiving new information, your Rheumatology doctor might consider it to be in your best interests to withdraw you from the study. He/she will explain the reasons and arrange for your care to continue.

#### **11. What happens when the research study stops?**

Once the study is over, your research doctor will decide whether you should continue to be monitored in the Rheumatology Unit or whether you can be followed by your GP.

#### **12. Other information**

All your written and computer records will be kept strictly confidential at all times. Data Protection Act regulations have been complied with to ensure confidentiality.

#### **13. What will happen to the results of the research study?**

The results from the study will be compiled on a database. These results will be analysed at various stages by Statisticians to look into markers of diagnosis, prognosis and treatment response. These anonymised results may be presented at Rheumatology meetings and published in the medical literature. They may also form part of a higher research degree being undertaken by one of the study doctors. Your clinical data will be stored within a planned secure BioBank and may be used for additional research activities. As part of this research we are additionally seeking your consent to update the clinical data from your medical records or other electronic data records. Researchers undertaking this additional research will not be able to identify you and all projects will be reviewed by a Research Ethics Committee.

#### **14. Who is organising and funding the research?**

This sub-study is being funded by Arthritis Research UK. The University of Leeds are organising and responsible for the conduct of the study. The study doctor and nurse will not receive any payment for conducting this research study.

#### **15. Are there any other third parties involved in the research?**

Researchers may collaborate with other researchers and third parties in the UK, the EU, Switzerland and the USA. They may work in universities, hospitals or the private sector. Your tissue or other samples will not, however, be sold for profit. Scientific

investigation will be undertaken using the samples provided by study participants. We may share anonymised samples and clinical data collected in this study with regional, national and international collaborators who are conducting studies with similar research themes. This will allow a greater range of technologies to be applied and help to facilitate specific research objectives.

**16. Who has reviewed the study?**

This study has been reviewed by the independent ethics committee called the Leeds West Research Ethics Committee. This committee is appointed to determine that research studies are ethical and do not impair the rights or well-being of patients. We have received approval by this committee to be able to do this research study.

**17. What if there is a problem?**

If you have a concern about any aspect of this study, you should ask to speak with the researchers who will do their best to answer your question. If you remain unhappy and wish to complain formally, you can do this through the NHS Complaints Procedure. Details can be obtained from the hospital.

**18. Contact for further information**

If you need any further information please do not hesitate to contact your study doctor or nurse. You should also contact your GP for independent advice should you so desire.

**CONSENT FORM**

Rheumatoid Arthritis Disease Research – Biological Sub-study (RADAR)

SUBJECT INITIALS 

--	--	--

 SUBJECT NO. 

--

 SUBJECT DOB 

--	--	--

Please initial line

1. I confirm that I have read and understand the information sheet dated 24.02.2017 (version 6.0) for the above study, and have had the opportunity to ask questions. I understand that my participation is voluntary and that I am free to withdraw at any time without my medical care or legal rights being affected. I agree to take part in the study. \_\_\_\_\_
  
2. I understand that my medical records may be looked at by authorised individuals from the Sponsor for the study, the UK Regulatory Authority or the Independent Ethics Committee in order to check that the study is being carried out correctly. I give permission, provided that strict confidentiality is maintained, for these bodies to have access to my medical records for the above study and any further research that may be conducted in relation to it. I also give permission for a copy of my consent form to be sent to the Sponsor for the study. \_\_\_\_\_
  
3. I understand that even if I withdraw from the clinical study, the data and samples collected from me will be used in analysing the results of the study, unless I specifically withdraw consent for the laboratory study. I understand that my identity will remain anonymous. \_\_\_\_\_
  
4. I consent to the storage including electronic and personal information for the purposes of this study. I understand that any information that could identify me will be kept strictly confidential and that no personal information will be included in the study report or other publication.  
 \_\_\_\_\_
  
5. I consent to the transfer abroad, under confidentiality agreements, of my samples and associated data to other public or private research groups of which the results \_\_\_\_\_ may be patented, published or used for commercial purposes
  
6. I agree to the samples being stored for future ethically approved research and for any remaining samples to be deposited in a Research Tissue Bank. \_\_\_\_\_
  
7. I agree to have genetic tests done on samples for research purposes. \_\_\_\_\_
  
8. I agree for any previous samples collected on me during previous ethically approved studies to be used for the purposes of the current study. \_\_\_\_\_
  
9. I agree for the results from this study to be submitted to the Rheumatology department BioBank and for this to be updated from my medical and electronic records in order to support additional research activities.  
 \_\_\_\_\_
  
10. I consent for my GP to be informed of my participation in this study





	Clotted Red Top	Sodium Citrate Blue top	EDTA Purple Top	Lithium Heparin Green Top	PAXgene Clear top	URINE White top
<b>No. tubes/mls</b>						
<b>Max. total vol.</b>						10mls

I confirm that the patient has been appropriately consented to the biological sub-study for sample collection and storage.

\*Consent confirmed by .....  
*(print name)*
*(signature)*

\*Samples collected by..... *(print name)*

RADAR\_BMC\_Research\_Blood\_Request\_Form\_ v4.0, 9\_May\_2017

**Chapel Allerton Hospital Laboratory- RESEARCH BLOOD REQUEST FORM**  
*Please contact laboratory staff on CAH ext 24730 to arrange collection, Thank you.*

**RA cohort: EAC & BMC**

<b>Researcher</b>	<b>Contact details</b>	<b>New onset RA</b>	<b>Pre/post biologic</b>
-------------------	------------------------	---------------------	--------------------------

**Additional instruction**

\*Please identify patients who may be agreeable to have ultrasound-guided synovial biopsy. Initially, to prioritise patients commencing tocilizumab and tofacitinib

\*\* Please identify patients who may be agreeable to have ultrasound-guided synovial biopsy – irrespective of treatment being commenced; early and/or later RA.  
 For patient attending joint aspiration, please also request bloods as indicated above  
 Note: Synovial fluid can be collected in a sterilin tube (white top universal container).

## Automated clean-up gating for single cell population script

```
##Script for automated clean up gating of cyTOF FCS using gaussian parametes in addition to DNA, CD45 and CD3 vs CD19 doublet removal
```

```
##Script adapted from Camilla Pang August 2019
```

```
##Edited by Emma Sutton
```

```
#load libraries
```

```
library(DT)
```

```
library(readxl)
```

```
library(rmarkdown)
```

```
library(ggplot2)
```

```
library(mixtools)
```

```
library(flowCore)
```

```
library(kza)
```

```
library(dplyr)
```

```
library(ggcyto)
```

```
library(tidyverse)
```

```
library(corrplot)
```

```
library(tidyr)
```

```
library( flowStats )
```

```
library( FlowSOM )
```

```
library(icesTAF)
```

```
#set working directory to find FCS files
```

```
setwd("Tejal/Batch_2_debarcoded/Test2")
```

```
fnames <- list.files( pattern = 'fcs' , full.names = TRUE , recursive = TRUE )
```

```

fnames <- fnames[ !grepl( fnames , pattern = 'gated' , ignore.case = TRUE ) ]

lapply( fnames , function( ifile ) {

#make output directry with file name

mkdir(paste( gsub( x = ifile , pattern = '\\.fcs' , replacement = " " )))

#set output directory name
output <- (paste( gsub( x = ifile , pattern = '\\.fcs' , replacement = " " )))

#read FCS file into flowFrame

file_name <- ifile
fcs_raw <- read.FCS(ifile, transformation=FALSE, truncate_max_range = FALSE)

#Map channels and epitopes

fcs_channel_descr <- fcs_raw@parameters@data$desc %>% as.data.frame()
fcs_channel_names <- fcs_raw@parameters@data$name %>% as.data.frame()
fcs_channel <- data.frame(fcs_channel_descr, fcs_channel_names)

# rename rows in data.frame using colnames.

colnames(fcs_channel) <- c('Epitope','Channel')
head(fcs_channel)

#Transform data using ArcSinH cofactor of 5 (for cyTOF data)

fcs_ch_name <- fcs_raw@parameters@data$name
tf <- transformList(from = fcs_ch_name, tfun = asinh)

```

```

fcs_trans <- tf %on% fcs_raw

## Automated cleanup

#Gate 1: Remove beads and clean up width

p <- ggcyto(fcs_trans, aes(x = 'Width', y = 'Ce140Di'))
p <- p + geom_hex(bins= 30)
tiff(file = paste(output,gsub( x = ifile , pattern = '\\.fcs' , replacement = " ) , "_width_pregate.tiff" , sep = " ) ,
width = 4, height = 4, units = "in", res = 300)

plot(p)
dev.off()

norm <- norm2Filter('Width', 'Ce140Di', scale = 4)

fcs_norm1 <- Subset(fcs_trans, norm)
p <- ggcyto(fcs_norm1, aes(x = 'Width', y = 'Ce140Di'))
p <- p + geom_hex(bins= 30)
tiff(file = paste(output,gsub( x = ifile , pattern = '\\.fcs' , replacement = " ) , "_width.tiff" , sep = " ) , width = 4,
height = 4, units = "in", res = 300)

plot(p)
dev.off()

##Gate 2: Clean up on offset

p <- ggcyto(fcs_norm1, aes(x = 'Offset', y = 'Ce140Di'))
p <- p + geom_hex(bins= 100)
tiff(file = paste(output, gsub( x = ifile , pattern = '\\.fcs' , replacement = " ) , "_offset_pregate.tiff" , sep = " ) ,
width = 4, height = 4, units = "in", res = 300)

plot(p)
dev.off()

norm2 <- norm2Filter('Offset', 'Ce140Di', scale = 2.8) # 1.35 works high intensity, but emma said 2.

```

```

fcs_norm2 <- Subset(fcs_norm1, norm2)

p <- ggcyto(fcs_norm2, aes(x = 'Offset', y = 'Ce140Di'))
p <- p + geom_hex(bins= 30)
tiff(file = paste(output, gsub( x = ifile , pattern = '\\.fcs' , replacement = " ) , "_offset.tiff" , sep = " ), width = 4,
height = 4, units = "in", res = 300)

plot(p)
dev.off()

###Gate 3: Clean up on center

p <- ggcyto(fcs_norm2, aes(x='Center', y = 'Ce140Di'))
p <- p + geom_hex(bins= 100)
tiff(file = paste(output, gsub( x = ifile , pattern = '\\.fcs' , replacement = " ) , "_center_pregate.tiff" , sep = " ),
width = 4, height = 4, units = "in", res = 300)

plot(p)
dev.off()

norm3 <- norm2Filter('Ce140Di', 'Center', scale = 5) # CAN BE MORE GENEROUS
fcs_norm3 <- Subset(fcs_norm2, norm3)

p <- ggcyto(fcs_norm3, aes(x = 'Center', y = 'Ce140Di'))
p <- p + geom_hex(bins= 30)
tiff(file = paste(output, gsub( x = ifile , pattern = '\\.fcs' , replacement = " ) , "_center.tiff" , sep = " ), width = 4,
height = 4, units = "in", res = 300)

plot(p)
dev.off()

###Gate 4: Clean up on residual

p <- ggcyto(fcs_norm3, aes(x='Residual', y = 'Ce140Di'))
p <- p + geom_hex(bins= 100)
tiff(file = paste(output, gsub( x = ifile , pattern = '\\.fcs' , replacement = " ) , "_residual_pregate.tiff" , sep = " ),
width = 4, height = 4, units = "in", res = 300)

plot(p)
dev.off()

```

```

norm4 <- norm2Filter('Ce140Di', 'Residual', scale = 3)
fcs_norm4 <- Subset(fcs_norm3, norm4)

p <- ggcyto(fcs_norm4, aes(x = 'Residual', y = 'Ce140Di'))
p <- p + geom_hex(bins = 30)
tiff(file = paste( output, gsub( x = ifile , pattern = '\\.fcs' , replacement = " ) , "_residual.tiff" , sep = " ) , width =
4, height = 4, units = "in", res = 300)

plot(p)
dev.off()

```

```

##Gate 5: DNA1 vs CD45

```

```

range2 <- rangeGate(fcs_norm4, 'Y89Di', alpha="min", sd=4, plot= TRUE, borderQuant=0.1, absolute=TRUE,
filterId="defaultRectangleGate", positive= TRUE, refLine= NULL)

p <- ggcyto(fcs_norm4, aes(x = 'Ir191Di', y = 'Y89Di'))
p <- p + geom_hex(bins = 30)
tiff(file = paste( output, gsub( x = ifile , pattern = '\\.fcs' , replacement = " ) , "_DNA_pregate.tiff" , sep = " ) ,
width = 4, height = 4, units = "in", res = 300)

plot(p + geom_gate(range2))
dev.off()

```

```

fcs_range2 <- Subset(fcs_norm4, range2) #LATEST FILTRED DATA

```

```

p <- ggcyto(fcs_range2, aes(x = 'Ir191Di', y = 'Y89Di'))
p <- p + geom_hex(bins = 30)
tiff(file = paste(output, gsub( x = ifile , pattern = '\\.fcs' , replacement = " ) , "_DNA.tiff" , sep = " ) , width = 4,
height = 4, units = "in", res = 300)

plot(p)
dev.off()

```

```

extradensity <- norm2Filter('Ir191Di', 'Y89Di', scale = 5)

```

```

fcs_range2 <- Subset(fcs_range2, extradensity)

p <- ggcyto(fcs_range2, aes(x = 'Ir191Di', y = 'Y89Di'))
p <- p + geom_hex(bins= 30)
tiff(file = paste(output, gsub( x = ifile , pattern = '\\.fcs' , replacement = " " ) , "_CD45.tiff" , sep = " " ), width = 4,
height = 4, units = "in", res = 300)

plot(p)
dev.off()

##Gate 6: Live cell gate (if needed, define if rhodamine or cisplatin)

#p <- ggcyto(fcs_range2, aes(x = 'Y89Di', y = 'Rh103Di'))
#p <- p + geom_hex(bins = 30)
#p

#rhod_range <- norm2Filter('Rh103Di', 'Y89Di', scale = 8) # 1.35 works high intensity, but emma said 2.

#subset FCS
#fcs_range3 <- Subset(fcs_range2, rhod_range)

#p <- ggcyto(fcs_range3, aes(x = 'Y89Di', y = 'Rh103Di'))
#p <- p + geom_hex(bins= 30)
#p

#Gate 7: Gate for singlets (NB object name change as skipped Rh gate (fcs_range2 instead of fcs_range3))

p <- ggcyto(fcs_range2, aes(x = 'Nd142Di', y = 'Er170Di'))
p <- p + geom_hex(bins = 30)
tiff(file = paste(output, gsub( x = ifile , pattern = '\\.fcs' , replacement = " " ) , "_singlets_pregate.tiff" , sep = " " ),
width = 4, height = 4, units = "in", res = 300)

plot(p)
dev.off()

quad <- quadrantGate(fcs_range2, c("Nd142Di", "Er170Di"), alpha=c(0.5, 0.25), plot=TRUE)

