

Bioinformatic Methods to Interrogate Flow
Cytometry and RNA Sequencing Data to
Understand T-cell Interactions in Disease

By

Daniel D. Griffiths

A thesis submitted

for the degree of

Master of Philosophy

Cardiff University

DECEMBER 2025

ACKNOWLEDGEMENTS

For the $\gamma\delta$ T cell gene expression analysis, RNA sequencing was performed by Angela Marchbank and Nick Kent. Cell culture experiments were carried out by Chris Tyler.

For the AL clinical study; Clinical lead was Jared Torkington. Sample preparation and staining carried out by Nicola Reeves and Sarah Baker. Data acquisition by flow cytometry was carried out by Nicola Reeves and Simone Cuff.

Principal investigator for both projects was Matthias Eberl.

Thank you to Matthias Eberl and Andreas Artemiou for their valuable assistance and advice during the conception of this project, as well as their review and corrections of this manuscript.

PUBLICATIONS

Journal

- "IL-21 conditions antigen-presenting human $\gamma\delta$ T-cells to promote IL-10 expression in naïve and memory CD4+ T-cells"

Christopher J Tyler , Inva Hoti , Daniel D Griffiths , Simone M Cuff , Robert Andrews , Maximilian Keisker , Raya Ahmed , Hinrich P Hansen , James O Lindsay , Andrew J Stagg , Bernhard Moser , Neil E McCarthy , Matthias Eberl
Discovery Immunology, **2024**, Vol. 3.

DOI: doi.org/10.1093/discim/kyae008

Other related works

- Workflow proposed with bioinformatics team to promote NF-Core usage at the department (training materials)

Daniel D. Griffiths. (2025).

dgrfs/NFCore_RNASeq_Cardiff_University_Hawk: (Version 2). Zenodo.

doi.org/10.5281/zenodo.16903920

- Data preparation and annotation of RNA sequencing data associated with the above publication. Published on EMBL-EBI under accession E-MTAB-13875

ABSTRACT

Conventional T cells are defined by expression of CD4⁺ or CD8⁺ surface antigens. These are the most abundant of the CD3⁺ T cell lineage, and are the most characterised in their role in immune response, regulation, and disease. Unconventional T cells do not present these conventional markers, and are not fully characterised. These cells are of low frequency in peripheral blood and are tissue resident, with phenotypes and roles that are tissue specific. Tissue specific analysis of immune cells allows these unconventional immune cells, which are of low frequency / absent from peripheral blood, to be analysed.

Colorectal cancer treatment often involves colorectal resection, removal of the tumour, and stitching the remaining colon back together; this process is called anastomosis. Anastomotic leak (AL) occurs in approximately 10% of patients undergoing anastomosis and is associated with significantly increased mortality and complications. Current diagnostic confirmation relies on elevated systemic C-reactive protein (CRP) and clinical symptoms, typically enabling detection only from postoperative day 2-3. Earlier and more specific markers are needed to improve patient outcomes. Evidence from peritonitis studies suggests that unconventional T cells in peritoneal fluid may predict complications, motivating investigation in AL. Innate immune cells such as MAIT and $\gamma\delta$ T cells are often the first immune cells at the site of a bacterial infection.

Previous work by the Eberl lab indicated a regulatory role of $\gamma\delta$ T cells on CD4⁺ T cells, and collected RNA sequencing data of $\gamma\delta$ T cells in different relevant tissue-specific cytokine environments for a more in depth mechanistic understanding. In this thesis, a workflow for quality control is proposed and analysis of the data is performed.

Previous researchers acquired many flow cytometry files in many individual batches and with variable quality. This thesis discusses the ability for these flow cytometry files to be analysed together and batch-corrected using publicly available packages,

within a reproducible bioinformatic pipeline. These findings suggest that unconventional T cells are in measurable frequency using flow cytometry, and sets a proof-of-concept for the use of peritoneal fluid to measure immune cell frequency, particularly of tissue resident unconventional T cells. Further, RNA sequencing analysis found CXCL13 and CD30 expression correlating with an inflammatory related IL-21 environment, with indications of the role of $\gamma\delta$ T cells in initiating an immune response at the site of infection.

This data further advances the understanding of innate immune response, and indicates potential candidates for novel, specific targets for immune therapy for chronic inflammatory conditions. These initial studies also highlight improvements to study design and data acquisition that can improve quality of data along the course of a longitudinal study.

ABBREVIATIONS

AL	Anastomotic leak
APC	Antigen presenting cell
BTN	Butyrophilin
CRP	C-reactive protein
CRC	Colorectal cancer
ERAS	Enhanced Recovery After Surgery
HMB-PP	(E)-4-hydroxy-3-methyl-but-2-enyl pyrophosphate
IBD	Inflammatory bowel disease
IFN	Interferon
ILC	Innate lymphoid cell
LOS	Length of stay
MAIT	Mucosal-associated invariant T cell
NICE	National Institute of Health and Care Excellence
NKT	Natural killer T cell
PD	Peritoneal dialysis
PRR	Pattern recognition receptors
SSI	Surgical site infection
TCR	T cell receptor
TLO	Tertiary lymphoid organ

CONTENTS

Introduction	1
Anastomotic leak	3
Improving indicators for anastomotic leak – current methods	3
Validation of peritoneal fluid for biomarker discovery	4
Study Rationale and Hypothesis	5
Aims and Objectives	6
 1 Improving the identification of high risk patients after colorectal resection	 7
1.1 Introduction: Colorectal cancer	7
1.1.1 Incidence	7
1.1.2 Treatment and survival	8
1.1.3 Perioperative recovery	10
1.1.4 Complications to colorectal surgery	11
1.1.5 Detecting anastomotic leak early	11
1.1.6 Innate immune cells and anastomotic leak	12
1.1.7 Aims	13
1.2 Methods	15

1.2.1	Ethics approval	15
1.2.2	Study design	15
1.2.3	AL clinical study data analysis	18
1.3	Results	21
1.3.1	Cohort	21
1.3.2	Flow cytometry analysis	23
1.3.3	Identifying technical effect caused by flow rate using FlowAI .	26
1.3.4	Peritoneal fluid monocytes and CD8 T cell frequencies correlate with perioperative complications	28
1.3.5	Supervised analysis	30
1.3.6	Unsupervised analysis	32
1.3.7	Preparation of flow cytometry data for unsupervised clustering	33
1.3.8	Factors correlating with highest variation in the data	35
1.3.9	Dimensionality reduction	41
1.4	Discussion	50
1.4.1	Flow cytometry analysis of peritoneal fluid	51
1.4.2	Key findings in immune profiling	56
1.4.3	Context of peritoneal fluid analysis in other diseases	57
1.4.4	Interpretation and clinical implications	58
1.4.5	Clinical cohort size and power calculations	58
1.4.6	Sample Size Power Calculation	59
1.4.7	CRP is an accurate predictor of AL	60
1.4.8	Future direction	62
2	Unconventional T cells: Response to acute bacterial infection	65
2.1	Introduction	65
2.1.1	Unconventional T cells	65
2.1.2	$\gamma\delta$ T cells facilitate immune response to bacterial infection . .	68
2.1.3	The role of $\gamma\delta$ T cells in inflammation	68

2.1.4	$\gamma\delta$ T cells at the intestinal barrier	70
2.1.5	Understanding the role of $\gamma\delta$ T cells in regulating inflammation	71
2.1.6	Selecting conditions for investigation	71
2.1.7	Aims	72
2.2	Methods	73
2.2.1	RNA sequencing	73
2.2.2	RNA Sequencing data analysis	73
2.2.3	Data availability	74
2.3	Results	75
2.3.1	Experiment design	75
2.3.2	Quality control	76
2.3.3	Sequencing saturation	78
2.3.4	DESeq2 Differential expression workflow	79
2.3.5	Identifying highly variable genes	80
2.3.6	Dimensionality reduction of RNA sequencing data	83
2.3.7	Identifying differentially expressed genes	84
2.4	Discussion	87
2.4.1	Cohort size	87
2.4.2	IL-21 condition – Differentially expressed genes	89
2.4.3	Tertiary lymphoid structure	90
2.4.4	Future direction	91
	Discussion	92
	Improving the quality of flow cytometry data	93
	Workflows and pipelines	94
	Conclusion	96

LIST OF FIGURES

1.1	Left anterior resection	9
1.2	Incidence of perioperative complications – Cohort	22
1.3	Gating strategy for T cell panel.	24
1.4	Gating strategy for neutrophil cell panel.	25
1.5	Anomaly detection	27
1.6	Correlogram of peritoneal fluid outputs with blood markers and patient demographics.	29
1.7	Mean frequency of monocytes, grouped by perioperative complication.	31
1.8	Grouped mean plot – CD8	32
1.9	Arcsinh data transformation	34
1.10	Dimensionality reduction of neutrophil panel samples.	36
1.11	Dimensionality reduction of T cell panel samples.	37
1.12	Random forest model results for neutrophil and T cell panel	40
1.13	Dimensionality reduction: T cell panel	42
1.14	Dimensionality reduction: Neutrophil panel	43
1.15	Heatmap of neutrophil panel meta-clusters	45
1.16	Heatmap of T cell panel meta-clusters	47

1.17	Values of clustering plotted against manual gating for identification of CD8 T cells, with a limma trend applied.	49
1.18	Grouped mean plot – Blood CRP	61
1.19	Preprocessing pipeline	64
2.1	Tissue specific honing of $\gamma\delta$ T cells	69
2.2	Representative example of sequencing saturation.	79
2.3	VST transformed RNA sequencing data	80
2.4	Heatmap showing variance across all samples	82
2.5	MDS plot of RNA sequencing dataset	84
2.6	Volcano plot IL-21 condition against IL-2 and IL-15	86

LIST OF TABLES

1.1	Patient demographics and characteristics	17
1.2	Table of neutrophil panel including cell type associated with marker .	19
1.3	Table of T cell panel	19
1.4	T cell panel gating strategy	20
1.5	Statistical Parameters for Sample Size Calculation	60
2.1	Defining features of conventional, unconventional and innate immune cells.	67
2.2	RNA sequencing sample metadata	76

INTRODUCTION

Conventional T cells are specialised lymphocytes that identify antigen-presenting cells (APCs). APCs present antigens via the MHC class I molecule for CD8⁺ T cells and the MHC class II molecule for CD4⁺ T cells, which is then recognised by the T-cell receptor (TCR) of the T cells. The TCR of the conventional T cell consists of highly variable $\alpha\beta$ chains, which recognise antigen-presenting peptide-MHC (pMHC) complexes. In contrast, unconventional T cells are, in part, defined by their ability to recognise antigens without reliance on pMHC complexes. These include natural killer (NK) T cells, Mucosal-associated invariant T (MAIT) cells, and $\gamma\delta$ T cells.

In humans, $\gamma\delta$ T cells are found in relatively low frequency in peripheral blood (1-5%), and enriched at barrier tissues such as gut mucosa, skin and lungs. Human $\gamma\delta$ T cells can be classified into different groups based on the expression of TCR γ chains, TCR δ chains, cell surface markers and tissue-specificity (Fichtner et al. 2020, Li et al. 2021). Unlike conventional T cells, $\gamma\delta$ T cells are not reliant on maturation in the thymus and undergo continuous extra-thymic clonal selection (Di Lorenzo et al. 2017). They are referred to as semi-invariant due to some γ and δ combinations more

prevalent than others (Davey et al. 2017). Though they are of low relative frequency, $\gamma\delta$ T cells are powerful drivers of pro- and anti-inflammatory responses. V γ 9/V δ 2 are the most abundant peripheral blood $\gamma\delta$ T cells, with this chain-pairing being highly conserved in humans. Many motifs are also conserved between individuals, including the CDR3 region of the TCR structure (Sherwood et al. 2011).

$\gamma\delta$ T cell in regulating immune response

$\gamma\delta$ T cells have been shown to regulate immune cells at the site of bacterial infection to induce an effective immune response (Wozniak et al. 2012). These cells are highly optimised for specific cytokine and tissue environments, and have a significant role in clearing infection, as shown by a decrease in survival of patients with low $\gamma\delta$ T cell frequency while suffering from sepsis (Andreu-Ballester et al. 2013). Due to their tissue specificity, rapid innate immune response and regulatory role, this immune cell subset is a logical candidate biomarkers and targeted immunotherapy. To further understanding of the mechanistic role of $\gamma\delta$ T cells, gene expression changes to specific cytokine environments were investigated *in vitro*. A further understanding of differentially expressed genes in these environments can provide insight into how $\gamma\delta$ T cells initiate an immune response and recruit conventional immune cells to the site of infection. Not only are $\gamma\delta$ T cells a potentially tissue and disease specific marker for early detection of an immune response (and therefore an acute infection), they may also have a more general role in the regulation of that immune response, and so could be a target for immunotherapy in cases of chronic inflammation.

Anastomotic leak

Clinical manifestations of AL include abdominal pain and tachycardia, with diagnosis relying on the detection of purulent or fecal discharge from peritoneal drains and radiological evidence, leading to peritonitis and sepsis in the absence of medical intervention (Wako et al. 1970). C-reactive protein (CRP) monitoring remains a standard method for detecting AL, with levels reaching significance 48 to 72 hours postoperatively. CRP levels, combined with clinical observations such as abdominal pain localisation, contribute to the Dutch leakage score, aiding AL detection (den Dulk et al. 2013).

Improving indicators for anastomotic leak

Although the involvement of immune markers, particularly lymphocytes, in AL is documented, the extent of their role during infection is not fully understood. Therefore, investigating the immunological profile of patients with AL may elucidate specific involvement of T cell subsets. Understanding the immunological profile at the anastomotic site can potentially identify AL-related markers, enhancing the current monitoring system's specificity and early detection capabilities, thereby improving patient outcomes.

Immunophenotyping drain fluid may provide predictive insights into clinical outcomes, as demonstrated in peritoneal dialysis (PD) contexts, where it aids in detecting peritonitis (Liuzzi et al. 2016). For patients requiring PD, methods such as bacterial 16S ribosomal RNA sequencing and leukocyte esterase reagent strips show promise in improving infection detection and by extension improving patient outcomes (Fung

& Li 2022).

Validation of peritoneal fluid for biomarker discovery

As discussed above, there is a clear gap in knowledge in the understanding of mechanism and biomarkers allowing for the early detection of AL. Previous work has identified that those requiring peritoneal dialysis (PD) have biomarkers detected at site of infection, detected through peritoneal fluid, which were not found in systemic measurements such as blood CRP. This provides a logical proof-of-concept that could be adapted to perioperative anastomosis patients.

Previous studies were used to inform the experimental design. Liuzzi et al. (2016) found that PD patients undergoing inflammation in response to bacterial infection had $V\gamma 9/V\delta 2$ T cells and MAIT cells accumulating at the site of infection, as detected through peritoneal fluid. With a cohort of 101 PD patients, an increase in innate T cells were detected in peritoneal fluid of patients with acute infection compared to when they were in a stable condition. They found a statistically significant increase in $V\gamma 9/V\delta 2$ and MAIT cells in PD patients during acute infection. This paper provides context and a foundation to apply to patients with AL, who are also undergoing acute infection. Similar processing methods can be applied in the context of AL. There are, however, some differences between previous analysis methods and the current AL study. This paper had matched control samples of the same individuals before and during acute infection, which cannot be done in this study. There is no typical peritoneal fluid volume in healthy individuals. In this specific case, it is not possible to obtain a control condition for each individual in this way. Therefore, the comparison is between groups of people – those who recover well, and those who

develop AL. This is more informative, as the difference between these two 'at risk' groups can be characterised.

Previous research suggests that analysing peritoneal fluid cytokine levels, particularly IL-6, alongside blood CRP levels, enhances AL detection accuracy and may enable earlier intervention (Cuff et al. 2023). The informative cytokine changes and clinical metadata (from peripheral blood) can be used to improve the analysis of the flow cytometry data and provide additional context.

In summary, investigating immune markers in peritoneal fluid can potentially improve AL detection, prognostics, and patient care; potentially guiding personalised treatment strategies and improving long-term outcomes in CRC surgery.

Study Rationale and Hypothesis

This thesis tests the hypothesis of detecting local immune response during AL using unconventional T cell markers, and potential for biomarkers for early detection of complications from colorectal resection. Alongside the clinical cohort, the specific mechanistic role of $\gamma\delta$ T cells is investigated by gene expression analysis.

Aims and Objectives

- Validate the use of peritoneal fluid for detecting unconventional T cells.
- Evaluate methods used for flow cytometry data and suggest improvements for standardisation and reproducibility in analysis
- Identify immune markers associated with AL.
- Investigate cytokine-driven transcriptional changes in $\gamma\delta$ T cells using RNA sequencing data.
- Propose candidate biomarkers for early detection and intervention during colorectal surgery recovery.

The first three aims are addressed in Chapter 1, through the anastomotic leak clinical study, and the final two aims are addressed in Chapter 2, through the $\gamma\delta$ T cell cytokine-stimulation study.

CHAPTER 1

IMPROVING THE IDENTIFICATION OF HIGH RISK PATIENTS AFTER COLORECTAL RESECTION

1.1 Introduction: Colorectal cancer

1.1.1 Incidence

Colorectal cancer (CRC) is the third most common cancer globally and ranks second in cancer-related mortality (Sung et al. 2021). In the UK, CRC is the fourth most common cancer, with 44,100 new cases each year between 2017 - 2019. Cancer Research UK (CRUK) shows the incidence of colorectal cancer is increasing, particularly with younger demographics (0 - 24 years). The reasoning behind this could be dependent on socio-economic factors and requires further study to understand root causes. In Wales, Scotland and England bowel cancer screening is now offered to everyone aged 50 - 74 years old. Targeted screening is also conducted in the UK. This involves increased monitoring of patients with familial-linked mutations in BRAF V600E and

RAS, which is closely associated with more aggressive forms of colorectal cancer that can resist usual treatment pathways.

1.1.2 Treatment and survival

The survival rates for CRC have greatly improved since the 1990s, with almost 60% of people living for at least 5 years after initial diagnosis. The National Institute for Health and Care Excellence (NICE) provide recommendations for healthcare within the UK. In cases with curative intent, most patients diagnosed with colorectal cancer are recommended for laparoscopic surgery for the removal of the tumour (NICE TA105 2006). NICE reporting shows that 70 - 80% of individuals diagnosed with colorectal cancer undergo complete surgical excision, with the remaining individuals likely to require palliative care and treatment without curative intent.

Surgical resection of the colon is a common treatment for CRC, aiming to remove the tumour and restore bowel function (Figure 1.1). The process of early detection of cancer and implementing appropriate treatments has greatly improved, especially for colorectal cancer. Depending on the stage of disease, various treatments are available.

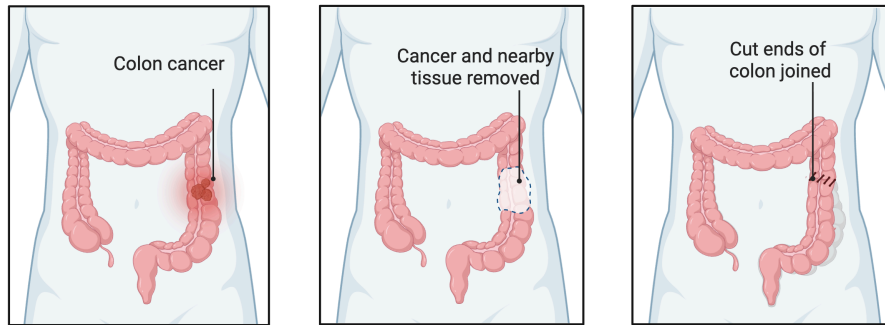


Figure 1.1: Left anterior resection

In order to maintain function of the bowel while removing the tumour, a colorectal resection and reattachment of the healthy ends of the colon is performed (anastomosis). The example shown here describes a low anterior resection. The three panels show the tumour location within the colon, the section of bowel containing the tumour and surrounding margin marked for removal, and the resulting anastomosis where the two cut ends of the colon are rejoined; it is at this anastomotic join that a leak, the principal complication studied in this chapter, can subsequently develop. Created using biorender (biorender.com).

Colon resection is conducted via laparoscopy or laparotomy, with the latter usually being higher risk. Recent improvements in patient recovery are related to standardising recommendations for specific cancers, depending on location of tumour and cancer stage - in both treatment options and perioperative care.

To improve treatment at a global scale a consortium was set up, called enhanced recovery after surgery (ERAS). ERAS was set up with an initial cohort in the late 90s and early 2000s, initially focused on colorectal surgery procedure. This began with a series of small studies and literature reviews led by Professor Henrik Kehlet, providing an evidence-based approach to reducing the length of stay at hospitals post surgery. The first article by Kehlet defines factors associated with recovery that can be monitored and improved upon (Kehlet 1997). In 2005, the first ERAS care protocol was published, which included reduced patient fasting both pre- and post-surgery, and encouraging early mobilisation after surgery (Fearon et al. 2005).

Since then, the recommendations have expanded towards many different types of surgery. The first ERAS study was borne out of an attempt to standardise care across different hospitals, which at this time had a range of many different approaches and recovery protocols. The most recent ERAS recommendations (Gustafsson et al. 2019) cover both colonic and rectal surgery, with systematic reviews indicating a reduction of about 40% in perioperative morbidities across different studies between 2015 - 2024 (Kannan et al. 2025). Not only have these recommendations improved perioperative care, there was a significant increase in 3-year survival (Tidadini et al. 2022). This, along with better tailored treatment, is crucial to improving patient outcomes both during immediate recovery and for longer term survival.

1.1.3 Perioperative recovery

In the immediate 2-3 days after surgery, some patients have an abdominal drain inserted to reduce swelling and the build up of fluid in the peritoneal cavity. There is ambivalence towards the usage of drains as part of perioperative care, and inconclusive guidance. The ERAS society did not find improved recovery with the use of an abdominal drain (Gustafsson et al. 2019), and this was corroborated following systematic review by other researchers (Emile & Abd El-Hamed 2017).

Retrospective studies show that the previous median length of stay (LOS) across different cohorts showed a range of 8-12 days in the early 2000s (Aravani et al. 2016), with a median LOS of 3 days more recently. The rate of recovery has greatly improved in recent decades, assisted by better recovery protocols and more detailed record keeping and monitoring. Not only does a shorter recovery time reduce burden on the healthcare service, shorter LOS correlated with better outcomes, with fewer

readmissions (Emmanuel et al. 2018).

1.1.4 Complications to colorectal surgery

Surgical resection (anastomosis) of the colon is a common treatment for CRC, aiming to remove the tumour while also restoring function. This is a preferred treatment option as opposed to removal of the tumour and creating a stoma – using a stoma reduces quality of life for the patient and increases a life-long risk of infection. Therefore, tumour removal and creating a stoma is usually avoided unless completely necessary.

Although anastomosis is a preferred treatment pathway, a notable complication of anastomosis is anastomotic leak (AL), in which intestinal contents leak into the peritoneal cavity at the site of resection. Clinical manifestations of AL include abdominal pain and tachycardia leading to peritonitis and sepsis in the absence of medical intervention (Rajabaleyan et al. 2024)

Various factors contribute to the increased risk of AL, including smoking, male gender, obesity, operative duration, blood transfusions and tumour proximity to the anal verge (Zarnescu et al. 2021). AL adversely affects patient quality of life; leading to prolonged hospital stays, increased likelihood of stoma requirement, and reduced 5-year survival rates (Hammond et al. 2014).