```

```

# get quadrant value splits for each channel
bounds <- quad@boundary %>% as.data.frame()
CD19_threshold <- bounds$.[1]
CD3_threshold <- bounds$.[2]

bound_table <- data.frame(CD19_threshold, CD3_threshold)
quadbounds <- bound_table %>% t() %>% as.data.frame()
colnames(quadbounds) <- c("Threshold")
quadbounds

quad_fcs <- split(fcs_range2, quad, population=list(keep=c("142Nd_CD19-
170Er_CD3+", "142Nd_CD19+170Er_CD3-", "142Nd_CD19-170Er_CD3-")))

p <- ggcyto(quad_fcs$keep, aes(x = 'Nd142Di', y = 'Er170Di'))
p <- p + geom_hex(bins = 30)
tiff(file = paste( output, gsub( x = ifile , pattern = '\\.fcs' , replacement = " ) , "_singlets.tiff" , sep = " ) , width =
4, height = 4, units = "in", res = 300)
plot(p + geom_gate(quad))
dev.off()

full_gated_fcs <- quad_fcs$keep

p <- ggcyto(quad_fcs$keep, aes(x = 'Nd142Di', y = 'Er170Di'))
p <- p + geom_hex(bins = 30)
tiff(file = paste( output, gsub( x = ifile , pattern = '\\.fcs' , replacement = " ) , "_final.tiff" , sep = " ) , width = 4,
height = 4, units = "in", res = 300)
plot(p)
dev.off()

# make a cell count table or each cleaning and gating step

raw_cells <- nrow(fcs_raw)
transformed_cells <- nrow(fcs_trans)
width_clean <- nrow(fcs_norm1)

```

```

offset_clean <- nrow(fcs_norm2)
center_clean <- nrow(fcs_norm3)
residual_clean <- nrow(fcs_norm4)
cd45_cells <- nrow(fcs_range2)
singlets <- nrow(full_gated_fcs)

gating_table <- data.frame(raw_cells, transformed_cells, width_clean, offset_clean, center_clean,
residual_clean, cd45_cells, singlets)

cell_counts <- gating_table %>% t() %>% as.data.frame()
colnames(cell_counts) <- c("Cell Counts")
cell_counts
write.table(cell_counts, file = paste(output, gsub(x = ifile, pattern = '\\.fcs', replacement = "" ), "_counts.csv"
, sep = "" ))

## Write out fully gated FCS file

write.FCS(full_gated_fcs, file = paste(output, gsub(x = ifile, pattern = '\\.fcs', replacement = "" ), "_gated.fcs"
, sep = "" ), what = "numeric")

})

```

# Automated compensation script

```
#load libraries
library(flowCore)
library(CATALYST)

fnames <- list.files( 'FCS_files' , pattern = 'fcs' , full.names = TRUE , recursive = TRUE )
fnames <- fnames[ !grepl( fnames , pattern = 'bead' , ignore.case = TRUE ) ]
fnames <- fnames[ !grepl( fnames , pattern = 'comp' , ignore.case = TRUE ) ]

#read data (change file path to correspond to single stain bead data)
data <- read.FCS('OneComp_141_176_EQ_01_1.fcs', transformation = FALSE, truncate_max_range = FALSE)

#assign single stain bead channels
channels <-
c(141,142,143,144,145,146,147,148,149,150,151,152,153,154,155,156,158,159,160,161,162,163,164,165,166,
167,168,169,170,171,172,173,174,175,176)

#determine single stain populations
comp <- assignPrelim(x=data, y=channels)
comp <- estCutoffs(x=comp)
comp <- applyCutoffs (x=comp)

#generate spill over matrix and plot
SpillMat <- computeSpillmat(x=comp)
na <- replace(SpillMat, is.na(SpillMat),0)
plotSpillmat(bc_ms=channels, SM=na, plotly=FALSE)

lapply( fnames , function( ifile ) {

#read in cell data for compensation (change file path to choose data to be compensated)
cells <- read.FCS(ifile ,transformation = FALSE, truncate_max_range = FALSE)
```

```

#apply compenstation based on spill matrix from single stain beads nnls
cells_nnls <- compCytof(x=cells, y=na, method = "nnls")

#apply compenstation based on spill matrix from single stain beads flow
#cells_flow <- compCytof(x=cells, y=SpillMat, method = "flow")

#write out compensated fcs file (change outpute path and file name as desired)
write.FCS(cells_nnls, file = paste( gsub( x = ifile , pattern = '\\.fcs' , replacement = " " ) , "_comp_nnls.fcs" , sep = " " ) ,what = "numeric")

#write out compensated fcs file (change outpute path and file name as desired)
#write.FCS(cells_flow, file = paste( gsub( x = ifile , pattern = '\\.fcs' , replacement = " " ) , "_comp_flow.fcs" , sep = " " ) ,what = "numeric")

})

#load libraries
library(flowCore)
library(CATALYST)

fnames <- list.files( 'FCS_files' , pattern = 'fcs' , full.names = TRUE , recursive = TRUE )
fnames <- fnames[ !grepl( fnames , pattern = 'bead' , ignore.case = TRUE ) ]
fnames <- fnames[ !grepl( fnames , pattern = 'comp' , ignore.case = TRUE ) ]

#read data (change file path to correspond to single stain bead data)
data <- read.FCS('OneComp_141_176_EQ_01_1.fcs', transformation = FALSE, truncate_max_range = FALSE)

#assign single stain bead channels
channels <-
c(141,142,143,144,145,146,147,148,149,150,151,152,153,154,155,156,158,159,160,161,162,163,164,165,166,
167,168,169,170,171,172,173,174,175,176)

#determine single stain populations
comp <- assignPrelim(x=data, y=channels)
comp <- estCutoffs(x=comp)
comp <- applyCutoffs (x=comp)

```

```

#generate spill over matrix and plot
SpillMat <- computeSpillmat(x=comp)
na <- replace(SpillMat, is.na(SpillMat),0)
plotSpillmat(bc_ms=channels, SM=na, plotly=FALSE)

lapply( fnames , function( ifile ) {

#read in cell data for compensation (change file path to choose data to be compensated)
cells <- read.FCS(ifile ,transformation = FALSE, truncate_max_range = FALSE)

#apply compenstation based on spill matrix from single stain beads nnls
cells_nnls <- compCytotf(x=cells, y=na, method = "nnls")

#apply compenstation based on spill matrix from single stain beads flow
#cells_flow <- compCytotf(x=cells, y=SpillMat, method = "flow")

#write out compensated fcs file (change outpute path and file name as desired)
write.FCS(cells_nnls, file = paste( gsub( x = ifile , pattern = '\\.fcs' , replacement = " " ) , "_comp_nnls.fcs" , sep = " " ) ,what = "numeric")

#write out compensated fcs file (change outpute path and file name as desired)
#write.FCS(cells_flow, file = paste( gsub( x = ifile , pattern = '\\.fcs' , replacement = " " ) , "_comp_flow.fcs" , sep = " " ) ,what = "numeric" ) })

```

# Automated normalisation script

```
## Load libraries NB requires updated version of FlowSOM in order to have GetClusters function. Install  
FlowSOM from github, remotes::install_github("saeyslab/FlowSOM)
```

```
library(CytoNorm)
```

```
library(FlowSOM)
```

```
dir <- ("Tejal/CytoNorm")
```

```
##data to be normalised has been downsampled so column names changed, need to change column names of  
batch controls to match - downsample batch controls to 25000 cells per file
```

```
fnames <- train_data$Path
```

```
fnames
```

```
fnames <- fnames[ !grepl( fnames , pattern = 'downsample' , ignore.case = TRUE ) ]
```

```
library(flowCore)
```

```
##downsample each FCS file to desired cell number per file and write out new FCS
```

```
for(i in fnames) {
```

```
  #read in FCS file
```

```
  fcs <- read.FCS(filename=i, transformation=FALSE)
```

```
  #creat matrix
```

```
  exprs <- fcs@exprs
```

```
  #define marker names and remove metals
```

```
  marker_names <- gsub(pattern = ".*_", replacement = "", x = as.vector(fcs@parameters@data$desc))
```

```
  #match colnames in matrix with marker names from original FCS
```

```
  colnames(exprs)[which(!is.na(marker_names))] <- marker_names[which(!is.na(marker_names))]
```

```

#downsample matrix to get desired number of cells
matrix_down <- exprs[sample(nrow(exprs), 25000),]

#create new flowFrame from downsampled matrix
new_fcs <- new("flowFrame", exprs=as.matrix(matrix_down))

#write out downsampled FCS files to folder
write.FCS(new_fcs, file = paste(gsub( x = i , pattern = '\\.fcs' , replacement = " ) , "_downsample.fcs" , sep = "
) ,what = "numeric")
}

##Load in example data for CytoNorm package

files <- list.files(dir, pattern = "fcs$")

## create data frame of data
data <- data.frame(File = files,
  Path = file.path(dir, files),
  Type = stringr::str_match(files, "([12]).fcs")[, 2],
  Batch = stringr::str_match(files, "batch[0-9]*")[, 1],
  stringsAsFactors = FALSE)

#assign data as training or validation
data$Type <- c("2" = "Train", "1" = "Validation")[data$Type]

train_data <- dplyr::filter(data, Type == "Train")
validation_data <- dplyr::filter(data, Type == "Validation")

## read in fcs files and transform channels
ff <- flowCore::read.FCS(data$Path[1])

## set channels to use for flowSOM - used major lineage markers for flowSOM

```

```
channels <-  
flowCore::colnames(ff)[c(3,13,14,15,16,17,18,19,20,21,22,23,24,25,26,27,28,29,30,31,32,33,34,35,36,37,38,39  
,40,41,42,43,44,45,46,47,54)]
```

```
## batch control not been transformed but healthy controls from run have - how to sort this?
```

```
transformList <- 0
```

```
transformList.reverse <- 0
```

```
#tdff <- flowCore::read.FCS(train_data$Path[1])
```

```
#tdff$desc <- gsub("-", "_", tdff$desc)
```

```
##Build FlowSOM map on training data
```

```
fsom <- prepareFlowSOM(train_data$Path,  
  channels,  
  nCells = 25000,  
  FlowSOM.params = list(xdim = 10,  
    ydim = 10,  
    nClus = 8,  
    scale = FALSE),  
  transformList = NULL,  
  seed = 1)
```

```
## check cv values for different cluster number to evaluate appropriateness of clustering (only used if clusters  
are not effected by batch effects)
```

```
## currently get error 'GetClusters' is not an exported object from namespace:FlowSOM
```

```
cvs <- testCV(fsom,  
  cluster_values = c(4:8))
```

```
##Training the model
```

```
model <- CytoNorm.train(files = train_data$Path,  
  labels = train_data$Batch,  
  channels = channels,  
  transformList = NULL,
```

```

FlowSOM.params = list(nCells = 25000,
                      xdim = 10,
                      ydim = 10,
                      nClus = 8,
                      scale = FALSE),
normMethod.train = QuantileNorm.train,
normParams = list(nQ = 101,
                  goal = "mean"),
seed = 1,
verbose = TRUE)

```

##Normalising the data

```

channels_validation <-
flowCore::colnames(ff)[c(3,13,14,15,16,17,18,19,20,21,22,23,24,25,26,27,28,29,30,31,32,33,34,35,36,37,38,39
,40,41,42,43,44,45,46,47,54)]

```

```

transform_list_validation <- flowCore::transformList(channels_validation,
                                                    cytofTransform)

```

```

transformList.reverse <- flowCore::transformList(channels_validation,
                                                cytofTransform.reverse)

```

```

CytoNorm.normalize(model = model,
                  files = validation_data$Path,
                  labels = validation_data$Batch,
                  transformList = NULL,
                  transformList.reverse = NULL,
                  normMethod.normalize = QuantileNorm.normalize,
                  outputDir = "Normalized_10012020",
                  prefix = "Norm_",
                  clean = TRUE,
                  verbose = TRUE)

```

# Diffcyt script

```
#load libraries
```

```
library(CATALYST)
```

```
library(flowCore)
```

```
library(diffcyt)
```

```
library(readxl)
```

```
library(caret)
```

```
#set working directory
```

```
setwd(" ")
```

```
#setwd("")
```

```
#getwd()
```

```
fnames <- list.files( pattern = 'fcs' , full.names = TRUE , recursive = TRUE )
```

```
fnames <- fnames[ !grepl( fnames , pattern = 'downsample' , ignore.case = TRUE ) ]
```

```
#make output directry to save downsampled FCS used for analysis
```

```
#mkdir("downsampled_files")
```

```
#set output directory name
```

```
#output <- ("downsampled_files")
```

```
#downsample each FCS file to desired cell number per file and write out new FCS
```

```
for(i in fnames) {
```

```
  #read in FCS file
```

```
  fcs <- read.FCS(filename=i, transformation=FALSE)
```

```

#creat matrix
exprs <- fcs@exprs

#define marker names and remove metals
marker_names <- gsub(pattern = ".*_", replacement = "", x = as.vector(fcs@parameters@data$desc))

#match colnames in matrix with marker names from original FCS
colnames(exprs)[which(!is.na(marker_names))] <- marker_names[which(!is.na(marker_names))]

#downsample matrix to get desired number of cells
matrix_down <- exprs[sample(nrow(exprs), 27000),]

#create new flowFrame from downsampled matrix
new_fcs <- new("flowFrame", exprs=as.matrix(matrix_down))

#write out downsampled FCS files to folder
write.FCS(new_fcs, file = paste(gsub( x = i , pattern = '\\.fcs' , replacement = " " ) , "_downsample.fcs" , sep = "
" ), what = "numeric")
}

#Import experiment metadata from excel file and format
##update names in metadata file to reflect downsampled file names from above if needed and change to
directory containing downsampled files
setwd(output)
metadata_filename <- "metadata.xlsx"
md <- read_excel
md <- read_excel(metadata_filename)
md
md$condition <- factor(md$condition, levels = c("PB", "SF"))
color_conditions <- c("#6A3D9A", "#FF7F00")
names(color_conditions) <- levels(md$condition)

# Import fcs files into flowset