1.1.5 Detecting anastomotic leak early

Previous work has identified that those requiring PD have biomarkers detected at site of infection, detected through peritoneal fluid, that was not found in systemic measurements by analysing blood. Liuzzi et al. (2016) found that PD patients undergoing inflammation in response to bacterial infection had V γ 9/V δ 2 T cells and MAIT cells accumulating at the site of infection, as detected through peritoneal fluid.

This provides an opportunity to validate this method of peritoneal fluid analysis in the context of AL with a particular focus on the implied role of MAIT and $V\gamma 9/V\delta 2$ T cells at the site of local infection. To be able to assess potential biomarkers, previous work in patients undergoing PD provides a proof of concept that is worth also investigating in other situations such as during perioperative colorectal surgery. Neutrophils and lymphocytes have been previously associated with AL, and higher relative neutrophil count has been shown to correlate with risk of AL (Liu et al. 2026). Therefore, investigating the immunological profile of patients with AL may elucidate specific involvement of immune cell subsets.

Peritoneal fluid was collected from patients with an abdominal drain during the period immediately after surgery. The efficacy of abdominal drains after colorectal surgery is under debate. ERAS suggests that an abdominal drain should not be used as standard procedure, but only when there is a specific indication (for example, when there is localised swelling). A recent review suggests the rate of abdominal drain usage is varied between countries, at around 40 - 70 %, and that the addition of a drain increased length of stay (Guadagni et al. 2024).

In this study, peritoneal fluid that would otherwise be a waste product becomes input material for analysis, and a possible route of measuring biomarkers indicating risk of developing AL before clinical observations are apparent.

1.1.6 Innate immune cells and anastomotic leak

MAIT cells are considered unconventional T cells. They are found in peripheral blood at a frequency of 10% of lymphocytes. They have semi-invariant TCRs and are restricted to the non-polymorphic MHC class I-like protein MR1. MAIT cells can

respond rapidly to infection with both TCR and non-TCR activation (Godfrey et al. 2019).

Previous research by the Eberl Lab suggests that analysing peritoneal fluid cytokine levels, particularly IL-6, alongside blood CRP levels, enhances AL detection accuracy and may enable earlier intervention (Cuff et al. 2023). This provides further insight into the ability to analyse peritoneal fluid to get similar outputs to markers found in blood samples. The next step is to further characterise the peritoneal fluid across multiple individuals, to determine if there is a consistent subset that correlates with postoperative complications. Crucially, this will allow the hypothesis to be tested that AL infection-related innate immune cells increase at the site of infection, measured using peritoneal fluid, faster than any changes in immune cells systemically, measured using blood samples.

This study complements the previous work on cytokine measurements, with the addition of flow cytometry analysis. This data was acquired previously, but was not included in the 2023 paper as the flow cytometry data had batch effect issues and proved difficult to gate rare populations consistently.

In summary, investigating immune markers in peritoneal fluid can improve AL detection, prognostics, and patient care; potentially guiding personalised treatment strategies and improving long-term outcomes in CRC surgery.

1.1.7 Aims

There are two main aims of this study. First, to determine whether peritoneal fluid samples can be reliably collected and analysed to quantify the frequency of innate immune cells. Second, to assess whether these immune cell profiles can be used to

predict the occurrence of anastomotic leak (AL). This will test the hypothesis that innate immune cells are among the earliest responders to acute infection and accumulate locally at the site of AL before systemic changes are detectable. Identifying such biomarkers could improve early detection of AL during perioperative recovery, potentially reducing the morbidity rate.

1.2 Methods

1.2.1 Ethics approval

All methods were carried out in accordance with relevant guidelines and regulations. Experimental protocols were approved by Cardiff University, and written informed consent was obtained from all subjects. Recruitment of patients was approved by the Wales Cancer Bank under reference no. 18/016 and conducted according to the principles expressed in the Declaration of Helsinki.

1.2.2 Study design

Samples were extracted, collected and analysed by previous lab members of the group. Storage and sample collection as described in Cuff et al. (2023). Antibody staining, flow cytometry and technical quality control was carried out by the same lab members.

The data contains longitudinal sampling of 41 patients, 3 of whom developed AL. This was complemented by blood measurements and other relevant clinical covariates such as BMI and age. The aim of this study is to process the flow cytometry data of these individuals, perform quality control, and to determine if the relative abundance of the cell types identified in the flow cytometry data correlates with anastomotic leak. The full description of the clinical metadata found below in Table 1.1.

Table 1.1: Patient demographics and characteristics

Characteristic	Anastomotic leaks ($n = 3$)	Other complications ($n = 6$)	Uncomplicated recoveries ($n = 31$)
Age in years, mean (range)	56 (31–82)	61 (54–70)	66 (40–87)
Sex			
Male	2	3	16
Female	1	3	15
Body mass index, mean (range)	32.1 (21.3–41.0)	27.1 (23.4–37.0)	27.5 (19.3–35.3)
Smoking status			
Smoker	1	0	6
Non-smoker	2	3	20
Ex	0	3	5
Chemotherapy			
Yes	1	1	3
No	2	5	28
Radiotherapy			
Yes	1	1	4
No	2	5	27
Approach			
Laparoscopic	2	5	26
Open	1	1	5
Operation time in minutes, mean (range)	285 (210–390)	225 (140–320)	250 (120–435)

One further patient, who consented and continued in the study but received no surgical intervention, is excluded from this table as they do not fit any of the three outcome groups; see the Cohort description in Results for details.

1.2.3 AL clinical study data analysis

Flow cytometry data were analysed using **FlowJo**TM v10.10.0 Software (BD Sciences). Relative frequencies of cell subsets compared to CD3⁺ live cells were calculated for each sample and exported as statistics. The exported statistics files containing frequencies of CD3⁺ live cells were imported into R version 4.3.1 (2023-06-16) using **RStudio** IDE version 2023.6.1.524 (Posit team 2023). The **tidyverse** package (version 2.0.0) was utilised for data manipulation, including filtering and summarising the data. **ggplot2** version 3.4.4 package was used to create scatter plots and other visualisation graphs. **ComplexHeatmap** package version 2.18.0 was utilised to generate heatmaps displaying the expression patterns of differentially expressed genes or cell populations.

Flow cytometry panel

The design of the panels is based on immune cells that would be expected to be at the site of acute bacterial infection. These were split over two separate panels, named the neutrophil panel (Table 1.2) and T cell panel (Table 1.3 and 1.4). For the neutrophil panel, peritoneal fluid was analysed same-day, as neutrophils are not suitable for cryopreservation. The T cell panel was run on PBMC extracted peritoneal fluid and run within the same week. Usually, 1-2 patients were recruited a week, and so one batch was one patient and all timepoints.

Table 1.2: Table of neutrophil panel including cell type associated with marker

Machine parameter	Marker	Description
APC-Cy7-A	CD3	T cell marker
Alexa Fluor 405-A	CD1c	Dendritic cell marker
Alexa Fluor 488-A	CD14	Monocyte marker
Alexa Fluor 647-A	SIGLEC8	Eosinophil marker
Alexa Fluor 700-A	CD45	Leukocyte marker
AmCyan-A	Live/Dead	Viability marker
BV605-A	CD15	Granulocyte marker
BV711-A	HLA-DR	MHC Class II marker
PE-A	CD116	Myeloid lineage marker
PE-Cy7-A	CD19	B cell marker
PerCP-A	CD16	Mature neutrophil marker

Table 1.3: Table of T cell panel

Machine parameter	Marker
7-AAD-A	Pan- $\gamma\delta$
APC-Cy7-A	CD3
Alexa Fluor 405-A	CCR7
Alexa Fluor 488-A	CXCR3
Alexa Fluor 647-A	CD161
AmCyan-A	Live/Dead
BV605-A	V α 7.2
BV711-A	CD8
PE-A	V δ 2
PE-CF594-A	CD45RA
PE-Cy5-5-A	CD4
PE-Cy7-A	CD27

Table 1.4: T cell panel gating strategy

Cell type	Gating definition
Live lymphocytes	Live/Dead ⁻
Total T cells	CD3 ⁺ /Live-Dead ⁻
$\gamma\delta$ T cells	CD3 ⁺ /Pan- $\gamma\delta$ ⁺
V δ 2 $\gamma\delta$ T cells	CD3 ⁺ /Pan- $\gamma\delta$ ⁺ /V δ 2 ⁺
Non-V δ 2 $\gamma\delta$ T cells	CD3 ⁺ /Pan- $\gamma\delta$ ⁺ /V δ 2 ⁻
Conventional T cells	CD3 ⁺ /Pan- $\gamma\delta$ ⁻
CD4 ⁺ T cells	CD3 ⁺ /Pan- $\gamma\delta$ ⁻ /CD4 ⁺ /CD8 ⁻
CD8 ⁺ T cells	CD3 ⁺ /Pan- $\gamma\delta$ ⁻ /CD8 ⁺ /CD4 ⁻
Naïve T cells	CD3 ⁺ /CD45RA ⁺ /CCR7 ⁺ /CD27 ⁺
MAIT cells	CD3 ⁺ /Pan- $\gamma\delta$ ⁻ /V α 7.2 ⁺ /CD161 ^{hi} /CD4 ⁻

Gating strategy for T cell subsets from panel shown in Table 1.3. Used on PBMC-isolated peritoneal fluid.

1.3 Results

This project has a primary aim of obtaining flow cytometry data of peritoneal fluid of perioperative patients after colorectal resection. Both manual gating and clustering were used to assess if there is any correlation with outcome by using associated clinical data. Clustering techniques were also used to identify batch effect. Discovery of such biomarkers in peritoneal fluid could aid in the accurate and early diagnosis of AL in a clinical setting.

1.3.1 Cohort

Peritoneal fluid was collected from 41 individuals, with 3 developing AL (Figure 1.2). One further patient consented and continued in the study but received no surgical intervention; as they did not fit any of the three outcome groups and represented a sample size of one, they were excluded from further analysis. Most individuals had timepoints for day 1 to day 3 post operation, and the drain was removed for the majority of patients by day 4. To accompany the peritoneal fluid, clinical metadata was available for the patients at each timepoint during their recovery.

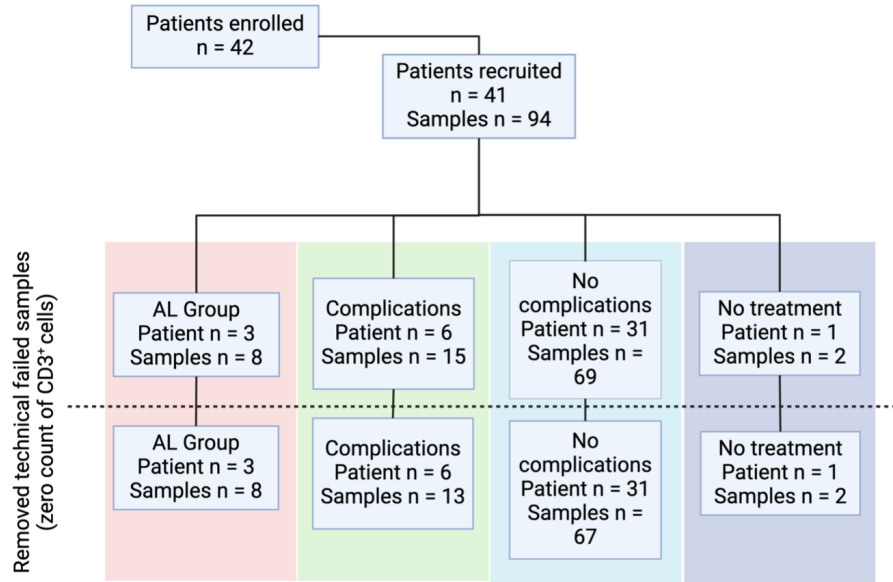


Figure 1.2: Incidence of perioperative complications – Cohort

42 patients were consented, with 41 continuing participation in the study. Of those, 3 patients went on to have AL, 6 had other complications, and 31 had no complications. One further patient received no surgical intervention and was excluded from the anastomotic-leak comparison. Samples with a failed CD3⁺ cell count of zero were removed at the technical quality-control stage (dashed line), leaving 8, 13, 67 and 2 samples available for analysis in the AL, complications, no-complications and no-treatment groups respectively.