```

```
fcs_filename <- md$file_name # issue with ordering of FCS files if read in as list? Read in using names directly from metadata file
```

```
#fcs_filename <- list.files(pattern = "\\\\.fcs")
```

```
fcs_raw <- read.flowSet(fcs_filename, transformation = FALSE, truncate_max_range = FALSE)
```

```
fcs_raw
```

```
#Import information on experiment panel and format
```

```
panel_filename <- "antibody.xlsx"
```

```
panel <- read_excel(panel_filename)
```

```
head(data.frame(panel))
```

```
panel$Antigen <- gsub("-", "_", panel$Antigen)
```

```
panel_fcs <- pData(parameters(fcs_raw[[1]]))
```

```
head(panel_fcs)
```

```
#rename markers in desc to conform with antigen names in panel DO NOT NEED FOR DOWNSAMPLED FILES AS METALS ALREADY REMOVED
```

```
#panel_fcs$name <- sapply(strsplit(as.character(panel_fcs$name), "_"), .subset, 2)
```

```
#clean up names in panel_fcs description
```

```
##panel_fcs$desc replced with panel_fcs$name for downsampled files as do not contain info in $desc slot - revert back if not using
```

```
panel_fcs$name <- gsub("-", "_", panel_fcs$name)
```

```
#define lineage markers
```

```
lineage_markers <- panel$Antigen[panel$Lineage==1]
```

```
lineage_markers
```

```
#define functional markers
```

```
functional_markers <- panel$Antigen[panel$Functional==1]
```

```
functional_markers
```

```
#check markers in panel match description parameters in fcs files
```

```
all(lineage_markers %in% panel_fcs$name)
```

```
all(functional_markers %in% panel_fcs$name)
```

```
#Would usually include arcsinh transformation here however already applied as data gated using auto-gating script
```

```
fcs <- fsApply(fcs_raw, function(x,cofactor=5){  
  colnames(x) <- panel_fcs$name  
  expr <- exprs(x)  
  expr <- (expr[,c(lineage_markers,functional_markers)])  
  exprs(x) <- expr  
  x  
})
```

```
expr <- fsApply(fcs,exprs)
```

```
dim(expr)
```

```
sample_ids <- rep(md$sample_id, fsApply(fcs_raw, nrow))
```

```
#Normalisation, mean = 0 SD +/- 1
```

```
library(matrixStats)
```

```
#rng <- colQuantiles(expr, probs = c(0.01,0.99))
```

```
#expr01 <- t((t(expr - rng[,1]) / rng[,2] - rng[,1]))
```

```
#expr01[expr01 <0] <- 0
```

```
#expr01[expr01 >1] <- 1
```

```
expr01 <- expr
```

```
#plot figure of cell numbers in each fcs file
```

```

library(ggplot2)
library(reshape2)

#create table of cell numbers

cell_table <- table(sample_ids)
cell_table

#cell_table <- table(downsample)

#plot cell counts

ggdf <- data.frame(sample_id = names(cell_table),
                  cell_counts = as.numeric(cell_table))
mm <- match(ggdf$sample_id, md$sample_id)
ggdf$condition <- md$condition[mm]
cell_count <- ggplot(ggdf, aes(x= sample_id, y=cell_counts, fill= condition)) +
  geom_bar(stat = "identity") +
  geom_text(aes(label = cell_counts), hjust = 0.5, vjust = 0.5, size = 2.5) +
  theme_bw() +
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust = 1)) +
  scale_fill_manual(values= color_conditions, drop=FALSE) +
  scale_x_discrete(drop=FALSE)
tiff("Cell_count.tiff", width = 8, height = 8, units = "in", res = 300)
plot(cell_count)
dev.off()

#plot marker expression

ggdf <- data.frame(sample_id = sample_ids, expr)
ggdf <- melt(ggdf, id.var = "sample_id",
            value.name = "expression", variable.name = "antigen")
mm <- match(ggdf$sample_id, md$sample_id)

```

```

ggdf$condition <- md$condition[mm]

marker_expression <- ggplot(ggdf, aes(x = expression, color = condition,
                                     group = sample_id)) +
  geom_density() +
  facet_wrap (~antigen, nrow = 4, scales = "free") +
  theme_bw()
theme(axis.text.x = element_text(angle = 90, hjust = 1),
      strip.text = element_text(size=7), axis.text = element_text(size = 5)) +
  scale_color_manual(values = color_conditions)
tiff("Marker_expression.tiff", width = 14, height = 7, units = "in", res = 300)
plot(marker_expression)
dev.off()

#plot MDS

library(dplyr)

expr_median_sample_tbl <- data.frame(sample_id = sample_ids, expr) %>%
  group_by(sample_id) %>% summarise_all(funs(median))

expr_median_sample <- t(expr_median_sample_tbl[,-1])
colnames(expr_median_sample) <- expr_median_sample_tbl$sample_id

library(limma)
mds <- plotMDS(expr_median_sample, plot = FALSE)

library(ggrepel)
ggdf <- data.frame(MDS1 = mds$x, MDS2 = mds$y,
                  sample_id = colnames(expr_median_sample))
mm <- match(ggdf$sample_id, md$sample_id)
ggdf$condition <- md$condition[mm]
mds <- ggplot(ggdf, aes(x = MDS1, y = MDS2, color = condition)) +
  geom_point(size = 2, alpha = 0.8) +

```

```

geom_label_repel(aes(label = sample_id)) +
theme_bw() +
scale_color_manual(values = color_conditions) +
coord_fixed()
tiff("MDS.tiff", width = 8, height = 8, units = "in", res = 300)
plot(mds)
dev.off()

```

#plot heatmap of median marker expression for each sample

```

library(RColorBrewer)
library(pheatmap)
mm <- match(colnames(expr_median_sample), md$sample_id)
annotation_col <- data.frame(condition = md$condition[mm],
                             row.names = colnames(expr_median_sample))
annotation_colors <- list(condition = color_conditions)
color <- colorRampPalette(brewer.pal(n=9, name = "YlGnBu"))(100)
heatmap_samples <- pheatmap(expr_median_sample, color = color, display_numbers = TRUE,
                             number_color = "black", fontsize_number = 5, annotation_col = annotation_col,
                             annotation_colors = annotation_colors, clustering_method = "average")
save_pheatmap_png <- function(x, filename, width=1024, height=1024, res = 300) {
  stopifnot(!missing(x))
  stopifnot(!missing(filename))
  png(filename, width=width, height=height)
  grid::grid.newpage()
  grid::grid.draw(x$gtable)
  dev.off()
}
save_pheatmap_png(heatmap_samples, "Heatmap_samples_res.png")

```

#plot NRS of marker expression

```

NRS <- function(x, ncomp = 3) {

```

```

pr <- prcomp(x, center = TRUE, scale. = FALSE)
score <- rowSums(outer(rep(1, ncol(x)),
                      pr$sdev[1:ncomp]^2) * abs(pr$rotation[,1:ncomp]))
return(score)
}

nrs_sample <- fsApply(fcs[, lineage_markers], NRS, use.exprs = TRUE)
rownames(nrs_sample) <- md$sample_id
nrs <- colMeans(nrs_sample, na.rm = TRUE)

lineage_markers_ord <- names(sort(nrs, decreasing = TRUE))
nrs_sample <- data.frame(nrs_sample)
nrs_sample$sample_id <- rownames(nrs_sample)

ggdf <- melt(nrs_sample, id.var = "sample_id",
            value.name = "nrs", variable.name = "antigen")
ggdf$antigen <- factor(ggdf$antigen, levels = lineage_markers_ord)
mm <- match(ggdf$sample_id, md$sample_id)

ggdf$condition <- md$condition[mm]
nrs_plot <- ggplot(ggdf, aes(x = antigen, y = nrs)) +
  geom_point(aes(color = condition), alpha = 0.9,
            position = position_jitter(width = 0.3, height = 0)) +
  stat_summary(fun.y = "mean", geom = "point", shape = 21, fill = "white") +
  theme_bw() +
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust = 1)) +
  scale_color_manual(values = color_conditions)
tiff("NRS_plot.tiff", width = 4, height = 4, unit = "in", res = 300)
plot(nrs_plot)
dev.off()

#flowSOM clustering - increased SOM grid to 20 x 20

```

```

library(FlowSOM)
fsom <- ReadInput(fcs, transform = FALSE, scale = FALSE)
set.seed(123)
som <- BuildSOM(fsom, colsToUse = lineage_markers, xdim = 20, ydim = 20)
cell_clustering_som <- som$map$mapping[,1]

#Build MST from SOM grid

tsom_functional <- BuildMST(som, tSNE = TRUE)

#Plot MST with starplots

tiff("MST.tiff", width = 8, height = 8, unit = "in",res = 300)
PlotStars(tsom_functional)
dev.off()

#Colour MST based on marker expression

print(colnames(tsom_functional$map$medianValues))

tiff("CD19_MST.tiff", width = 8, height = 8, unit = "in",res = 300)
PlotMarker(tsom_functional, "CD19", main = "CD19")
dev.off()

tiff("CD3_MST.tiff", width = 8, height = 8, unit = "in",res = 300)
PlotMarker(tsom_functional, "CD3", main = "CD3")
dev.off()

tiff("CD4_MST.tiff", width = 8, height = 8, unit = "in",res = 300)
PlotMarker(tsom_functional, "CD4", main = "CD4")
dev.off()

```

```
tiff("CD14_MST.tiff", width = 8, height = 8, unit = "in",res = 300)
PlotMarker(tsom_functional, "CD14", main = "CD14")
dev.off()
```

```
tiff("CD8a_MST.tiff", width = 8, height = 8, unit = "in",res = 300)
PlotMarker(tsom_functional, "CD8a", main = "CD8a")
dev.off()
```

```
tiff("CD11b_MST.tiff", width = 8, height = 8, unit = "in",res = 300)
PlotMarker(tsom_functional, "CD11b", main = "CD11b")
dev.off()
```

```
tiff("CD11c_MST.tiff", width = 8, height = 8, unit = "in",res = 300)
PlotMarker(tsom_functional, "CD11c", main = "CD11c")
dev.off()
```

```
tiff("CD16_MST.tiff", width = 8, height = 8, unit = "in",res = 300)
PlotMarker(tsom_functional, "CD16", main = "CD16")
dev.off()
```

```
tiff("CD161_MST.tiff", width = 8, height = 8, unit = "in",res = 300)
PlotMarker(tsom_functional, "CD161", main = "CD161")
dev.off()
```

```
tiff("CD127_MST.tiff", width = 8, height = 8, unit = "in",res = 300)
PlotMarker(tsom_functional, "CD127", main = "CD127")
dev.off()
```

```
tiff("CD45RO_MST.tiff", width = 8, height = 8, unit = "in",res = 300)
PlotMarker(tsom_functional, "CD45RO", main = "CD45RO")
dev.off()
```

```
#Metaclustering
```

```

library(ConsensusClusterPlus)

codes <- som$map$codes
plot_outdir <- "consensus_plots"
nmc <- 40

mc <- ConsensusClusterPlus(t(codes), maxK = nmc, reps = 100,
  pltem = 0.9, pFeature = 1, title = plot_outdir, plot = "png",
  clusterAlg = "hc", innerLinkage = "average", finalLinkage = "average",
  distance = "euclidean", seed = 1234)

code_clustering <- mc[[nmc]]$consensusClass
cell_clustering <- code_clustering[cell_clustering_som]

#Plotting - plotting heatmap of median marker expression for metaclusters

color_clusters <-
c("#E6E6FA", "#483D8B", "#D8BFD8", "#00FFFF", "#00FF7F", "#800000", "#008000", "#000080", "#000000",
"#DC050C", "#FB8072", "#1965B0", "#7BAFDE", "#882E72", "#B17BA6", "#FF7F00", "#FDB462",
"#E7298A", "#E78AC3", "#33A02C", "#B2DF8A", "#55A1B1", "#8DD3C7", "#A6761D", "#E6AB02", "#7570B3",
"#BEAED4", "#666666", "#999999", "#aa8282", "#d4b7b7", "#8600bf", "#ba5ce3", "#808000", "#aeae5c",
"#1e90ff", "#00bfff", "#56ff0d", "#ffff00", "#FF7F50")

plot_clustering_heatmap_wrapper <- function(expr, expr01, cell_clustering, color_clusters, cluster_merging =
NULL){

#calculate median expression

expr_median <- data.frame(expr, cell_clustering = cell_clustering) %>%
  group_by(cell_clustering) %>% summarise_all(funs(median))

expr01_median <- data.frame(expr01, cell_clustering = cell_clustering) %>%
  group_by(cell_clustering) %>% summarise_all(funs(median))

#calculate cluster frequencies

clustering_table <- as.numeric(table(cell_clustering))
clustering_prop <- round(clustering_table / sum(clustering_table) * 100, 2)

d <- dist(expr_median[, colnames(expr)], method = "euclidean")