The first step of quality control was to attempt to gate CD3⁺ cells for each of the files. If there was a recognisable positive and negative, then the sample would be analysed. If not, the sample was removed from further analysis. As most of the cell types of interest were downstream of T cells, the files that had no CD3⁺ cells would not provide any further information of T cell subsets.

1.3.2 Flow cytometry analysis

Gating strategy

The data was first analysed with manual gating. The definitions of the manual gating is shown below (Figure 1.3, Figure 1.4). There were some markers such as SIGLEC8 which could not identify a consistent positive eosinophil population. There were relatively few immune cells for the neutrophil panel as this was processed fresh and had more contaminants compared to the samples processed for the T cell panel. These gating strategies enable identification of the innate T cells in the T panel, with monocytes and neutrophils being the primary output from the neutrophil panel.

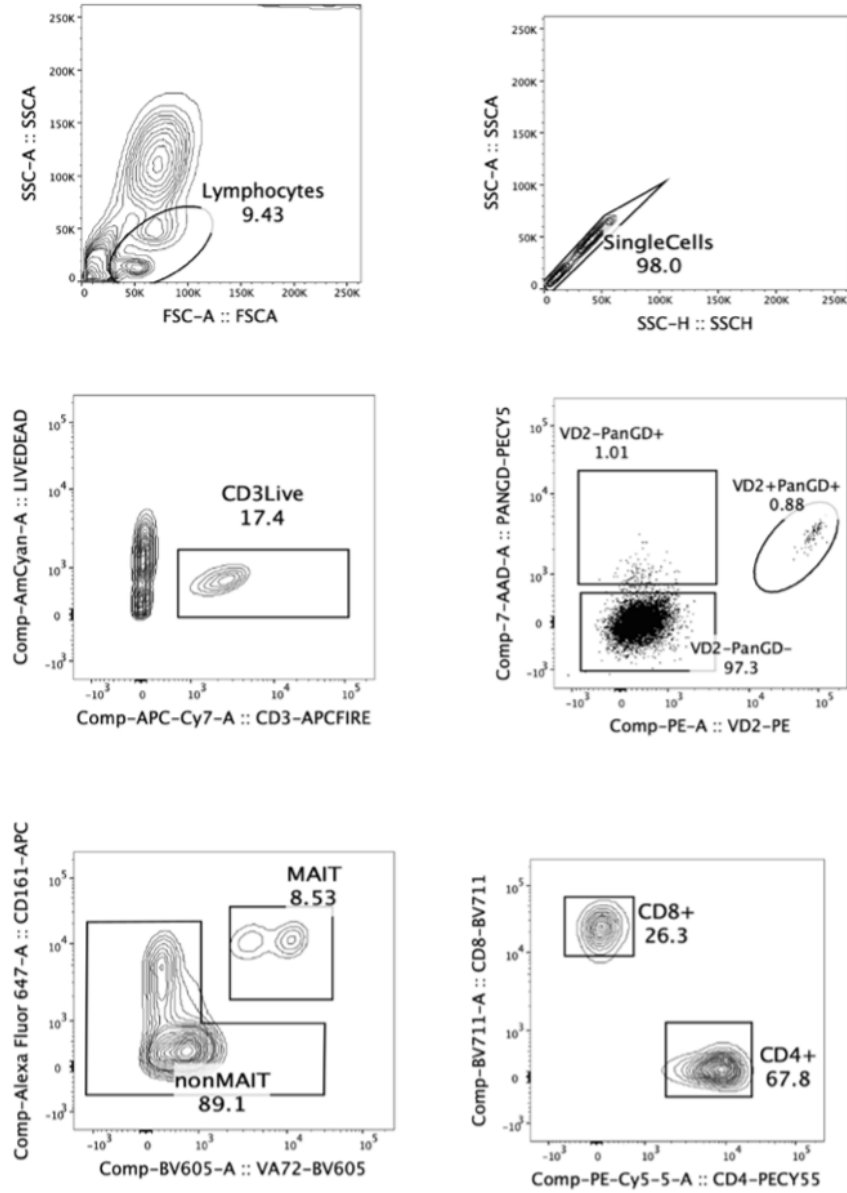


Figure 1.3: Gating strategy for T cell panel.

Lymphocytes were identified by forward and side scatter, followed by exclusion of doublets and gating on live, CD3⁺ T cells. $\gamma\delta$ T cells were identified using pan- $\gamma\delta$ and V δ 2 antibodies, distinguishing V δ 2⁺ from V δ 2-negative $\gamma\delta$ T cell subsets, while MAIT cells were excluded from the remaining T cells by CD161 and V α 7.2 co-expression. The non-MAIT population was then divided into CD4⁺ and CD8⁺ T cells.

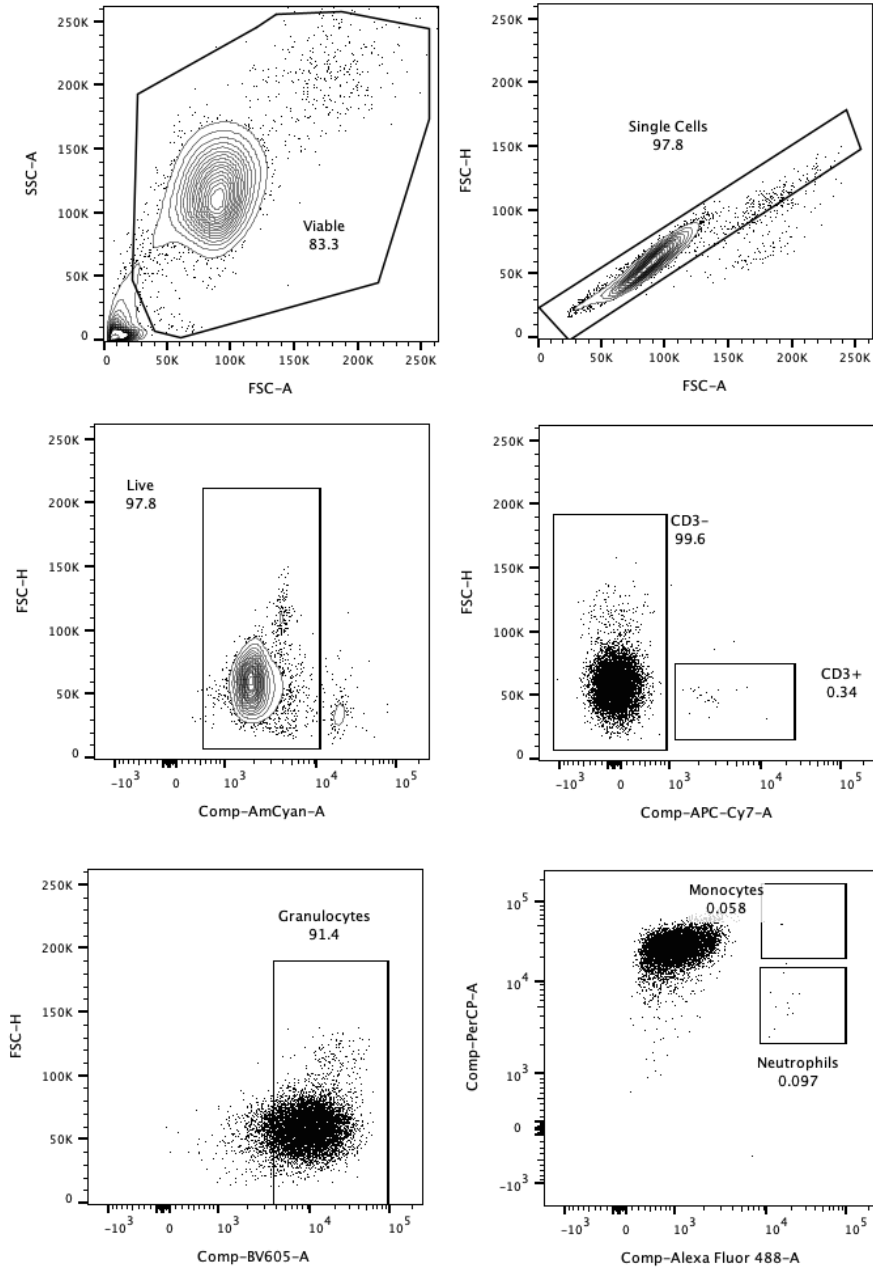


Figure 1.4: Gating strategy for neutrophil cell panel.

Viable cells were identified by forward and side scatter, followed by exclusion of doublets on single cells, and gating on live, CD3-negative events to exclude T cells. Granulocytes were then identified by forward scatter, within which monocytes and neutrophils were resolved by surface marker staining, though both represented only a small proportion of the gated events.

1.3.3 Identifying technical effect caused by flow rate using FlowAI

There are common technical effects that impact the quality of the data acquisition. This includes flow rate, signal acquisition and dynamic range. These can be identified and removed by R package / FlowJo plugin FlowAI (Monaco et al. 2016). FlowAI reads an `.fcs` file, identifies anomalies, exports an `.fcs` file with the anomalies removed, and creates a report of the number of anomalies per file. FlowAI defines anomalies as recorded events that have one or more of the following characteristics: 1) changes in flow rate 2) instability of signal or 3) outliers in upper/lower limit of dynamic range. Using these exported statistics, an overview can be obtained of the number of anomalies over the course of the project (Figure 1.5). With this dataset, most of the anomalies are due to changes in flow rate. Changes in flow rate will cause the events to have altered expression profiles, and so are not comparable to the rest of the data collected during stable flow rate. Changes in flow rate are common at the beginning and end of data acquisition, particularly if the sample is being ‘run dry’ and the whole sample is used. Small clogs can form in the fluidics that can be detected during this step, and the sections of data affected by clogs can be removed.

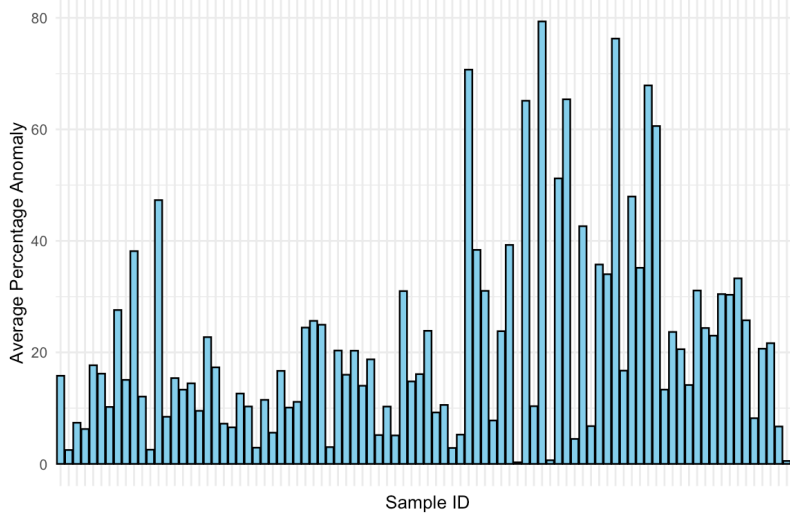


Figure 1.5: Anomaly detection

Percentage of anomalies detected by flowAI for each file. The order of the files are in chronological order in terms of date of acquisition, from left to right on the x axis. This indicates that there is some variability in the anomalies detected, with a range of 0.28% to 79.35%. For any samples that were repeated, the value was averaged. High anomaly percentages can arise from batch effects, instrument fluctuations or handling artefacts introduced during sample acquisition, and this metric was used alongside the other quality-control steps described in this section to flag samples requiring closer inspection rather than automatic exclusion.