```

```

cluster_rows <- hclust(d, method = "average")

#sort cell clusters with hierarchical clustering
expr_heat <- as.matrix(expr01_median[, colnames(expr01)])
rownames (expr_heat) <- expr01_median$cell_clustering

#colours for the heatmap
color_heat <- colorRampPalette(rev(brewer.pal(n=9, name = "RdYlBu"))) (100)
legend_breaks <- seq(from = 0, to = 10, by = 2)
labels_row <- paste0(expr01_median$cell_clustering, " (", clustering_prop, "%)")

#annotation of original clusters
annotation_row <- data.frame(Cluster = factor(expr01_median$cell_clustering))
rownames(annotation_row) <- rownames(expr_heat)
color_clusters1 <- color_clusters[1:nlevels(annotation_row$Cluster)]
names (color_clusters1) <- levels(annotation_row$Cluster)
annotation_colors <- list(Cluster = color_clusters1)

#annotation for merged clusters
if(! is.null(cluster_merging)) {
  cluster_merging$new_cluster <- factor(cluster_merging$new_cluster)
  annotation_row$Cluster_merging <- cluster_merging$new_cluster
  color_clusters2 <- color_clusters[1:nlevels(cluster_merging$new_cluster)]
  names(color_clusters2) <- levels(cluster_merging$new_cluster)
  annotation_colors$Cluster_merging <- color_clusters2
}

pheatmap(expr_heat, color = color_heat, cluster_cols = FALSE,
  cluster_rows = cluster_rows, labels_row = labels_row,
  display_numbers = TRUE, number_color = "black",
  fontsize = 12, fontsize_number = 6, legend_breaks = legend_breaks,
  annotation_row = annotation_row, annotation_colors = annotation_colors)
}

```

```
tiff("Heatmap_cluster_40.tiff", width = 8, height = 11, unit = "in", res = 300)
plot_clustering_heatmap_wrapper(expr = expr[, lineage_markers_ord],
                                expr01 = expr01[, lineage_markers_ord],
                                cell_clustering = cell_clustering, color_clusters = color_clusters)
dev.off()
```

#wrapper function to plot distribution of markers intensities for flowSOM clusters

```
library(ggribes)
```

```
plot_clustering_distr_wrapper <- function(expr, cell_clustering){
  #calculate median expression
  cell_clustering <- factor(cell_clustering)
  expr_median <- data.frame(expr, cell_clustering = cell_clustering) %>%
    group_by(cell_clustering) %>% summarize_all(funs(median))

  #sort cell clustering with hierarchical clustering
  d <- dist(expr_median[, colnames(expr)], method = "euclidean")
  cluster_rows <- hclust(d, method="average")

  #calculate cluster frequencies
  freq_clust <- table(cell_clustering)
  freq_clust <- round(as.numeric(freq_clust)/sum(freq_clust)*100,2)
  cell_clustering <- factor(cell_clustering,
                            labels = paste0(levels(cell_clustering), "(", freq_clust, "%)"))

  ##data organised by cluster
  ggd <- melt(data.frame(cluster = cell_clustering, expr),
              id.vars = "cluster", value.name = "expression",
              variable.name = "antigen")
  ggd$antigen <- factor(ggd$antigen, levels = colnames(expr))
  ggd$reference <- "no"
  ## reference data
  ggd_bg <- ggd
```

```

ggd_bg$cluster <- "reference"
ggd_bg$reference <- "yes"

ggd_plot <- rbind(ggd,ggd_bg)
ggd_plot$cluster <- factor(ggd_plot$cluster,
                          levels = c(levels(cell_clustering)[rev(cluster_rows$order)], "reference"))

ggplot() +
  geom_density_ridges(data = ggd_plot, aes(x = expression, y=cluster,
                                           color = reference, full = reference), alpha = 0.3) +
  facet_wrap(~antigen, scales = "free_x", nrow = 2) +
  theme_ridges() +
  theme(axis.text = element_text(size = 7),
        strip.text = element_text(size = 7), legend.position = "none")
}

tiff("Marker_dist_cluster.tiff", width = 8, height = 8, unit = "in", res = 300)
plot_clustering_distr_wrapper(expr = expr[,lineage_markers_ord],
                              cell_clustering = cell_clustering)
dev.off()

#tSNE analysis ( analysis on 140,000 cells takes 40 minutes in R)

library(Rtsne)

dups <- which(!duplicated(expr[,lineage_markers]))
inds <- split(1:length(sample_ids), sample_ids)

tsne_ncells <- pmin(table(sample_ids), 10000) # downsampling
set.seed(1234)
tsne_inds <- lapply(names(inds), function(i) {
  s <- sample(inds[[i]], tsne_ncells[i], replace = FALSE)
  intersect(s,dups)
})

```

```

})
tsne_inds <- unlist(tsne_inds)
tsne_expr <- expr[tsne_inds, lineage_markers]

set.seed(1234)
tsne_out <- Rtsne(tsne_expr, check_duplicates = FALSE, pca = FALSE)

#color tsne by marker expression

dr <- data.frame(tSNE1 = tsne_out$Y[,1], tSNE2 = tsne_out$Y[,2],
                expr[tsne_inds,lineage_markers])

test <- ggplot (dr, aes(x=tSNE1, y=tSNE2, color = CD4)) +
  geom_point(size = 0.8) +
  theme_bw() +
  scale_color_gradientn("CD4", colors = colorRampPalette(rev(brewer.pal(n=11, name = "Spectral")))(50))
tiff("CD4_tSNE.tiff", width = 4, height = 4, unit = "in", res = 300)
plot(test)
dev.off()

test <- ggplot (dr, aes(x=tSNE1, y=tSNE2, color = CD14)) +
  geom_point(size = 0.8) +
  theme_bw() +
  scale_color_gradientn("CD14", colors = colorRampPalette(rev(brewer.pal(n=11, name = "Spectral")))(50))
tiff("CD14_tSNE.tiff", width = 4, height = 4, unit = "in", res = 300)
plot(test)
dev.off()

test <-ggplot (dr, aes(x=tSNE1, y=tSNE2, color = CD19)) +
  geom_point(size = 0.8) +
  theme_bw() +

```

```

scale_color_gradientn("CD19", colors = colorRampPalette(rev(brewer.pal(n=11, name = "Spectral")))(50))
tiff("CD19_tSNE.tiff", width = 4, height = 4, unit = "in", res = 300)
plot(test)
dev.off()

```

```

test <- ggplot (dr, aes(x=tSNE1, y=tSNE2, color = CD15)) +
  geom_point(size = 0.8) +
  theme_bw() +
  scale_color_gradientn("CD15", colors = colorRampPalette(rev(brewer.pal(n=11, name = "Spectral")))(50))
tiff("CD15_tSNE.tiff", width = 4, height = 4, unit = "in", res = 300)
plot(test)
dev.off()

```

```

test <- ggplot (dr, aes(x=tSNE1, y=tSNE2, color = CD8a)) +
  geom_point(size = 0.8) +
  theme_bw() +
  scale_color_gradientn("CD8a", colors = colorRampPalette(rev(brewer.pal(n=11, name = "Spectral")))(50))
tiff("CD8a_tSNE.tiff", width = 4, height = 4, unit = "in", res = 300)
plot(test)
dev.off()

```

```

test <- ggplot (dr, aes(x=tSNE1, y=tSNE2, color = CD56)) +
  geom_point(size = 0.8) +
  theme_bw() +
  scale_color_gradientn("CD56", colors = colorRampPalette(rev(brewer.pal(n=11, name = "Spectral")))(50))
tiff("CD56_tSNE.tiff", width = 4, height = 4, unit = "in", res = 300)
plot(test)
dev.off()

```

```

test <- ggplot (dr, aes(x=tSNE1, y=tSNE2, color = HLA_DR)) +
  geom_point(size = 0.8) +
  theme_bw() +
  scale_color_gradientn("HLA_DR", colors = colorRampPalette(rev(brewer.pal(n=11, name = "Spectral")))(50))
tiff("HLA_DR_tSNE.tiff", width = 4, height = 4, unit = "in", res = 300)

```

```
plot(test)
dev.off()
```

```
test <- ggplot (dr, aes(x=tSNE1, y=tSNE2, color = CD16)) +
  geom_point(size = 0.8) +
  theme_bw() +
  scale_color_gradientn("CD16", colors = colorRampPalette(rev(brewer.pal(n=11, name = "Spectral")))(50))
tiff("CD16_tSNE.tiff", width = 4, height = 4, unit = "in", res = 300)
```

```
plot(test)
dev.off()
```

```
test <- ggplot (dr, aes(x=tSNE1, y=tSNE2, color = CD11c)) +
  geom_point(size = 0.8) +
  theme_bw() +
  scale_color_gradientn("CD11c", colors = colorRampPalette(rev(brewer.pal(n=11, name = "Spectral")))(50))
tiff("CD11c_tSNE.tiff", width = 4, height = 4, unit = "in", res = 300)
```

```
plot(test)
dev.off()
```

```
test <- ggplot (dr, aes(x=tSNE1, y=tSNE2, color = CD45RO)) +
  geom_point(size = 0.8) +
  theme_bw() +
  scale_color_gradientn("CD45RO", colors = colorRampPalette(rev(brewer.pal(n=11, name = "Spectral")))(50))
tiff("CD45RO_tSNE.tiff", width = 4, height = 4, unit = "in", res = 300)
```

```
plot(test)
dev.off()
```

```
test <- ggplot (dr, aes(x=tSNE1, y=tSNE2, color = CD68)) +
  geom_point(size = 0.8) +
  theme_bw() +
  scale_color_gradientn("CD68", colors = colorRampPalette(rev(brewer.pal(n=11, name = "Spectral")))(50))
tiff("CD68_tSNE.tiff", width = 4, height = 4, unit = "in", res = 300)
```

```
plot(test)
dev.off()
```

```

test <- ggplot (dr, aes(x=tSNE1, y=tSNE2, color = CD11b)) +
  geom_point(size = 0.8) +
  theme_bw() +
  scale_color_gradientn("CD11b", colors = colorRampPalette(rev(brewer.pal(n=11, name = "Spectral")))(50))
tiff("CD11b_tSNE.tiff", width = 4, height = 4, unit = "in", res = 300)
plot(test)
dev.off()

```

```

test <- ggplot (dr, aes(x=tSNE1, y=tSNE2, color = CD161)) +
  geom_point(size = 0.8) +
  theme_bw() +
  scale_color_gradientn("CD161", colors = colorRampPalette(rev(brewer.pal(n=11, name = "Spectral")))(50))
tiff("CD161_tSNE.tiff", width = 4, height = 4, unit = "in", res = 300)
plot(test)
dev.off()

```

#plot tsne coloured by clusters

```

dr$sample_id <- sample_ids[tsne_inds]
mm <- match(dr$sample_id, md$sample_id)
dr$condition <- md$condition[mm]
dr$cell_clustering <- factor(cell_clustering[tsne_inds], levels = 1:nmc)

```

```

ggp <- ggplot(dr, aes(x = tSNE1, y = tSNE2, color = cell_clustering)) +
  geom_point(size = 0.8) +
  theme_bw() +
  scale_color_manual(values = color_clusters) +
  guides(color = guide_legend(override.aes = list(size=4), ncol = 2))

tiff("tSNE_cluster.tiff", width = 6, height = 6, unit = "in", res = 300)
plot(ggp)
dev.off()

```

```

#tsne for each sample

ggp_sample <- ggp + facet_wrap(~ sample_id)
tiff("tSNE_sample.tiff", width = 8, height = 8, unit = "in", res = 300)
plot(ggp_sample)
dev.off()

#tsne for each condition

ggp_condition <- ggp + facet_wrap(~ condition)
tiff("tSNE_condition.tiff", width = 8, height = 6, unit = "in", res = 300)
plot(ggp_condition)
dev.off()

## investigate metaclusters

code_sizes <- table(factor(som$map$mapping[,1], levels = 1:nrow(codes)))
code_sizes <- as.numeric(code_sizes)

#run tsne on cell codes

library(Rtsne)
set.seed(1234)
tsne_out <- Rtsne(codes, perplexity = 5, pca = TRUE)

#run pca on codes

pca_out <- prcomp(codes, center = TRUE, scale = FALSE)

codes_dr <- data.frame(tSNE1 = tsne_out$Y[,1], tSNE2 = tsne_out$Y[,2],
                      PCA1 = pca_out$x[,1], PCA2 = pca_out$x[,2])
codes_dr$code_clustering <- factor(code_clustering)
codes_dr$size <- code_sizes

```

```
#plot tsne based on codes
```

```
gg_tsne_codes <- ggplot(codes_dr, aes(x = tSNE1, y = tSNE2,  
                                     color = code_clustering, size = size)) +  
  geom_point(alpha = 0.9) +  
  theme_bw() +  
  scale_color_manual(values = color_clusters) +  
  guides(color = guide_legend(override.aes = list(size = 4), ncol = 2))  
tiff("tSNE_codes.tiff", width = 4, height = 4, unit = "in", res = 300)  
plot(gg_tsne_codes)  
dev.off()
```

```
#plot pca based on codes
```

```
gg_pca_codes <- ggplot(codes_dr, aes(x = PCA1, y = PCA2,  
                                     color = code_clustering, size = size )) +  
  geom_point(alpha = 0.9) +  
  theme_bw() +  
  scale_color_manual(values = color_clusters) +  
  guides (color = guide_legend(override.aes = list(size = 4), ncol = 2)) +  
  theme(legend.position = "right", legend.box = "vertical")  
tiff("PCA_codes.tiff", width = 4, height = 4, unit = "in", res = 300)  
plot(gg_pca_codes)  
dev.off()
```

```
#wrapper function to plot median marker expression heatmap for metaclusters along with functional markers
```

```
library(ComplexHeatmap)
```

```
plot_clustering_heatmap_wrapper2 <- function(expr, expr01,  
                                             lineage_markers, functional_markers = NULL, sample_ids = NULL,  
                                             cell_clustering, color_clusters, cluster_merging = NULL,  
                                             plot_cluster_annotation = TRUE) {
```

```

#calculate median expression of lineage markers
expr_median <- data.frame(expr[, lineage_markers],
  cell_clustering = cell_clustering) %>%
  group_by(cell_clustering) %>% summarize_all(funs(median))
expr01_median <- data.frame(expr01[,lineage_markers],
  cell_clustering = cell_clustering) %>%
  group_by(cell_clustering) %>% summarize_all(funs(median))

#calculate cluster frequencies
clustering_table <- as.numeric(table(cell_clustering))
clustering_prop <- round(clustering_table / sum(clustering_table) * 100, 2)

#sort the cell clusters with hierarchical clustering
d <- dist(expr_median[, lineage_markers], method = "euclidean")
cluster_rows <- hclust(d,method = "average")

expr_heat <- as.matrix(expr01_median[, lineage_markers])

#Median expression of functional markers in each sample per cluster
expr_median_sample_cluster_tbl <- data.frame(expr01[, functional_markers,
  drop = FALSE], sample_id = sample_ids, cluster = cell_clustering) %>%
  group_by(sample_id, cluster) %>% summarize_all(funs(median))

#Colors for the heatmap
color_heat <- colorRampPalette(rev(brewer.pal(n=9, name = "RdYlBu")))(100)
legend_breaks <- seq(from = 0, to = 1, by = 0.2)
labels_row <- paste0(expr01_median$cell_clustering, "(", clustering_prop, "%)")

##Annotation for the original clusters
annotation_row1 <- data.frame(Cluster = factor(expr01_median$cell_clustering))
color_clusters1 <- color_clusters[1:nlevels(annotation_row1$Cluster)]
names(color_clusters1) <- levels(annotation_row1$Cluster)