Detection of anomalous events using **flowAI** can be done before any other analysis, including any gating. Some samples had up to 79.35% of events removed due to poor quality, suggesting that there are technical issues affecting the data over the course of the study. Subsequent analysis focuses on immune cell subsets of higher frequency that are less likely to be affected by technical effect. For this reason, subsequent analysis focused on CD8 T cells and monocytes, which had a detectable cell type frequency, found in all **.fcs** files after the **flowAI** quality control step.

1.3.4 Peritoneal fluid monocytes and CD8 T cell frequencies correlate with perioperative complications

After appropriately gating the flow cytometry files, the data was exported from the gated subsets. The statistics were of each subset as a percentage of live singlets. Using the `cor` package in R, it is possible to identify any potential markers that would be of further investigation. The correlations were calculated using Kendall's τ non-parametric test for a rank-based measure of association (Figure 1.6). Kendall's τ test was most appropriate for the data, compared to other correlation algorithms as the data was ranked and the flow cytometry marker frequencies were not normally distributed.

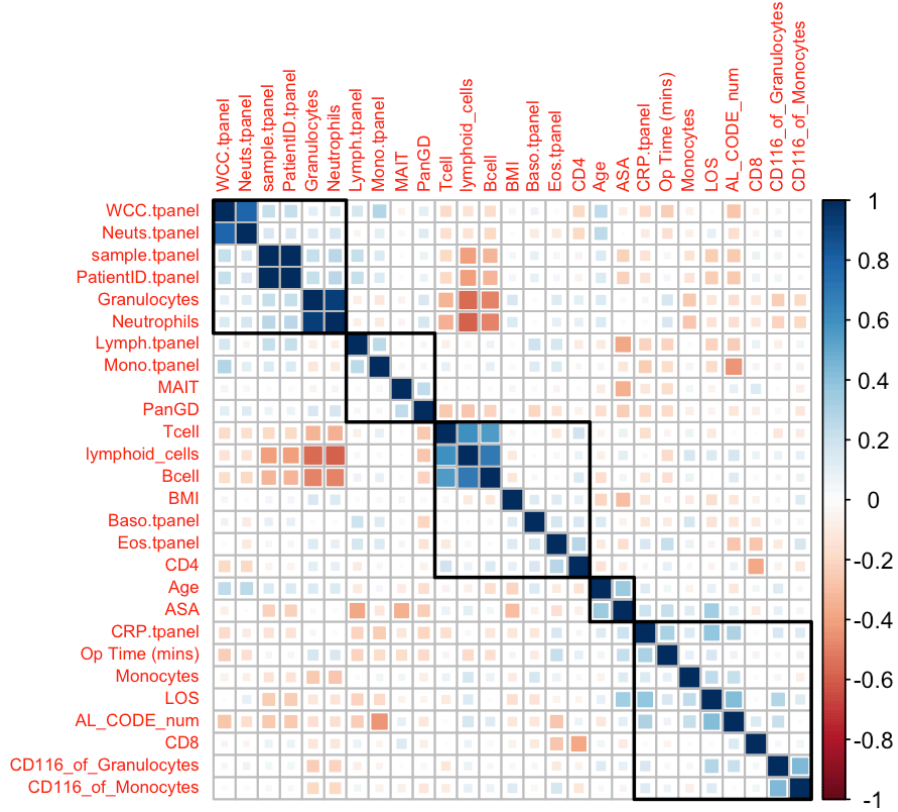


Figure 1.6: Correlogram of peritoneal fluid outputs with blood markers and patient demographics.

Pairwise Kendall's τ correlations are shown between clinical, demographic and flow cytometry-derived variables, with the size and colour intensity of each square indicating the strength and direction of the correlation. Variables cluster into four blocks (black outlines), broadly corresponding to white cell and granulocyte counts, lymphocyte-panel counts, cell-population frequencies, and clinical or operative measures. The lymphoid and granulocyte/neutrophil measures show a strong negative correlation with one another, consistent with these representing complementary proportions of the same leucocyte population.

The correlogram of flow cytometry data and clinical metadata provides insight into potential candidates for further investigation, as factors that associate with clinical complications have the potential to be a biomarker. Most of the blood markers correlate with each other independently, including the granulocyte, neutrophils, and white blood cell count. CRP strongly correlates with those that develop AL, Age,

ASA and length of stay (LOS). These factors are already known to correlate with AL, and CRP is often used in current clinical procedure as a marker for complications. Age is associated with worse outcomes and so is also expected to correlate with complications. Length of stay correlates with complications.

The CRP measurement associating with the clinical outcomes suggests that the cohort here is reflective of previous literature and observations. This measurement can be used as a proxy for the likelihood of AL, as it is so closely correlated – this can help with comparisons with continuous data rather than categorical data.

1.3.5 Supervised analysis

Of the peritoneal fluid markers, relative frequency of monocytes and CD8 T cells has the highest correlation with CRP and with the AL related observations of ‘AL_CODE’, Age, and length of stay.

First, the frequency of monocytes was examined, as it had the highest correlation to clinical complications (Figure 1.7). Visually, the data shows a higher mean for the complications and AL group, compared to the group with no complications. In this data, the group with the complications has the higher mean compared to the AL group. However, with this smaller sample size, it is not possible to determine if these means would change with more data collected for individuals with AL.

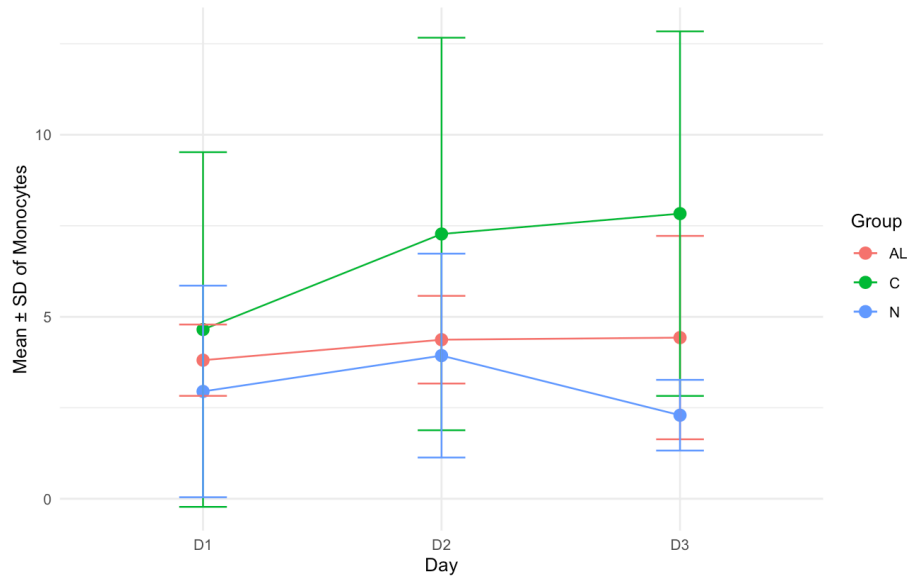


Figure 1.7: Mean frequency of monocytes, grouped by perioperative complication.

Grouped plot of mean percentage of monocytes per each perioperative complication group, with standard error bars for each timepoint per group. AL = anastomotic leak (n=3), C = complications (n=6), N = no complications (n=31). Monocyte frequency was broadly similar across the AL and no-complications groups at all three timepoints, while the complications group trended towards a higher mean frequency by day 3, although the wide and overlapping error bars indicate this difference is not well resolved with the available sample sizes.

The next factor of the peritoneal fluid data associating with clinical complications is the relative frequency of CD8 (Figure 1.8). Similar to the frequency of monocytes (Figure 1.7), the overlap of standard deviation for each of the groups makes it difficult to determine if CD8 would be a good marker for further investigation. With the current data available, it is not possible to determine if there is a unique correlation with the AL group with the frequency of CD8 T cells. CD8 T frequency appears to drop in the AL group compared to the other groups. Due to the sample size of the AL group it is difficult to determine if this pattern is applicable to a broader population. The ability of CD8 T cells to be measured, and indicating a difference at day 2, would make this an interesting marker for further work.

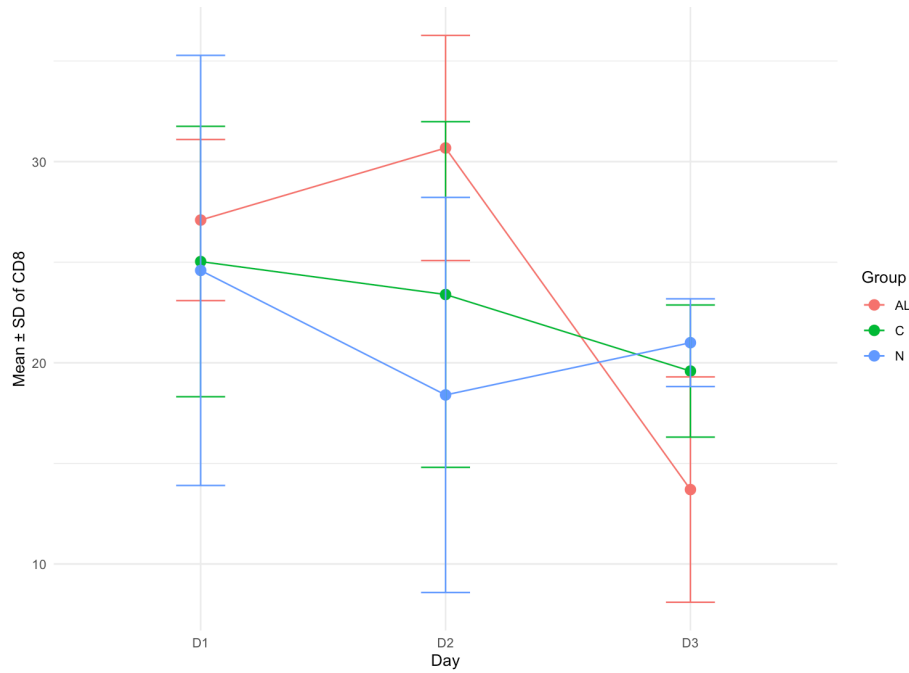


Figure 1.8: Grouped mean plot – CD8

Grouped mean plot of percentage of CD8⁺ population per each perioperative complication group, with standard error bars for each timepoint per group. AL = anastomotic leak (n=3), C = complications (n=6), N = no complications (n=31). The AL group shows a distinct trajectory relative to the other two groups, rising to the highest mean CD8⁺ frequency of all groups at day 2 before falling sharply to the lowest value by day 3, whereas the complications and no-complications groups both decline more gradually and remain closer together throughout.

1.3.6 Unsupervised analysis

The previous sections discuss the potential uses of manual gating. Another way of analysis is by clustering the data which allows identification of cell subsets without predefined gating strategy.

This type of analysis works best with state markers, and with a large panel in which many different subsets can be identified within the dataset. In this case, the panels are narrowed to describe very specific cell subsets, so it is not possible to find subsets that were undefined. The full list of immune subsets are spread across two

separate panels, and must be analysed separately in this particular dataset. Adding another method of analysis to the data can also be a test for robustness; if the results agree, then the data is robust to provide similar results regardless of analysis strategy.

1.3.7 Preparation of flow cytometry data for unsupervised clustering

Before clustering the data, it must first be transformed and normalised, so that the differences between cells are by marker expression rather than if a particular marker is consistently of higher fluorescence. When using FlowJo, the raw data is automatically transformed. Transformation must be applied manually for this analysis.