```

```

##Annotation of the merged clusters
if(!is.null(cluster_merging)){
  mm <- match(annotation_row1$Cluster, cluster_merging$original_cluster)
  annotation_row2 <- data.frame(Cluster_merging =
    factor(cluster_merging$new_cluster[mm]))
  color_clusters2 <- color_clusters[1:nlevels(annotation_row2$Cluster_merging)]
  names(color_clusters2) <- levels(annotation_row2$Cluster_merging)
}

## Heatmap for original clusters
ha1 <- Heatmap(annotation_row1, name = "Cluster",
  col = color_clusters1, cluster_columns = FALSE,
  cluster_rows = cluster_rows, row_dend_reorder = FALSE,
  show_row_names = FALSE, width = unit(0.5, "cm"),
  rect_gp = gpar(col = "grey"))

## Heatmap annotation for merged clusters
if(!is.null(cluster_merging)){
  ha2 <- Heatmap(annotation_row2, name = "Cluster \nmerging",
    col = color_clusters2, cluster_columns = FALSE,
    cluster_rows = cluster_rows, row_dend_reorder = FALSE,
    show_row_names <- FALSE, width = unit(0.5, "cm"),
    rect_gp = gpar(col = "grey"))
}

## Cluster names and sizes - text
ha_text <- rowAnnotation(text = row_anno_text(labels_row,
  gp = gpar(fontsize = 6)), width = max_text_width(labels_row))

###cluster sizes - bar plot
ha_bar <- rowAnnotation("Frequency (%)" = row_anno_barplot (
  x = clustering_prop, border = FALSE, axis = TRUE,
  axis_gp = gpar(fontsize = 5), gp = gpar(fill = "696969", col= "696969"),
  bar_width = 0.9), width = unit(0.7, "cm"), show_annotation_name = TRUE,
  annotation_name_rot = 0, annotation_name_offset = unit(5, "mm"),

```

```

annotation_name_gp = gpar(fontsize = 7))
## Heatmap for the lineage markers
ht1 <- Heatmap(expr_heat, name = "Expr", column_title = "Lineage markers",
  col = color_heat, cluster_columns = FALSE, cluster_rows = cluster_rows,
  row_dend_reorder = FALSE, heatmap_legend_param = list(at = legend_breaks,
    labels = legend_breaks, color_bar = "continuous"),
  show_row_names = FALSE, row_dend_width = unit(2, "cm"),
  rect_gp = gpar(col = "grey"), column_names_gp = gpar(fontsize = 8))

if(plot_cluster_annotation){
  draw_out <- ha1
}else{
  draw_out <- NULL
}
if(!is.null(cluster_merging)){
  draw_out <- draw_out + ha2 + ha_bar + ha_text
}else{
  draw_out <- draw_out + ht1 + ha_bar + ha_text
}

##heatmaps for the signalling markers
if(!is.null(functional_markers)){
  for(i in 1:length(functional_markers)){
    #re arrange so the row represent clusters
    expr_heat_fun <- as.matrix(dcast(expr_median_sample_cluster_tbl[,
      c("sample_id", "cluster", functional_markers[i])],
      cluster ~ sample_id, value.var = functional_markers[i]),-1))

    draw_out <- draw_out + Heatmap(expr_heat_fun,
      column_title = functional_markers[i], col = color_heat,
      cluster_columns = FALSE, cluster_rows = cluster_rows,
      row_dend_reorder = FALSE, show_heatmap_legend = FALSE,
      show_row_names = FALSE, rect_gp = gpar(col = "grey"),
      column_names_gp = gpar(fontsize = 8))
  }
}

```

```

}
}
draw(draw_out, row_dend_side = "left")
}

tiff("lineage_heatmap_functional_PD-1.tiff", width = 10, height = 8, unit = "in", res = 300)
plot_clustering_heatmap_wrapper2(expr = expr, expr01 = expr01,
    lineage_markers = lineage_markers, functional_markers = "PD-1",
    sample_ids = sample_ids, cell_clustering = cell_clustering,
    color_clusters = color_clusters, cluster_merging = NULL)
dev.off()

tiff("lineage_heatmap_functional_b.tiff", width = 10, height = 8, unit = "in", res = 300)
plot_clustering_heatmap_wrapper2(expr = expr, expr01 = expr01,
    lineage_markers = lineage_markers, functional_markers = "b",
    sample_ids = sample_ids, cell_clustering = cell_clustering,
    color_clusters = color_clusters, cluster_merging = NULL)
dev.off()

tiff("lineage_heatmap_functional_pdl1.tiff", width = 10, height = 8, unit = "in", res = 300)
plot_clustering_heatmap_wrapper2(expr = expr, expr01 = expr01,
    lineage_markers = lineage_markers, functional_markers = "pdl1",
    sample_ids = sample_ids, cell_clustering = cell_clustering,
    color_clusters = color_clusters, cluster_merging = NULL)
dev.off()

tiff("lineage_heatmap_functional_perforin.tiff", width = 10, height = 8, unit = "in", res = 300)
plot_clustering_heatmap_wrapper2(expr = expr, expr01 = expr01,
    lineage_markers = lineage_markers, functional_markers = "perforin",
    sample_ids = sample_ids, cell_clustering = cell_clustering,
    color_clusters = color_clusters, cluster_merging = NULL)
dev.off()

tiff("lineage_heatmap_functional_ctla4.tiff", width = 10, height = 8, unit = "in", res = 300)

```

```

plot_clustering_heatmap_wrapper2(expr = expr, expr01 = expr01,
    lineage_markers = lineage_markers, functional_markers = "ctla4",
    sample_ids = sample_ids, cell_clustering = cell_clustering,
    color_clusters = color_clusters, cluster_merging = NULL)
dev.off()

tiff("lineage_heatmap_functional_IL17a.tiff", width = 10, height = 8, unit = "in", res = 300)
plot_clustering_heatmap_wrapper2(expr = expr, expr01 = expr01,
    lineage_markers = lineage_markers, functional_markers = "IL_17A",
    sample_ids = sample_ids, cell_clustering = cell_clustering,
    color_clusters = color_clusters, cluster_merging = NULL)
dev.off()

tiff("lineage_heatmap_functional_IL10.tiff", width = 10, height = 8, unit = "in", res = 300)
plot_clustering_heatmap_wrapper2(expr = expr, expr01 = expr01,
    lineage_markers = lineage_markers, functional_markers = "IL_10",
    sample_ids = sample_ids, cell_clustering = cell_clustering,
    color_clusters = color_clusters, cluster_merging = NULL)
dev.off()

tiff("lineage_heatmap_functional_IL4.tiff", width = 10, height = 8, unit = "in", res = 300)
plot_clustering_heatmap_wrapper2(expr = expr, expr01 = expr01,
    lineage_markers = lineage_markers, functional_markers = "IL_4",
    sample_ids = sample_ids, cell_clustering = cell_clustering,
    color_clusters = color_clusters, cluster_merging = NULL)
dev.off()

tiff("lineage_heatmap_functional_IL8.tiff", width = 10, height = 8, unit = "in", res = 300)
plot_clustering_heatmap_wrapper2(expr = expr, expr01 = expr01,
    lineage_markers = lineage_markers, functional_markers = "IL_8",
    sample_ids = sample_ids, cell_clustering = cell_clustering,
    color_clusters = color_clusters, cluster_merging = NULL)
dev.off()

```

```
tiff("lineage_heatmap_functional_IL21.tiff", width = 10, height = 8, unit = "in", res = 300)
plot_clustering_heatmap_wrapper2(expr = expr, expr01 = expr01,
    lineage_markers = lineage_markers, functional_markers = "IL_21",
    sample_ids = sample_ids, cell_clustering = cell_clustering,
    color_clusters = color_clusters, cluster_merging = NULL)
dev.off()
```

```
tiff("lineage_heatmap_functional_IL2.tiff", width = 10, height = 8, unit = "in", res = 300)
plot_clustering_heatmap_wrapper2(expr = expr, expr01 = expr01,
    lineage_markers = lineage_markers, functional_markers = "IL_2",
    sample_ids = sample_ids, cell_clustering = cell_clustering,
    color_clusters = color_clusters, cluster_merging = NULL)
dev.off()
```

```
tiff("lineage_heatmap_functional_IL32.tiff", width = 10, height = 8, unit = "in", res = 300)
plot_clustering_heatmap_wrapper2(expr = expr, expr01 = expr01,
    lineage_markers = lineage_markers, functional_markers = "IL_32",
    sample_ids = sample_ids, cell_clustering = cell_clustering,
    color_clusters = color_clusters, cluster_merging = NULL)
dev.off()
```

```
tiff("lineage_heatmap_functional_Osteoactivin.tiff", width = 10, height = 8, unit = "in", res = 300)
plot_clustering_heatmap_wrapper2(expr = expr, expr01 = expr01,
    lineage_markers = lineage_markers, functional_markers = "Osteoactivin",
    sample_ids = sample_ids, cell_clustering = cell_clustering,
    color_clusters = color_clusters, cluster_merging = NULL)
dev.off()
```

#plot heatmap to visualise median marker expression for all clusters in SOM grid - not particularly helpful due to number of clusters in SOM

```
#plot_clustering_heatmap_wrapper2(expr = expr, expr01 = expr01,
# lineage_markers = lineage_markers, functional_markers = "GM_CSF",
```

```

# sample_ids = sample_ids, cell_clustering = cell_clustering_som,
# color_clusters = color_clusters, cluster_merging = NULL)

#merging of metaclusters

cluster_merging1_filename <- "cluster_merging1.xlsx"
cluster_merging1 <- read_excel(cluster_merging1_filename)
data.frame(cluster_merging1)

#convert to factor with clusters in desired order

levels_cluster_merged <- c("naïve CD4 CD27 CD38", "memory CD4 CD27", "naïve CD4", "CD3 NK cells",
"Neutrophils", "naïve T cell CD27 CD38", "T cells", "naïve T cell", "naïve CD8", "memory CD8", "memory CD8
PD-1", "cells", "MAIT cells", "CD8 NK cells", "NK cells", "basophils", "mast cells", "CD14 monocytes",
"eosinophils", "plasmacytoid dendritic cells", "dendritic cells", "B cells", "memory B cells")

cluster_merging1$new_cluster <- factor(cluster_merging1$new_cluster,
levels = levels_cluster_merged)

#New clustering
mm <- match(cell_clustering, cluster_merging1$original_cluster)
cell_clustering1 <- cluster_merging1$new_cluster[mm]

mm <- match(code_clustering, cluster_merging1$original_cluster)
code_clustering1 <- cluster_merging1$new_cluster[mm]

#Update tSNE with annotated cell populations
dr$cell_clustering1 <- cell_clustering1[tsne_inds]
tsne_cell_pop_cond <- ggplot(dr, aes(x = tSNE1, y = tSNE2, color = cell_clustering1)) +
  geom_point(size = 0.8) +
  theme_bw() +
  scale_color_manual(values = color_clusters) +
  guides(color = guide_legend(override.aes = list(size = 4)))
tiff("tSNE_cell_populations_cond.tiff", width = 10, height = 8, unit = "in", res = 300)
plot(tsne_cell_pop_cond)
dev.off()

```

```

#TSNE split by condition annotated with cell populations
ggp_condition_2 <- tsne_cell_pop_cond + facet_wrap(~ condition)
tiff("tSNE_condition_labels.tiff", width = 12, height = 4, unit = "in", res = 300)
plot(ggp_condition_2)
dev.off()

#plot heatmaps showing cluster merging

tiff("Heatmap_cluster_merging.tiff", width = 12, height = 8, unit = "in", res = 300)
plot_clustering_heatmap_wrapper(expr= expr[, lineage_markers_ord],
                                expr01 = expr01[, lineage_markers_ord], cell_clustering = cell_clustering,
                                color_clusters = color_clusters, cluster_merging = cluster_merging1)
dev.off()

#heatmap of annotated cell pop

tiff("Heatmap_annotated_cell_pops.tiff", width = 10, height = 8, unit = "in", res = 300)
plot_clustering_heatmap_wrapper(expr = expr[, lineage_markers_ord],
                                expr01 = expr01[, lineage_markers_ord], cell_clustering = cell_clustering1,
                                color_clusters = color_clusters)
dev.off()

#Differential analysis

library(lme4)
library(multcomp)
#Model formula
model.matrix(~condition, data = md)

#create contrasts
contrast_names <- c("earlyvshealthy")
k1 <- c(0,1)
K <- matrix(k1, nrow =1, byrow = TRUE, dimnames = list(contrast_names))