The cofactor for each transformation can be adjusted for each parameter. FlowVS (Azad et al. 2016) was used to determine the best cofactors for hyperbolic arcsine (arcsinh) transformation of the data (Figure 1.9). FlowVS corrects for a phenomenon of higher fluorescent values within a population having higher variance. FlowVS calculates the best possible cofactor with the highest homogeneity of variance using Bartlett's test statistic. This is calculated on a representative subset of files, and applied across all the data. It is possible to clearly identify positive and negative values after transformation (Figure 1.9). Transformation is common practice, and crucial, in flow cytometry analysis.

In FlowJo and similar software, transformation is done automatically and uniformly across all parameters. By using this method, transformation can be applied to each parameter individually, providing clearer negative and positive peaks tailored to each expression profile of each parameter.

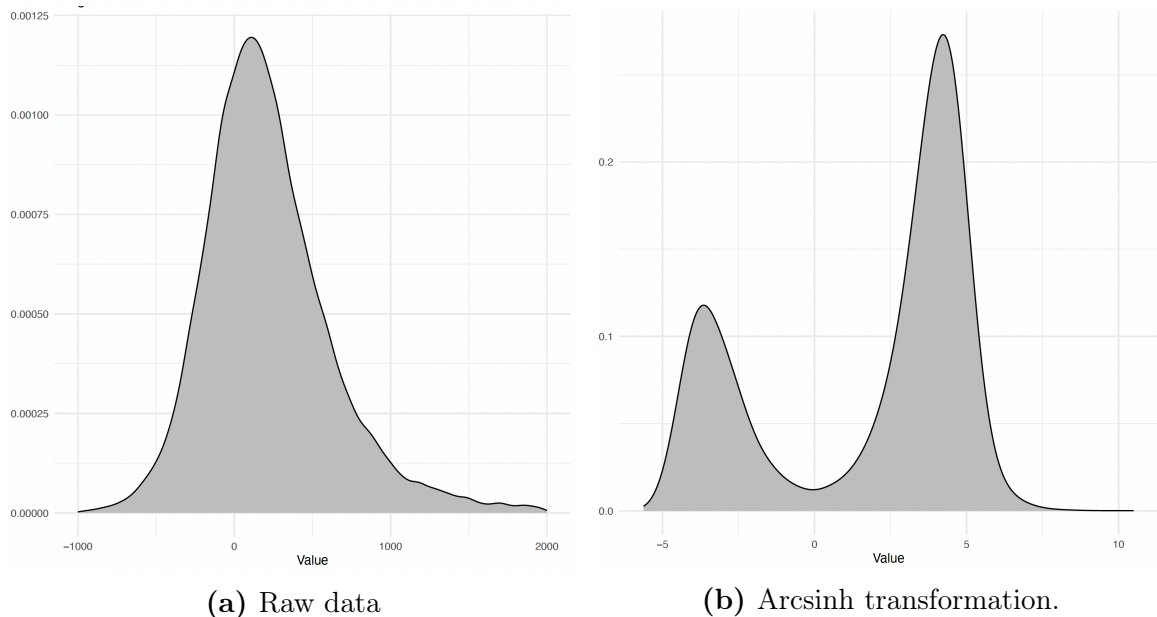


Figure 1.9: Arcsinh data transformation

Representative example of target distribution of intensity for each parameter, using cofactors calculated through FlowVS and arcsinh transformation. Transformation of the data is a required preprocessing step for flow cytometry data to be able to identify negative and positive peaks. Before transformation, the raw fluorescence intensity values are heavily right-skewed, with the bulk of events compressed near zero and a long tail extending to high values, making negative and positive populations difficult to distinguish on a linear scale.

It is possible to repeat the same sample across every batch to be able to measure technical effects in the data acquisition over time – this is referred to as an anchor sample. There are fewer options for normalising data without anchor samples, and is less likely to completely correct for batch effect as it is unknown to what extent the fluorescence intensities are attributed to technical effect.

There are some normalisation techniques that can reduce some technical effect. This data was normalised to reduce batch effect using CytoNorm (Van Gassen et al. 2020). The median fluorescence intensity of the positive and negative for each parameter was normalised to shift towards a similar pattern across all data – making

the data more comparable but retaining the difference in frequency between the samples. The assumptions are that fluorescence intensity changes are related to technical effect, and that there are similar phenotypes across all samples. As the samples are all from humans, with the same input material of peritoneal fluid, it can be assumed that the cell types discovered are comparable.

1.3.8 Factors correlating with highest variation in the data

The two separate panels (Figure 1.10 and Figure 1.11) should be considered separately. It is possible to combine panels that have been analysed on the same biological samples, but in this case, there are limited shared markers, and the samples have been treated differently (PBMC extraction for the T cell panel), and so this would likely introduce errors into the dataset that cannot be corrected.

In this study, the three most common dimensionality reduction techniques were applied. The goal is to find a dimensionality reduction technique that produces a graph that separates the cells into discrete clusters. It is then possible to determine if those clusters correlate with marker, cell type, and clinical metadata. The clustering is applied solely on the markers, with the clinical metadata applied afterwards; therefore making the analysis unsupervised and led by the data.

The three dimensionality reduction techniques were PCA (Principal Component Analysis), UMAP (Uniform Manifold Approximation and Projection) and MDS (MultiDimensional Scaling). In terms of single cell transcriptomics, UMAP is the most popular, and is known for preserving global structure of data (Becht et al. 2019)

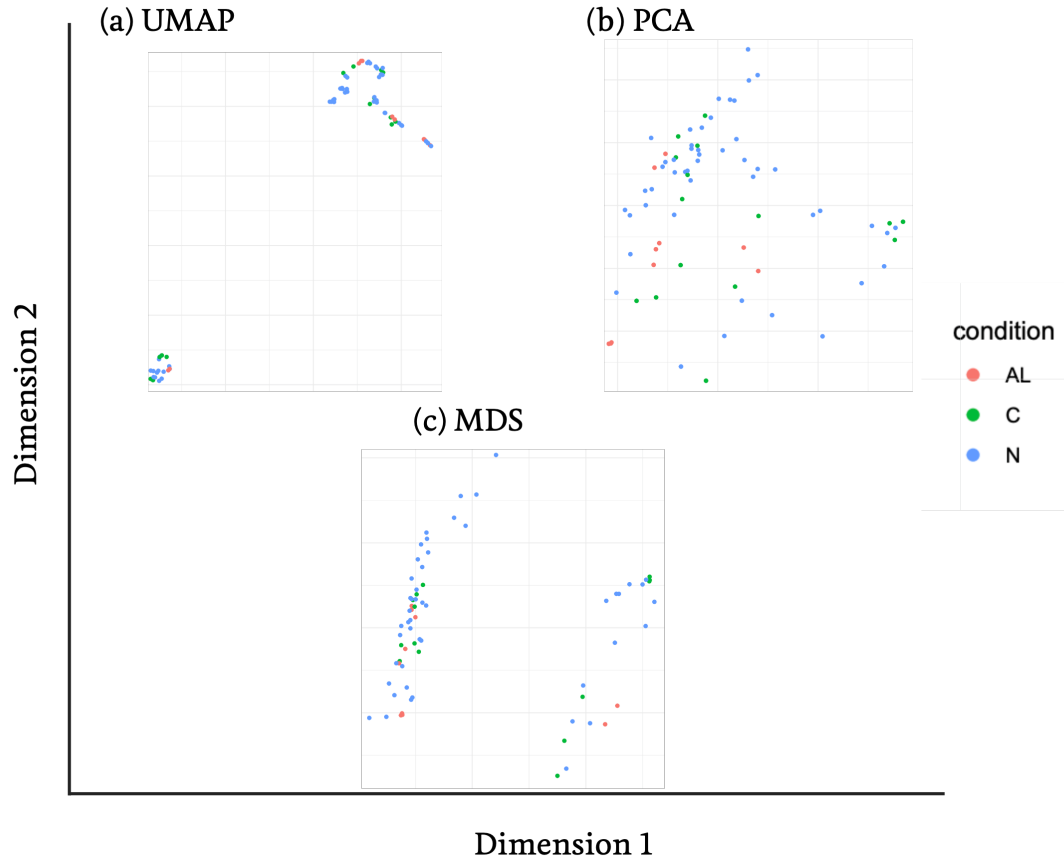


Figure 1.10: Dimensionality reduction of neutrophil panel samples.

Plots showing a visualisation of each dimensionality reduction technique, at a sample level, for the neutrophil panel data. colour corresponding to clinical endpoint (AL = anastomotic leak, C = other complications, N = no complications). No plots showed separation of samples by the clinical endpoint. Across UMAP, PCA and MDS, samples from the three clinical groups are interspersed throughout the projection rather than forming separate regions, and the small number of AL samples (red) fall within the same range of variation as the larger no-complications group (blue).

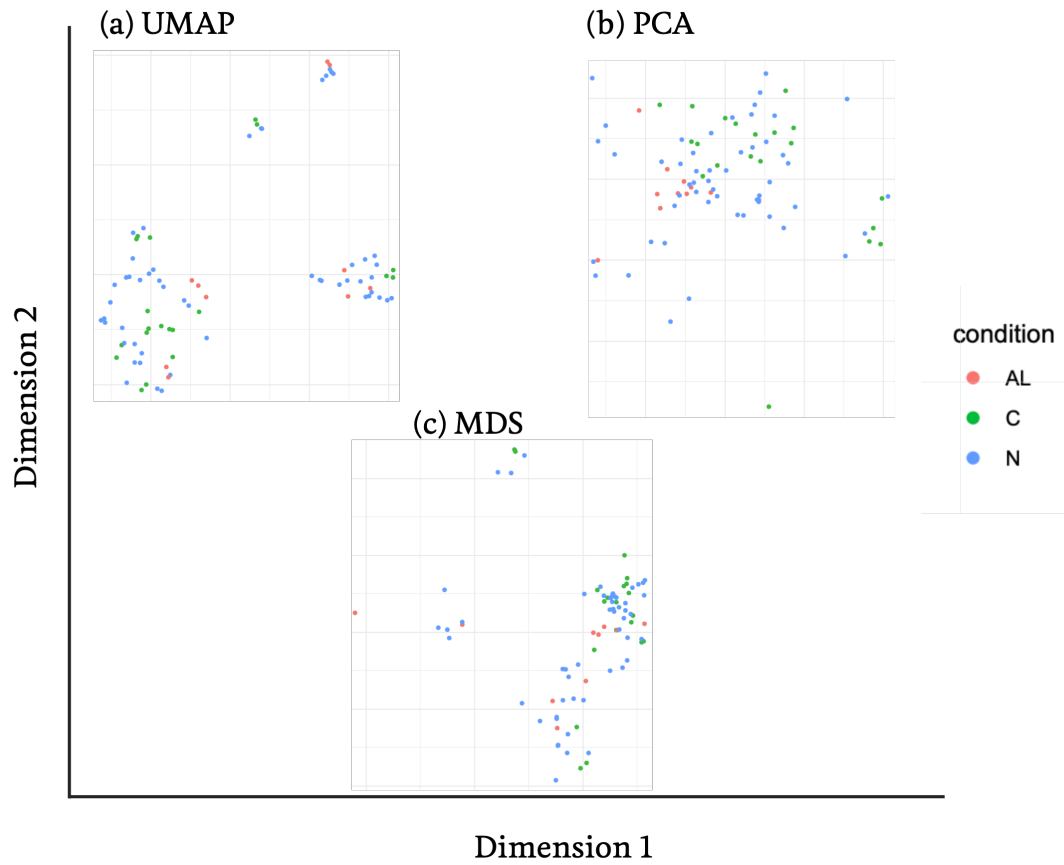


Figure 1.11: Dimensionality reduction of T cell panel samples.