```

K

```
FDR_cutoff <- 0.05
```

```
#Differential cell population abundance
```

```
counts_table <- table(cell_clustering1, sample_ids)
```

```
props_table <- t(t(counts_table) / colSums(counts_table))*100
```

```
counts <- as.data.frame.matrix(counts_table)
```

```
props <- as.data.frame.matrix(props_table)
```

```
#plot relative abundance of cell populations in each sample
```

```
ggdf <- melt(data.frame(cluster = rownames(props), props),
```

```
          id.vars = "cluster", value.name = "proportion", variable.name = "sample_id")
```

```
ggdf$cluster <- factor(ggdf$cluster, level = levels_clusters_merged)
```

```
#add condition information
```

```
mm <- match(ggdf$sample_id, md$sample_id)
```

```
ggdf$condition <- factor(md$condition[mm])
```

```
proportion_PBMC <- ggplot(ggdf, aes(x=sample_id, y=proportion, fill=cluster)) +
```

```
  geom_bar(stat = "identity") +
```

```
  facet_wrap(~condition, scales = "free_x") +
```

```
  theme_bw() +
```

```
  theme(axis.text.x = element_text(angle = 90, hjust = 1)) +
```

```
  #scale_fill_manual(values = color_clusters)
```

```
  tiff("Proportion_PBMC.tiff", width = 10, height = 8, unit = "in", res = 300)
```

```
plot(proportion_PBMC)
```

```
dev.off()
```

```
#box plots relative abundance PBMC
```

```
ggdf$patient_id <- factor(md$patient_id[mm])
```

```
abundance_boxplot <- ggplot(ggdf) +
```

```

geom_boxplot(aes(x= condition, y = proportion, color = condition,
               fill = condition), position = position_dodge(), alpha = 0.5,
            outlier.color = NA) +
geom_point(aes(x = condition, y= proportion, color = condition
), alpha = 0.8, position = position_jitterdodge()) +
facet_wrap(~cluster, scales = "free", nrow =2) +
theme_bw() +
theme(axis.text.x = element_blank(), axis.ticks.x = element_blank(),
      axis.title.x = element_blank(), strip.text = element_text(size = 10)) +
scale_color_manual(values = color_conditions) +
scale_fill_manual(values = color_conditions) +
#scale_shape_manual(values = c(16,17,8,3,12,0,1,2,18,5,6))
tiff("Box_plot_abundance_increase_font.tiff", width = 20, height = 8, unit = "in", res = 300)
plot(abundance_boxplot)
dev.off()

#get median functional marker expression for each annotated cell population in each sample

expr_median_sample_cluster_tbl <- data.frame(expr[, functional_markers],
      sample_id = sample_ids, cluster = cell_clustering1) %>%
group_by (sample_id, cluster) %>%
summarize_all(funs(median))

expr_median_sample_cluster_melt <- melt(expr_median_sample_cluster_tbl,
      id.vars = c("sample_id", "cluster"), value.name = "median_expression",
      variable.name = "antigen")

#rearrange so that rows represent clusters and markers
expr_median_sample_cluster <- dcast(expr_median_sample_cluster_melt,
      cluster + antigen ~ sample_id, value.var = "median_expression")
rownames(expr_median_sample_cluster) <- paste0(expr_median_sample_cluster$cluster,
      "_", expr_median_sample_cluster$antigen)

#eliminate clusters with low frequency

```

```

clusters_keep <- names(which((rowSums(counts <1) == 0 )))
keep_lf <- expr_median_sample_cluster$cluster %in% clusters_keep
expr_median_sample_cluster <- expr_median_sample_cluster[keeplf, ]
##eliminate cases where there is 0 expression in all samples
keep0 <- rowSums(expr_median_sample_cluster[, md$sample_id]) > 0
expr_median_sample_cluster <- expr_median_sample_cluster[keep0, ]

#plot marker expressions for cell populations
ggdf <- expr_median_sample_cluster_melt[expr_median_sample_cluster_melt$cluster %in% clusters_keep, ]
mm <- match(ggdf$sample_id, md$sample_id)
ggdf$condition <- factor(md$condition[mm])
ggdf$patient_id <- factor(md$patient_id[mm])
func_marker_plot <- ggplot(ggdf) +
  geom_boxplot(aes(x= antigen, y = median_expression,
    color = condition, fill = condition),
    position = position_dodge(), alpha = 0.5, outlier.color = NA) +
  geom_point(aes(x=antigen, y=median_expression, color = condition
), alpha = 0.8, position = position_jitterdodge(),
size = 0.7) +
  facet_wrap(~cluster, scales = "free_y", ncol = 2) +
  theme_bw() +
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust =1, size = 12)) +
  theme(axis.title.x = element_blank(), strip.text = element_text(size = 12)) +
  scale_color_manual(values = color_conditions) +
  scale_fill_manual(values = color_conditions) +
  #scale_shape_manual(values = c(16,17,8,3,12,0,1,2,18,5,6)) +
  #guides(shape = guide_legend(override.aes = list(size = 2)))
  tiff("Functional_markers_cell_pop_increase_font.tiff", width = 12, height = 18, unit = "in", res = 300)
plot(func_marker_plot)
dev.off()
write out csv file that can be used to view clusters and tsne in FlowJo

mm <- match(dr$sample_id, md$sample_id)
mm

```

```

dr$patient_id <- md$patient_id[mm]
write.csv(dr, "seed60_HCvEarlyRA_cytnorm_analysis.csv")

#attempt differential analysis significance testing

library(lme4)
library(multcomp)

#create model matrix to include condition and batch information

#model.matrix( ~condition + batch_id, data = md)

#create contrast

contrast_names <- c("earlyvshealthy")
contrast_names
k1 <- c(0,1)
k1
K <- matrix(k1,nrow = 1, byrow = TRUE, dimnames = list(contrast_names))
K
FDR_cutoff <- 0.05

#define generalized mixed model accounting for sample ID or sample ID and batch ID

formula_glmer_binomial1 <- y/total ~ condition + (1|sample_id)
#formula_glmer_binomial2 <- y/total ~condition + (1|sample_id) + (1|batch_id)
#formula_glmer_binomial2 <- y/total ~condition

#plot wrapper function to input data frame of cell counts and perform differential analysis specified in contrast
matrix K for each population seperately

differential_abundance_wrapper <- function(counts,md, formula, K) {
  #Fit GLM for each cluster seperately
  ntot <- colSums(counts)

```

```

fit_binomial <- lapply(1:nrow(counts), function(i){

  data_tmp <- data.frame(y = as.numeric(counts[i, md$sample_id]),
                        total = ntot[md$sample_id], md)

  fit_tmp <- glmer(formula, weights = total, family = binomial, data = data_tmp)

  ##Fit contrasts one by one
  out <- apply(K, 1, function(k){
    contr_tmp <- glht(fit_tmp, linfct = matrix(k,1))
    summ_tmp <- summary(contr_tmp)
    pval <- summ_tmp$test$pvalues
    return(pval)
  })
  return(out)
})

pvals <- do.call(rbind,fit_binomial)
colnames(pvals) <- paste0("pval_", contrast_names)
rownames(pvals) <- rownames(counts)

#Adjust the p values
adjp <- apply(pvals, 2, p.adjust, method = "BH")
colnames(adjp) <- paste0("adjp_", contrast_names)
return(list(pvals = pvals, adjp = adjp))
}

differential_abundance_wrapper <- function(counts, md, formula, K){
  ## Fit the GLMM for each cluster separately
  ntot <- colSums(counts)
  fit_binomial <- lapply(1:nrow(counts), function(i){

    data_tmp <- data.frame(y = as.numeric(counts[i, md$sample_id]),
                          total = ntot[md$sample_id], md)

    fit_tmp <- glmer(formula, weights = total, family = binomial,

```

```

data = data_tmp)

## Fit contrasts one by one
out <- apply(K, 1, function(k){
  contr_tmp <- glht(fit_tmp, linfct = matrix(k, 1))
  summ_tmp <- summary(contr_tmp)
  pval <- summ_tmp$test$pvalues
  return(pval)
})
return(out)
})
pvals <- do.call(rbind, fit_binomial)
colnames(pvals) <- paste0("pval_", contrast_names)
rownames(pvals) <- rownames(counts)
## Adjust the p-values

adjp <- apply(pvals, 2, p.adjust, method = "BH")
colnames(adjp) <- paste0("adjp_", contrast_names)
return(list(pvals = pvals, adjp = adjp))
}

da_out1_test <- differential_abundance_wrapper(counts, md = md, formula = formula_glmr_binomial1, K =
K)

apply(da_out1_test$adjp < FDR_cutoff, 2, table)

da_out1_test
##result was FALSE = 18 TRUE = 3

# repeat using contrast matrix accounting for batch
## got same result as previous - good sign as data has been batch normalised

#da_out2 <- differential_abundance_wrapper(counts, md = md, formula = formula_glmr_binomial2, K = K)

```

```

#apply(da_out2$adjp < FDR_cutoff,2, table)

#da_out2

da_output1 <- data.frame(cluster = rownames(props), props, da_out1_test$pvals, da_out1_test$adjp,
row.names = NULL)
print(head(da_output1), digits = 2)

#Plot results of differential abundance analysis as heatmap

#Define normalisation wrapper

normalisation_wrapper <- function(expr, th = 2.5) {
  expr_norm <- apply(expr,1, function(x){
    sdx <- sd(x, na.rm = TRUE)
    if(sdx ==0) {
      x <- (x - mean(x,na.rm = TRUE))
    } else {
      x <- (x - mean(x, na.rm = TRUE)) / sdx
    }
    x[x > th] <- th
    x[x < - th] <- -th
    return(x)
  })
  expr_norm <- t(expr_norm)
}

# wrapper to plot heatmap

plot_differential_heatmap_wrapper <- function(expr_norm, sign_adjp, condition, color_conditions, th = 2.5){
  #order samples by condition
  oo <- order(condition)
  condition <- condition[oo]
  expr_norm <- expr_norm[, oo, drop = FALSE]
}

```

```

#create row labels with adjp values and other objects
labels_row <- paste0(rownames(expr_norm), "(",
  sprintf("%.02e", sign_adjp), ")")
labels_col <- colnames(expr_norm)
annotation_col <- data.frame(condition = factor(condition))
rownames(annotation_col) <- colnames(expr_norm)
annotation_colors <- list(condition = color_conditions)
color <- colorRampPalette(c("#87CEFA", "#56B4E9", "#56B4E9", "#0072B2",
  "#000000", "#D55E00", "#E69F00", "#E69F00", "#FFD700"))(100)
breaks = seq(from = -th, to = th, length.out = 101)
legend_breaks = seq(from = -round(th), to = round(th), by = 1)
gaps_col <- as.numeric(table(annotation_col$condition))

pheatmap(expr_norm, color = color, breaks = breaks,
  legend_breaks = legend_breaks, cluster_cols = FALSE, cluster_rows = FALSE,
  labels_col = labels_col, labels_row = labels_row, gaps_col = gaps_col, annotation_col = annotation_col,
  annotation_colors = annotation_colors, fontsize = 8)
}

##apply the arcsine squareroot transformation to the proportions

asin_table <- asin(sqrt((t(t(counts_table)/ colSums(counts_table))))))
asin <- as.data.frame.matrix(asin_table)
##get significant clusters and sort them by significance
sign_clusters <- names(which(sort(da_out1_test$adjp,"adjp_earlyvshealthy") < FDR_cutoff))
## get the p values for significant clusters
sign_adjp <- da_out1_test$adjp[sign_clusters, "adjp_earlyvshealthy", drop = FALSE]
##normalise the transformed proportions to mean and sd +/- 1
asin_norm <- normalisation_wrapper(asin[sign_clusters,])

mm <- match(colnames(asin_norm), md$sample_id)

#plot differential abundance heatmap showing clusters with significantly different abundances in earlyRA or
HC

```

```
tiff("Heatmap_differential_abundance_new_labels.tiff", width = 14, height = 8, unit = "in", res = 300)

plot_differential_heatmap_wrapper(expr_norm = asin_norm, sign_adj = sign_adj, condition =
md$condition[mm], color_conditions = color_conditions)

dev.off()
```

```
#Differential expression testing of functional markers
```

```
differential_expression_wrapper <- function(expr_median, md, model = "lmer", formula, K){
  ## Fit LMM or LM for each marker separately
  fit_gaussian <- lapply(1:nrow(expr_median), function(i){
    data_tmp <- data.frame(y = as.numeric(expr_median[i, md$sample_id]), md)
    switch(model,
      lmer = {
        fit_tmp <- lmer(formula, data = data_tmp)
      },
      lm = {
        fit_tmp <- lm(formula, data = data_tmp)
      })
    ## Fit contrasts one by one
    out <- apply(K, 1, function(k){
      contr_tmp <- glht(fit_tmp, linfct = matrix(k, 1))
      summ_tmp <- summary(contr_tmp)
      pval <- summ_tmp$test$pvalues
      return(pval)
    })
    return(out)
  })
  pvals <- do.call(rbind, fit_gaussian)
  colnames(pvals) <- paste0("pval_", contrast_names)
  rownames(pvals) <- rownames(expr_median)
  ## Adjust the p-values
  adjp <- apply(pvals, 2, p.adjust, method = "BH")
  colnames(adjp) <- paste0("adjp_", contrast_names)
```

```

    return(list(pvals = pvals, adjp = adjp))
}

#create formula

formula_lm <- y ~ condition
formula_lmer <- y ~ condition + batch_id

de_out1 <- differential_expression_wrapper(expr_median = expr_median_sample_cluster,
                                         md = md, model = "lm", formula = formula_lm, K = K)

## Repeat differential expression using sampleID as a random variable

de_out2 <- differential_expression_wrapper(expr_median = expr_median_sample_cluster,
                                         md = md, model = "lmer", formula = formula_lmer, K = K)

#try to include batch information

de_out3 <- differential_expression_wrapper(expr_median = expr_median_sample_cluster,
                                         md = md, model = "lm", formula = formula_lmer, K = K)

apply(de_out1$adjp < FDR_cutoff, 2, table)

de_output <- data.frame(expr_median_sample_cluster, de_out1$pvals, de_out1$adjp, row.names = NULL)
print(head(de_output), digits = 2)

##Keep the significant markers and group by cluster

sign_clusters_markers <- names(which(de_out1$adjp[, "adjp_earlyRAvshealthy"] < FDR_cutoff))
oo <- order(expr_median_sample_cluster[sign_clusters_markers, "cluster"],
            de_out1$adjp[sign_clusters_markers, "adjp_earlyRAvshealthy"])

```

```
sign_clusters_markers <- sign_clusters_markers[oo]

#get the significant adjust P values

sign_adj <- de_out1$adjp[sign_clusters_markers, "adjp_earlyRAvshealthy"]

##normalise to mean 0 sd = 1

expr_s <- expr_median_sample_cluster[sign_clusters_markers, md$sample_id]
expr_median_sample_cluster_norm <- normalisation_wrapper(expr_s)

mm <- match(colnames(expr_median_sample_cluster_norm), md$sample_id)

tiff("differential_expression_heatmap.tiff", width = 14, height = 8, unit = "in", res = 300)
plot_differential_heatmap_wrapper(expr_norm = expr_median_sample_cluster_norm, sign_adj = sign_adj,
condition = md$condition[mm], color_conditions = color_conditions)
dev.off()
```