Plots showing a visualisation of each dimensionality reduction technique, at a sample level, for the T cell panel data, with colour corresponding to clinical endpoint (AL = anastomotic leak, C = other complications, N = no complications). No plots showed separation of samples by the clinical endpoint. As with the neutrophil panel, samples from the three groups overlap substantially across all three ordination methods, with no evidence that clinical outcome drives a distinguishable transcriptional or phenotypic signature at the whole-sample level.

This analysis is conducted at the sample level, not the cell level, to more effectively measure any variation as determined by technical and biological effects. This is done by calculating the median expression for each marker, for each sample, and then conducting dimensionality reduction on this data to show the differences between

samples. This removes a lot of nuance from the data and so cannot be used to determine any specific changes in cell type frequencies. This method can, however, be used to map covariates to the data and see which factors cause closely organised cells.

MDS plots show more variation between the samples than PCA because it conducts pairwise comparisons, whereas PCA plots the highest variation. This is why, for per-sample comparisons, MDS plots can be more useful in detecting batch effect over PCA.

MDS worked best, in this cohort, in clustering the samples into discrete groups. Therefore, this reduced data was taken for further analysis, to see what clinical data correlates with the groups defined in the MDS plot.

The MDS plot allows correlation of variability within the data with:

- Technical issues during acquisition
- Clinical endpoint
- Other covariates; including age, sex, and BMI

Next, the covariates associated with these groupings can be evaluated, as found with the MDS plot (Figures 1.10(c), 1.11(c)). Using random forest modelling, it is possible to determine which covariates, as collected in the clinical and technical metadata, have the highest correlation with the variation in the data. This compared the two clusters identified by MDS plot, with the following model:

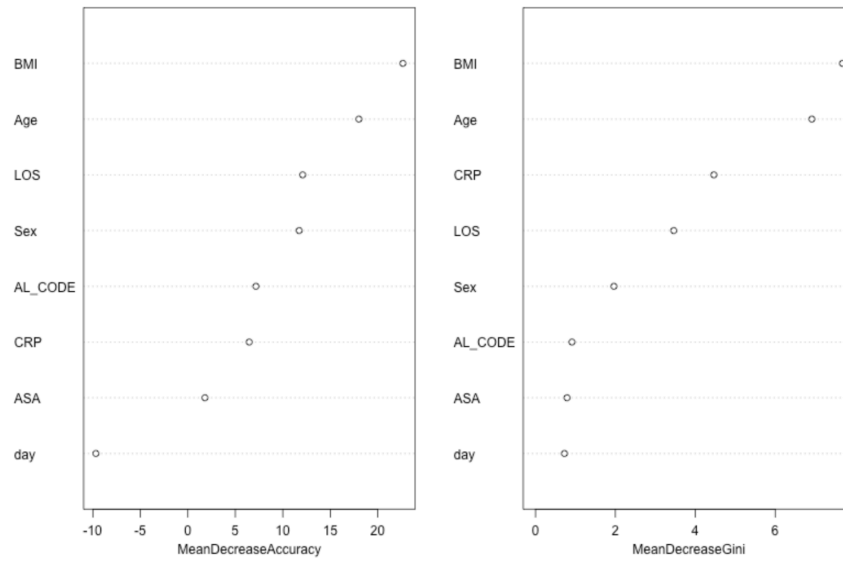
$$\text{Cluster} \sim \text{Sex} + \text{CRP} + \text{Age} + \text{AL_CODE} + \text{BMI} + \text{ASA} + \text{LOS} + \text{day}$$

The random forest model classifies the data using the covariate values, in relation

to one of the two clusters. The model can then remove one variable at a time, and run the model again to see how accurate the model is to the original with a modified or removed variable. The greater the effect of the removal of the variable, the more it is considered important. Showing importance of each variable, alongside the factors that explain the most variance, can be useful to identify singular factors of interest, removing less informative covariates (Figure 1.12).

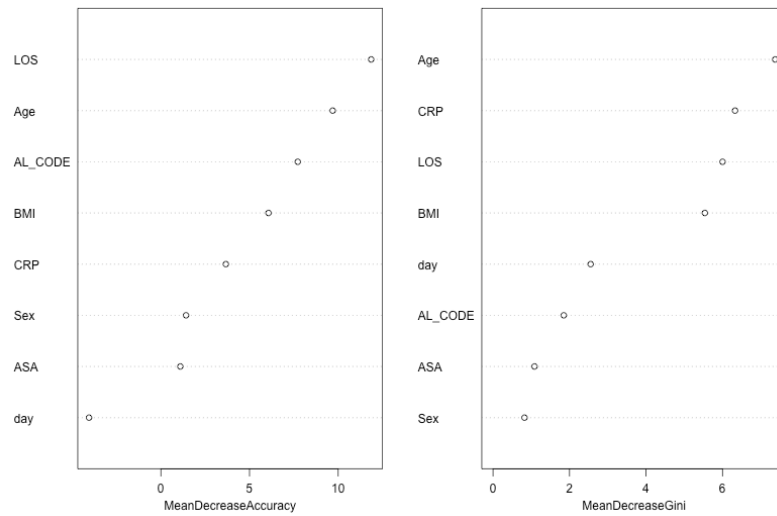
The mean decrease accuracy plot (Figure 1.12) measures the accuracy of the model if each variable was removed from the model. The Gini coefficient measures how much each variable helps to make useful splits in decision trees (reducing impurity) - therefore, the higher the mean decrease in Gini, the more useful that factor is for splits in decision trees, and therefore is considered more important.

Both measurements are useful in determining the variables that explain the most amount of variation within the dataset. In the analysis of both panels, BMI, age and length of stay are considered the most important variables. Other, hidden variables - including, but not limited to technical effects - can also be driving these changes to a greater degree, but the metadata to measure this is not available (as batches are the same as biological samples so it would split one-for-one). As discussed earlier with manual gating, length of stay is correlative, not causative. Age, BMI and sex are well known covariates that should be accounted for during analysis. In this case, it is not possible to correct for these covariates, as this would require an increase of sample size.



(a) Random forest model - Neutrophil panel

rf_model



(b) Random forest model - T cell panel.

Figure 1.12: Random forest model results for neutrophil and T cell panel

Random forest model, showing covariates that drive the difference between the two clusters as defined by the MDS plot. Each of the 8 most important factors in the data are shown on the y-axis. The left plot shows mean decrease accuracy, which is the global importance (if this factor is dropped, the model becomes less accurate). The right plot shows the mean decrease gini, which is how much the factor, when used to split a node in the decision tree, provides two groups that have most of the classes separated. BMI and age were consistently the most influential covariates by both measures, while clinical variables directly related to outcome, such as AL_CODE and CRP, ranked comparatively low, suggesting that the two clusters identified by MDS were more strongly driven by demographic differences between patients than by the perioperative complication itself.

1.3.9 Dimensionality reduction

Although there is no means to control for covariates as identified in the previous section (in particular, age and length of stay), it is still possible to investigate if there are any markers of interest in the flow cytometry data that stand out for further investigation in the future.

Different dimensionality reduction techniques were run to visualise which dimensionality reduction technique separates the data reasonably and into clusters of suspected cell types (Figures 1.13, 1.14).

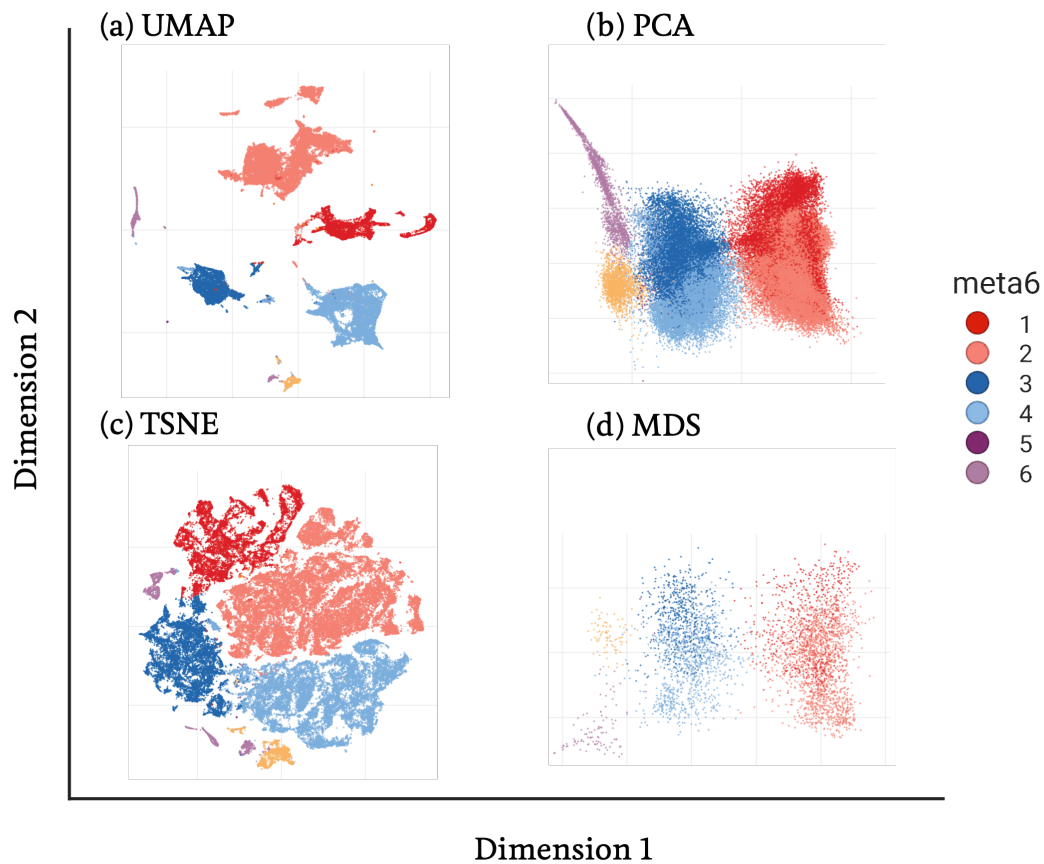


Figure 1.13: Dimensionality reduction: T cell panel

UMAP, TSNE, PCA, MDS dimensionality reduction for all cells within the dataset. Each sample was randomly downsampled to 1000 cells. Most informative are UMAP and TSNE. Six meta-clusters (coloured) were resolved with varying degrees of separation depending on the method: UMAP and t-SNE produced the most distinct, well-separated groupings, whereas PCA and MDS showed greater overlap between neighbouring clusters.