## Appendix C

### Patient cohort and demographic details

#### Early RA

Patient Study Number	CRP	ESR	ACPA	Rheumatoid Factor	Swollen Joint Count 28	Tender Joint Count 28	VAS Global Health	Smoking status	Number smoked per day
3486	NULL	NULL	Positive	Negative	1	2	20	Current	15
3549	NULL	NULL	Positive	Positive	9	25	85	Current	15
3611	NULL	NULL	Positive	Negative	1	2	30	Previous	15
3709	15.5	NULL	Positive	Negative	0	0	5	Previous	7
3939	NULL	NULL	NULL	NULL	NULL	NULL	40	Current	10
3967	NULL	NULL	NULL	NULL	1	7	60	Never	NULL
3968	NULL	NULL	NULL	NULL	0	0	5	Never	NULL
4043	NULL	NULL	NULL	NULL	NULL	NULL	30	Previous	NULL
3863	NULL	Not done	Positive	Positive	1	0	NULL	NULL	NULL
4123	<5	NULL	Positive	Positive	2	1	10	Never	NULL
4154	NULL	NULL	NULL	NULL	10	21	85	Current	15
5015	NULL	NULL	NULL	NULL	2	5	100	Never	NULL
4153	NULL	NULL	Negative	Negative	6	13	85	Never	NULL
4195	NULL	NULL	NULL	NULL	24	28	80	Previous	20
4202	11.3	NULL	NULL	NULL	NULL	NULL	10	Never	NULL
4196	36	NULL	NULL	NULL	18	20	30	Current	3
4217	NULL	NULL	Negative	Negative	8	10	7	Current	15
4209	NULL	NULL	NULL	NULL	14	9	30	Never	NULL
4212	NULL	NULL	NULL	NULL	13	3	20	Previous	40
4219	NULL	NULL	NULL	NULL	6	0	5	Previous	10
4226	NULL	NULL	NULL	NULL	1	12	75	Previous	1
4225	121	NULL	Negative	Positive	10	20	90	Never	NULL
4233	1	NULL	Positive	Positive	5	6	60	Previous	20
4227	NULL	NULL	NULL	NULL	3	2	50	Never	NULL
4229	NULL	NULL	NULL	NULL	16	0	20	Never	NULL
4228	NULL	NULL	NULL	NULL	2	3	10	Current	14
4230	55	Not done	Positive	Positive	2	16	80	Previous	20
4247	NULL	NULL	NULL	NULL	12	22	90	Previous	20
4302	5.5	27	Positive	Positive	11	21	80	Previous	15
4331	5	NULL	Positive	Positive	1	0	20	Never	NULL
4333	42	NULL	Negative	Negative	19	27	70	Never	NULL
4334	10.2	30	Positive	Positive	6	22	70	Previous	15

## Established RA

missing data denoted by dash (-)

Sample ID	DAS CRP	DAS ESR	TJC	SJC	VAS	CRP	ESR	Serology results	Erosions at baseline	Previous targeted therapies
RADAR BMC 5013	-	-	-	-	-	-	-	-	-	-
RADAR BMC 1808	-	-	-	-	-	-	-	-	-	-
RADAR BMC 246	-	-	-	-	-	-	-	-	-	-
RADAR BMC 5014	6.77	7.23	13	8	80	67	67	ccp+ RF+	y	-
RADAR BMC 4078	-	-	-	-	-	-	-	-	-	-
RADAR BMC 805	6.49	6.67	16	8	90	37	33	CCP+ RF+	y	5 (ADA, RTX, TCZ,CRT, ABT)
RADAR BMC 074	5.85	5.55	13	13	60	16.2	11	CCP+ RF+	y	-
RADAR BMC 5016	4.7	5.05	5	3	65	20	31	CCP+ RF+	n	1 (ADA)
RADAR BMC 1044	-	-	-	-	-	-	-	-	-	-
RADAR BMC 3538	-	-	-	-	-	-	-	-	-	-
RADAR BMC 1996	-	-	-	-	-	-	-	-	-	-
RADAR BMC 4039	5.08	5.31	16	1	55	9.1	18	CCP-RF+	n	1 (ETN)
RADAR BMC 3051	-	-	-	-	-	-	-	-	-	-
RADAR BMC 345	-	-	-	-	-	-	-	-	-	-
RADAR BMC 3024	-	-	-	-	-	-	-	-	-	-
RADAR BMC 4100	5.77	6.02	15	4	78	14.4	23	CCP+ RF+	n	4 (ETN, ADA, TCZ, ABT)
RADAR BMC 3948	5.96	6.4	14	7	80	17	33	CCP+ RF+	y	7 (ADA, ETN, RTX, ABT, CTZ, IFX, TOC)
RADAR BMC 5017	5.06	5.99	11	6	70	<5	34	CCP+ RF+ DsDNA+	n	-
RADAR BMC 4182	-	-	-	-	-	-	-	-	-	-
RADAR BMC 3208	5.63	5.9	14	5	80	8.9	19	CCP-RF-	n	1 (ADA)
RADAR BMC 5009	7.3	7.58	19	13	90	92	60	CCP+ RF+		1 (Bari)
RADAR BMC 5007	-	-	-	-	-	-	-	-	-	-

RADAR BMC 4185	-	-	-	-	-	-	-	-	-	-	-
RADAR BMC 3351	3.91	4.29	4	4	20	14.6	28	CCP+ RF+	y	1 (ADA)	
RADAR BMC 5020	6.13	6.57	16	8	83	13.9	30	CCP+ RF+	n	-	
RADAR BMC 4100	5.77	6.02	15	4	78	14.4	23	CCP+RF+	n	4 (ETN, ADA, TCZ, ABT)	
RADAR BMC 2430	6.1	6.51	18	11	90	<5	16	CCP+RF-	y	3(RTX, ETN, ABT)	
RADAR BMC 3593	4.46		2	3	60	45		CCP+RF+	y	2( ADA, GOL)	
RADAR BMC 2415	5.39	4.62	16	5	70	<5	3	CCP+RF-	n	2 (IFX, GOL)	
RADAR BMC 2914	5.85	6.95	22	6	50	10.5	66	CCP+RF+Ro+	y	3 (RTX, ABT, TCZ)	
RADAR BMC 926	5.81	6.13	6	11	80	52	48	CCP+RF-	y	3 (ETN, TCZ, RTX)	
RADAR BMC 407	3.72	2.6	2	3	60	<5	2	CCP+RF+	n	3 (RXT, ETN, TCZ)	
RADAR BMC 0072 (in this document defined as 721)	5.28	6.05	20	2	60	<5	27	CCP+RF+	n	1 (IFX)	
RADAR BMC 3896	5.31	3.77	15	3	80	<5	1	CCP-RF-	y	1 (TCZ)	
RADAR BMC 770	2.93	2.52	3	1	10	<5	5	CCP+RF-	y	1 (ADA)	
RADAR BMC 5021	4.11	4.27	0	14	60	32	30	CCP- RF+	n	2 (ADA, GOL)	
RADAR BMC 4079	5.39	6.15	12	5	70	10.5	41	CCP+ RF +	y	5 ( ABT, TCZ, RTX, ETN, TOFA)	
RADAR BMC 2888	5.67	5.1	12	8	100	<5	4	CCP+ RF +	y	3 (RTX, TCZ, ABT)	
RADAR BMC 2576	-	-	-	-	-	-	-	-	-	-	
RADAR BMC 2760	5.98	6.16	8	3	90	110	57	CCP- RF-	y	4 (ADA, ABT, TCZ, IFX)	
RADAR BMC 0907	2.31	3.03	0	0	20	18.5	51	CCP+ RF +	n	-	
RADAR BMC 074	1.61	1.05	0	0	5	<5	3	CCP+ RF+	y	-	
RADAR BMC 4237	-	-	-	-	-	-	-	-	-	-	
RADAR BMC 2624	3.7	4.94	6	0	80	<5	35	CCP+ RF +	n	1 (CTZ, ETN)	
RADAR BMC 4238	-	-	-	-	-	-	-	-	-	-	
RADAR BMC 3562	4.5	4.85	10	5	40	<5	15	CCP- RF-	n	-	

RADAR BMC 3407	5.58	5.08	10	6	90	11.3	7	CCP+ RF+	n	3 ADA, TCZ, RTX
RADAR BMC 805	6.51	6.41	13	9	99	37	22	CCP- RF-	y	5 (ADA, RTX,TCZ, CTZ, ABT)
RADAR BMC 5009	8.02	8.46	27	16	90	134	93	CCP+RF+	n	1 (Bari)
RADAR BMC 820	5.18	5.38	7	4	70	27	29	CCP-RF-	n	4 (ADA, RTX, TCZ, CTZ)
RADAR BMC 5014	3.08	3.56	1	1	50	<5	18	CCP+ RF+	y	1 ETN
RADAR BMC 2047	7.67	8.37	26	19	80	67	93	CCP+RF+	n	2 (IFX, RTX)
RADAR BMC 1963	6.29	6.78	21	5	80	16.1	34	CCP+ RF-	n	6 (IFX, ADA, ETN, RTX, TCZ, ABT)
RADAR BMC 5016	4.26	4.61	5	5	60	<5	15	CCP+ RF+	n	1 (ADA)
RADAR BMC 3105	4.4	4.31	5	5	70	<5	8	CCP+ RF+	n	4 (ETN, ADA,RTX, ABT)
RADAR BMC 1158	4.15	2.61	10	4	20	<5	1	CCP-RF+	y	3 (ETN, RTX, TCZ)
RADAR BMC 5023	-	-	-	-	-	-	-	-	-	-
RADAR BMC 5023	-	-	-	-	-	-	-	-	-	-
RADAR BMC 4171	-	-	-	-	-	-	-	-	-	-
RADAR BMC 453	5.01	4.44	12	6	60	<5	4	CCP-RF-	n	1 (TCZ)
RADAR BMC 4039	2.78	2.73	9	0	40	<5	2	CCP-RF+	n	1 (ETN)
RADAR BMC 4285	-	-	-	-	-	-	-	-	-	-
RADAR BMC 4285	-	-	-	-	-	-	-	-	-	-
RADAR BMC 3208	3.44	2.39	7	0	30	<5	2	CCP-RF-	n	1 (ADA)
RADAR BMC 4033	4.33	4.33	5	5	65	<5	9	CCP+RF+ DsDNA+	y	2 (ADA, RTX)
RADAR BMC 579	6.6	7.21	25	7	75	17.3	42	CCP+RF+Ro/La+	y	2 (ADA, RTX)
RADAR BMC 2518	4.72	3.18	11	3	60	<5	1	CCP+RF-	y	3 (ADA, RTX, TCZ)
RADAR BMC 2241	6.15	6.53	16	13	80	8.8	22	CCP+RF-	y	7 (IFX, ETN, ANA, TCZ, RYX, ADA, ABT)
RADAR BMC 3588	6.39	6.47	14	9	75	55	35	CCP+RF+	y	6 (ETN enb, ABT,TCZ,RTX, TOFA, ETN ber)
RADAR BMC 407	3.28	3.58	1	6	35	<5	14	CCP+RF+	n	3 (RXT, ETN, TCZ)
RADAR BMC 4288	-	-	-	-	-	-	-	-	-	-

RADAR BMC 4288	-	-	-	-	-	-	-	-	-	-	-
RADAR BMC 3046	5.81	4.76	16	6	96	<5	2	CCP+RF+	y	3 (IFX, RTX, TCZ)	
RADAR BMC 4086	5.9	6.04	20	7	80	<5	11	CCP+RF+	n	-	
RADAR BMC 5020	2.21	2.41	0	0	2	29	30	CCP+ RF+	n	-	
RADAR BMC 2415	3.4	3.29	1	1	60	7.3	10	CCP+RF-	n	2 (IFX, GOL)	
RADAR BMC 2878	6.05		17	10	40	40		CCP- RF+	n	3 (RTX,TCZ,ABT)	
RADAR BMC 3449	4.11	4.1	4	2	75	<5	9	CCP+ RF+	y	2 (IFX, ADA)	
RADAR BMC 3948	5.96	6.4	14	7	80	17	33	CCP+ RF+	y	7 (ADA, ETN, RTX,ABT,CTZ,IFX,Tofa)	
RADAR BMC 1484	-	-	-	-	-	-	-	-	-	-	
RADAR BMC 5022	6.23	6.89	10	10	80	63	85	CCP+ RF+	n	3 (ETN, ADA,RTX)	
RADAR BMC 1746	4.65		18	0	76	<5		CCP-RF-	n	1 ETN	
RADAR BMC 3449	4.11	4.1	4	2	75	<5	9	CCP+RF+	n	2 (IFX, ADA)	
RADAR BMC 3449	-	-	-	-	-	-	-	-	-	-	
RADAR BMC 4100	4.85	5.29	12	1	67	6.7	21	CCP+RF+	n	4 (ETN, ADA, TCZ, ABT)	
RADAR BMC 770	2.47	2.67	2	0	10	<5	12	CCP+RF-	y	1 (ADA)	
RADAR BMC 1431	5.66	6.03	12	6	75	16	29	CCP+RF+	n	1 (TCZ)	
RADAR BMC 4310	-	-	-	-	-	-	-	-	-	-	
RADAR BMC 926	1.68	2.08	0	0	10	<5	16	CCP+RF-	y	3 (ETN, TCZ, RTX)	
RADAR BMC 2914	4.5	6.05	4	5	60	6.9	104	CCP+RF+Ro+	y	3 (RTX, ABT, TCZ)	
RADAR BMC 180	4.01	4.08	10	4	10	<5	10	CCP+ RF + ANA+	n	1 (IFX)	
RADAR BMC 907	3.77	4.6	5	3	12	12.3	49	CCP+ RF +	n	-	
RADAR BMC 3593	3.15	3.8	2	2	30	<5	23	CCP+RF+	y	2( ADA, GOL)	
RADAR BMC 1020	5.06	5.2	9	11	65	<5	11	CCP+RF+	y	5 (ETN, RTX, ADA,TCZ,ABT)	
RADAR BMC 3887	6.01	5.92	23	2	99	<5	8	CCP+RF-	n	-	
RADAR BMC 074	1.57	1.15	0	0	2	<5	5	CCP+ RF+	y	-	

RADAR BMC 4312	4.96	5.78	16	4	68	<5	18	CCP+RF+	n	-
RADAR BMC 2514	2.65	2.71	3	0	20	<5	8	CCP+RF+	y	1 (ETN)
RADAR BMC 3565	3.56	4.1	5	0	10	28	48	CCP+RF+Sm/RNP+	y	-
RADAR BMC 1740	4.5	3.99	4	1	70	24	10	CCP+RF-	y	4 (ADA, RTX, TCZ,ABT)
RADAR BMC 874	4.49	3.92	11	2	50	<5	4	CCP- RF+	n	1 (TCZ)
RADAR BMC 072 (in this document defined as 722)	5.22	5.47	12	6	75	<5	13	CCP+RF+	n	2 (IFX, ADA)
RADAR BMC 3896	3.49	3.56	5	0	50	<5	10	CCP-RF-	y	1 (TCZ)
RADAR BMC 2742	-	-	-	-	-	-	-	-	-	-
RADAR BMC 5016	4.26	4.17	7	2	60	<5	8	CCP+ RF+	n	1 (ADA)
RADAR BMC 2430	4.45	3.88	8	5	50	<5	4	CCP+RF-	y	3(RTX, ETN, ABT)
RADAR BMC 4039	3.03	2.59	4	1	30	<5	3	CCP-RF+	n	1 (ETN)
RADAR BMC 5021	2.32	1.17	0	4	9	<5	2	CCP- RF+	n	2 (ADA, GOL)
RADAR BMC 4013	4.03	4.65	4	1	50	13.9	38	CCP-RF-	y	1 (ADA)
RADAR BMC 3329	-	-	-	-	-	-	-	-	-	-
RADAR BMC 3982	4.74	4.74	10	2	74	<5	9	CCP- RF+	n	2 (ETN, RTX)
RADAR BMC 2067	6.53	5.82	26	5	70	21	7	CCP- RF-	n	2 (IFX, RTX)
RADAR BMC 3407	3.94	4.31	5	2	60	9.9	8	CCP+RF+	y	3 (ADA,TCZ,RTX)
RADAR BMC 3351	2.35	3.34	2	1	5	<5	23	CCP+ RF+	y	1 (ADA)
RADAR BMC 3208	4.86	4.29	12	2	70	<5	4	CCP-RF-	n	1 (ADA)
RADAR BMC 407	2.87	3.01	1	4	15	<5	11	CCP+RF+	n	3 (RXT, ETN, TCZ)
RADAR BMC 4045	6.72	7.18	28	2	90	23	39	CCP+RF-	n	-
RADAR BMC 424	5.58	4.81	19	6	65	<5	3	CCP+RF+	n	4 (ETN, RTX,ABT,TCZ)
RADAR BMC 5017	2.31	3.43	1	2	10	<5	28	CCP+ RF+ DsDNA+	n	-
RADAR BMC 2415	5.01	4.02	8	2	56	35	6	CCP+RF-	n	2 (IFX, GOL)

RADAR BMC 2047	5.19	5.02	16	6	52	<5	7	CCP+RF+	n	2 (IFX, RTX)
RADAR BMC 2711	5.39	5.62	13	4	65	12.6	21	CCP+RF+	y	7 (ETN, ADA, RTX, TCZ, ABT, CTZ, TOFA)
RADAR BMC 1826	-	-	-	-	-	-	-	-	-	-
RADAR BMC 820	4.01	4.32	2	1	75	12.2	23	CCP-RF-	n	4 (ADA, RTX, TCZ, CTZ)
RADAR BMC 50	4.99	3.43	10	3	85	<5	1	CCP+RF-	y	3 (ETN, RTX, TCZ)
RADAR BMC 5020	3.35	3.55	3	0	60	<5	2	CCP+ RF+	n	-

## Appendix D

### Barcoded batches 1 -10

#### Barcode batch 1

<u>Sample Identifier</u>	<u>Barcode</u>
RADAR BMC 5014 27.2.2018	1
RADAR BMC 5014 10.7.2018	2
RADAR BMC 770 5.6.2018	3
RADAR BMC 770 29.8.2018	4
RADAR BMC 3948 17.3.2018	5
RADAR BMC 3948 16.8.2018	6
RADAR BMC 805 27.2.18	7
RADAR BMC 805 10.7.18	8
RADAR BMC 820 10.7.18	9
RADAR BMC 820 30.10.18	10
RADAR BMC 407 30.5.18	11
RADAR BMC 407 3.8.18	12
RADAR BMC 407 24.10.18	13
RADAR BMC 074 27.2.18	14
RADAR BMC 074 19.6.18	15
RADAR BMC 074 20.9.18	16
RADAR BMC 3861 healthy	17
RADAR BMC 4264 10.7.18	18
RADAR BMC 4275 17.7.18 healthy	19
Batch control	20

#### Barcode batch 2

<u>Sample Identifier</u>	<u>Barcode</u>
RADAR BMC 4100 14.3.18	1

RADAR BMC 4100 10.4.18	2
RADAR BMC 4100 29.8.18	3
RADAR BMC 3208 20.3.18	4
RADAR BMC 3208 31.7.18	5
RADAR BMC 3208 16.10.18	6
RADAR BMC 5009 20.3.18	7
RADAR BMC 5009 10.7.18	8
RADAR BMC 3449 14.8.18	9
RADAR BMC 3449 24.8.18	10
RADAR BMC 3449 SFMCs 24.8.18	11
RADAR BMC 907 19.6.18	12
RADAR BMC 907 19.9.18	13
RADAR BMC 072 1.6.18	14
RADAR BMC 072 2.10.18	15
RADAR BMC 4266	16
RADAR BMC 3982	17
RADAR BMC 4283 healthy	18
RADAR BMC 4284 healthy	19
Batch control	20

### **Barcode batch 3**

<b><u>Sample Identifier</u></b>	<b><u>Barcode</u></b>
RADAR BMC 3351 27.3.18	1
RADAR BMC 3351 16.10.18	2
RADAR BMC 4039 2.3.18	3
RADAR BMC 4039 24.7.18	4
RADAR BMC 4039 10.10.18	5
RADAR BMC 5020 11.4.18	6
RADAR BMC 5020 14.8.18	7
RADAR BMC 5020 16.11.18	8
RADAR BMC 5021 5.6.18	9
RADAR BMC 5021 10.10.18	10

RADAR BMC 3407 6.7.18	11
RADAR BMC 3407 16.10.18	12
RADAR BMC 5016 2.3.18	13
RADAR BMC 5016 17.7.18	14
RADAR BMC 5016 10.10.18	15
RADAR BMC 2430 10.4.18	16
RADAR BMC 2430 10.10.18	17
RADAR BMC 4265 healthy 10.7.18	18
RADAR BMC 4276 healthy 17.7.18	19
Batch control	20

#### **Barcode batch 4**

<b><u>Sample Identifier</u></b>	<b><u>Barcode</u></b>
RADAR BMC 5017 19.3.18	1
RADAR BMC 5017 30.10.18	2
RADAR BMC 3593 17.4.18	3
RADAR BMC 3593 19.9.18	4
RADAR BMC 2415 20.4.18	5
RADAR BMC 2415 14.8.18	6
RADAR BMC 2415 30.10.18	7
RADAR BMC 926 9.5.18	8
RADAR BMC 926 4.9.18	9
RADAR BMC 2914 24.4.18	10
RADAR BMC 2914 11.9.18	11
RADAR BMC 3896 5.6.18	12
RADAR BMC 3896 2.10.18	13
RADAR BMC 2047 17.7.18	14
RADAR BMC 2047 30.10.18	15
<b>BLANK</b>	16
<b>BLANK</b>	17
RADAR BMC 4269 healthy 11.7.18	18

RADAR BMC 4270 healthy 11.7.18	19
Batch control	20

### **Barcode batch 5**

<b><u>Sample Identifier</u></b>	<b><u>Barcode</u></b>
RADAR 4153 1.3.18	1
RADAR 4154 1.3.18	2
RADAR 3846 25.4.18	3
RADAR 5015 1.3.18	4
RADAR 4123 25.4.18	5
RADAR BMC 1808 21.2.18	6
RADAR BMC 246 21.2.18	7
RADAR BMC 2067 16.10.18	8
RADAR 4302 12.9.18	9
RADAR 4333 31.10.18	10
RADAR 4334 31.10.18	11
RADAR 4331 31.10.18	12
NHS CCP 590 3.9.18	13
NHS CCP 188 3.9.18	14
RADAR BMC 3538 7.3.18	15
RADAR BMC 1158 18.7.18	16
RADAR BMC 5022 16.8.18	17
RADAR BMC 4274 healthy 17.7.18	18
RADAR BMC 3861 17.7.18	19
Batch control	20

### **Barcode batch 6**

<b><u>Sample Identifier</u></b>	<b><u>Barcode</u></b>
RADAR 4195 25.4.18	1
RADAR 4196 9.5.18	2
RADAR 4202 9.5.18	3

RADAR 4209 23.5.18	4
RADAR 3709 23.5.18	5
RADAR BMC 5013 21.2.18	6
RADAR BMC 1826 30.10.18	7
RADAR BMC 50 6.11.18	8
RADAR BMC 2760 19.6.18	9
RADAR BMC 1963 17.7.18	10
RADAR BMC 3105 17.7.18	11
RADAR BMC 3562 4.7.18	12
RADAR BMC 453 24.7.18	13
RADAR BMC 4086 14.8.18	14
RADAR BMC 1746 21.8.18	15
RADAR BMC 1431 30.8.18	16
RADAR BMC 2514 25.9.18	17
RADAR BMC 4268 healthy 11.7.18	18
RADAR BMC 4281 healthy 17.7.18	19
Batch control	20

### **Barcode batch 7**

<b><u>Sample Identifier</u></b>	<b><u>Barcode</u></b>
RADAR 4212 23.5.18	1
RADAR 4217 23.5.18	2
RADAR 4219 6.6.18	3
RADAR 3549 6.6.18	4
RADAR 3863 20.6.18	5
RADAR 2241 31.7.18	6
RADAR BMC 3588 1.8.18	7
RADAR BMC 2878 14.8.18	8
RADAR BMC 1020 19.9.18	9
RADAR BMC 3887 19.9.18	10
RADAR 4312 20.9.18	11
RADAR BMC 180 19.9.18	12

RADAR BMC 1740 25.9.18	13
RADAR BMC 4045 25.10.18	14
RADAR BMC 1044 7.3.18	15
RADAR BMC 1996 7.3.18	16
RADAR BMC 4266 healthy 10.7.18	17
RADAR BMC 4271 healthy 12.7.18	18
RADAR BMC 4263 healthy 10.7.18	19
Batch control	20

### **Barcode batch 8**

<b><u>Sample Identifier</u></b>	<b><u>Barcode</u></b>
RADAR 4226 20.6.18	1
RADAR 4233 20.6.18	2
RADAR 4227 20.6.18	3
RADAR 3967 20.6.18	4
RADAR 3939 20.6.18	5
<b>BLANK</b>	6
RADAR BMC 2711 30.10.18	7
RADAR BMC 3726 .11.18	8
RADAR BMC 3051 7.3.18	9
RADAR BMC 345 14.3.18	10
RADAR BMC 3565 25.9.18	11
RADAR BMC 4033 31.7.18	12
RADAR BMC 579 31.7.18	13
RADAR BMC 2518 31.7.18	14
RADAR BMC 3046 7.8.18	15
RADAR BMC 4185 22.3.18	16
RADAR BMC 874 25.9.18	17
RADAR BMC 4273 healthy 12.7.18	18
RADAR BMC 3849 healthy 17.7.18	19

Batch control	20
---------------	----

### **Barcode batch 9**

<b><u>Sample Identifier</u></b>	<b><u>Barcode</u></b>
RADAR 4225 20.6.18	1
RADAR 4228 28.6.18	2
RADAR 4229 28.6.18	3
RADAR 4230 4.7.18	4
RADAR 3968 18.7.18	5
RADAR 3611 18.7.18	6
RADAR 4247 18.7.18	7
RADAR BMC 4043 5.9.18	8
RADAR BMC 5007 20.3.18	9
RADAR BMC 2888 18.6.18	10
RADAR BMC 2576 19.6.18	11
RADAR BMC 4078 27.2.18	12
RADAR BMC 3024 14.3.18	13
RADAR BMC 4079 5.6.18	14
RADAR BMC 424 28.10.18	15
RADAR BMC 4013 10.10.18	16
RADAR BMC 2742 8.10.18	17
RADAR BMC 4267 healthy 11.7.18	18
RADAR BMC 4262 healthy 12.7.18	19
Batch control	20

### **Barcode batch 10**

<b><u>Sample Identifier</u></b>	<b><u>Barcode</u></b>
RADAR BMC 4285 SFMCs 30.7.18	1
RADAR BMC 4285 blood 30.7.18	2
RADAR BMC 5023 SFMCs 23.7.18	3

RADAR BMC 5023 blood 23.7.18	4
RADAR BMC 4288 SFMCs 30.7.18	5
RADAR BMC 4288 blood 6.8.18	6
RADAR BMC 4184 SFMCs 22.3.18	7
TECSPA 27.2.18	8
RADAR BMC 3329 SFMCs 12.10.18	9
RADAR BMC 4310 SFMCs 3.9.18	10
SF01 9.7.18	11
SF02 9.7.18	12
RADAR BMC 4171 SFMCs 24.7.18	13
RADAR BMC 4238 SFMCs 21.6.18	14
RADAR BMC 4237 blood 21.6.18	15
RADAR BMC 2624 blood 21.6.18	16
RADAR BMC 4182 blood 20.3.18	17
RADAR BMC 4272 healthy 12.7.18	18
RADAR BMC 4282 healthy 10.7.18	19
Batch control	20