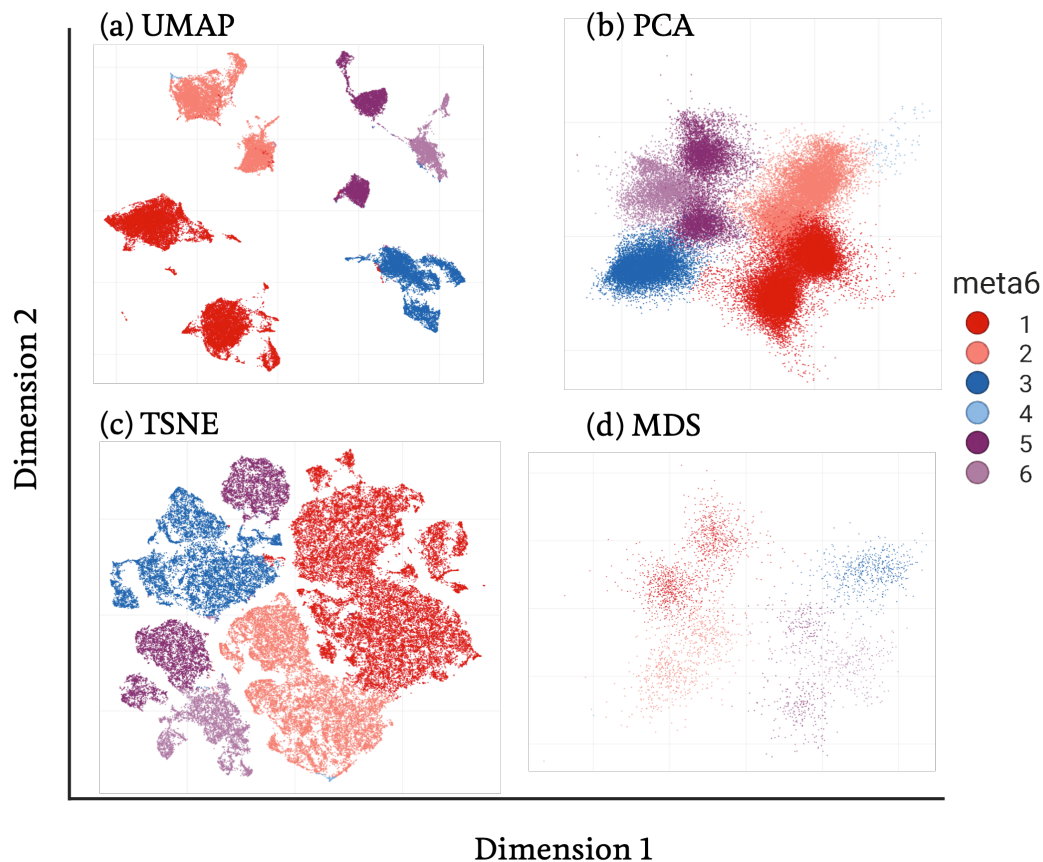


Figure 1.14: Dimensionality reduction: Neutrophil panel

UMAP, TSNE, PCA, MDS dimensionality reduction for all cells within the dataset. Each sample was randomly downsampled to 1000 cells. As with the T cell panel, UMAP and t-SNE gave the clearest separation of the six meta-clusters, while PCA and MDS showed more overlap between clusters, particularly among the smaller clusters.

MDS works particularly poorly at a single-cell level as it requires an extremely high amount of computing power, as it attempts to conduct pairwise analysis of every cell compared to every other cell - for this reason, fewer cells are shown in this plot (100 cells per sample). In all cases, the clustering is calculated using FlowSOM.

UMAP and TSNE worked well for dimensionality reduction corresponding with

the FlowSOM clustering – with UMAP being slightly easier to discern different clusters as they are more polarised compared to TSNE. Using the UMAP, there are a handful of clusters that align with FlowSOM clusters. After clustering, it is then possible to calculate relative abundance of each cluster, split by either by cell type, sample or by clinical complication group.

The best way to visualise which markers drive the expression profile of each cluster is with a heatmap of the T panel and neutrophil panel samples (Figures 1.15, 1.16). These show the relative marker expression for each FlowSOM cluster.

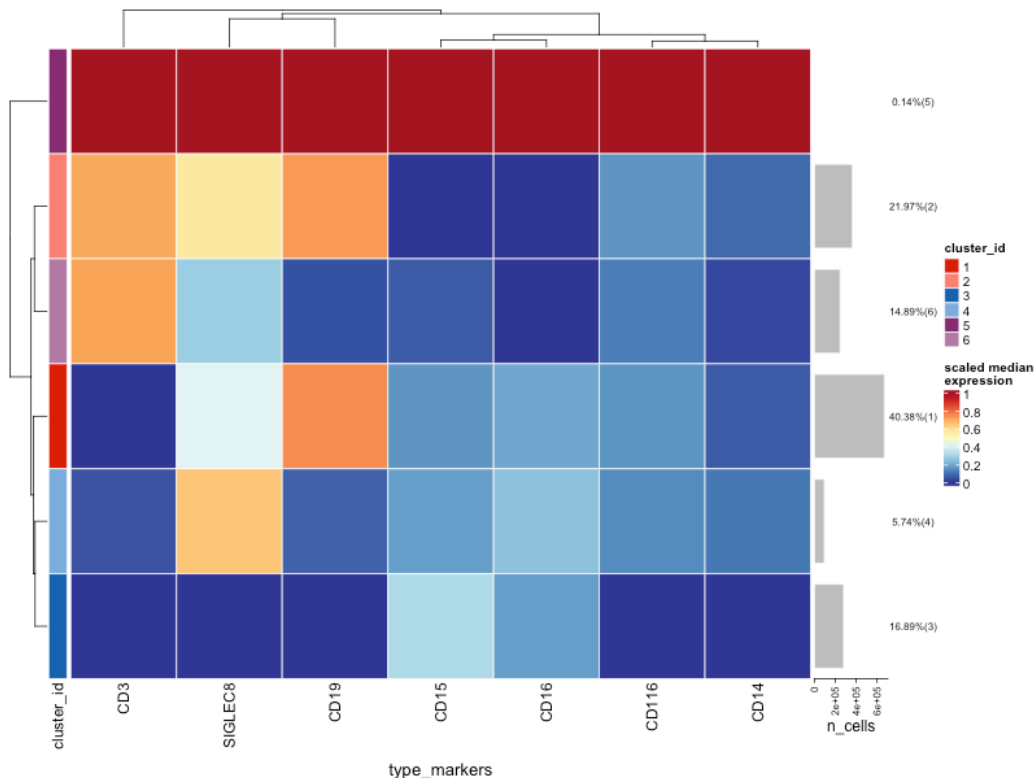


Figure 1.15: Heatmap of neutrophil panel meta-clusters

Heatmap showing the scaled median expression of each of the type markers in the neutrophil panel. The left side of the heatmap shows the cluster ID (the same ID as in Figure 1.14). The right side of the plot shows the number of cells that are in that cluster ID, as a percentage of all cells. The x-axis shows the cell type markers in the neutrophil panel, which should be used with the heatmap to define distinct cell types for each of the clusters. The markers and cluster IDs are ordered in similarity to the other cell type makers and cluster IDs, similarity shown in dendrograms on the border of the heatmap. Cluster 1, the largest population at 40.38% of cells, was distinguished from the others chiefly by higher relative CD19 expression alongside low CD3, while clusters 2 and 6 together showed the highest CD3 and SIGLEC8 expression, consistent with these six clusters resolving several major leucocyte lineages rather than neutrophils alone.

The first heatmap is for the neutrophil panel (Figure 1.15). This showed very little variation in the dataset overall, but some differences could be seen with this normalised heatmap. Cluster five represents 0.14% of the data and appears to have high expression in all markers. This is very likely to be debris, but only represents a small proportion of the data. 5.74% of the data is SIGLEC⁺ only, which could be

eosinophils – though CD15 would also be expected to be high if that were the case.

The two clusters which have a seem the most logical, in terms of their relative abundance and with the markers they express, are cluster 1 for CD19⁺ B cells and cluster 6 for CD3⁺ T cells. The remaining clusters do not have an expression profile that connects with any known cell subset.

The next heatmap is of the T cell panel (Figure 1.16). Compared to the neutrophil panel, this showed high variation between the clusters and some unique cell types could be identified. The greater variance in this dataset compared to the neutrophil panel is likely due to higher quality data acquisition and a higher number of immune cells analysed due to the enrichment of PBMCs with gradient separation (a processing step that was not possible for the neutrophil panel samples).

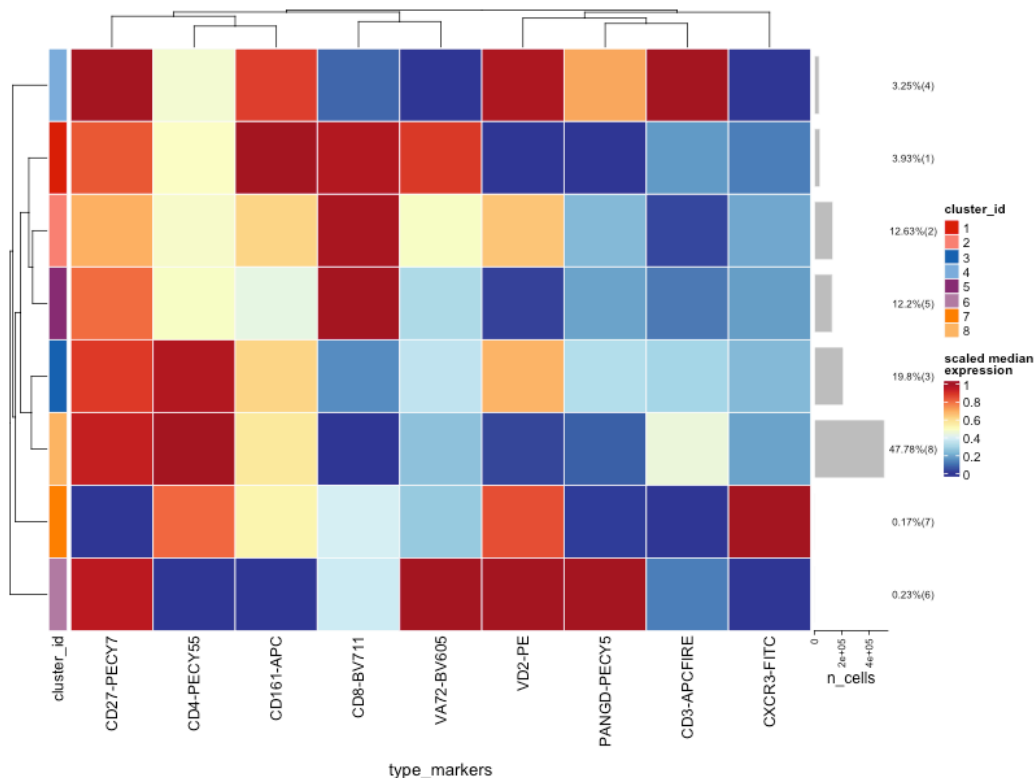


Figure 1.16: Heatmap of T cell panel meta-clusters

Heatmap showing each cluster with the relative expression of each marker. The right side of the plot shows the number of cells that are in that cluster ID, as a percentage of all cells. The x-axis shows the cell type markers in the T cell panel, which should be used with the heatmap to define distinct cell types for each of the clusters. The markers and cluster IDs are ordered in similarity to the other cell type makers and cluster IDs, similarity shown in dendrograms on the border of the heatmap. Cluster 8, the largest population at 47.78% of cells, showed relatively high expression of CD27, CD4 and CD161, consistent with a conventional CD4⁺ T cell phenotype, while the smaller clusters were distinguished mainly by differential expression of CD8, Vδ2 and PanGD, reflecting the panel's design to resolve both conventional and $\gamma\delta$ T cell populations.

This heatmap shows some potentially interesting markers, that map to immune cell subsets to be expected. However, for the more rare subsets, the number is too low to infer any statistical meaning between the groups. The most abundant cell cluster in this dataset is cluster 8 and cluster 3 representing a combined 68% of the cells. This contains CD4⁺ CD27⁺ cells. Both CD27 and CD3 appears to be of

consistent expression in nearly all clusters, and so indicates that these markers do not have variance across the dataset and cannot be used to effectively define cell types. Cluster 5 and 2 represents a combined 25% of the dataset and shows CD8⁺ (though low CD3 expression as with all clusters). Cluster 1 shares markers associated with MAIT cell markers and CD8, which is not a defined cell type. Cluster 4 has a relatively high expression of CD3, V δ 2, CD161 and CD26. This cluster could be a $\gamma\delta$ T cell subset, although potentially also having somewhat high CD4 expression which would not be compatible with the definition of $\gamma\delta$ T cells. Cluster 7 is defined by equally high V δ 2 expression, not compatible with current understanding of $\gamma\delta$ T cells. Finally, cluster 6 has both V δ 2 and V α 7.2 markers, which again is not compatible with current definitions of immune cell types. Overall, it is not possible to be confident with these findings with just this dataset.

The dataset with the highest variance in the data was the T cell panel data. This also had more observations with more cells being recorded and higher quality data acquisition. The most compelling definition of a particular cluster was the definition of CD8 T cells as CD8⁺ in cluster 5 (Figure 1.16). Therefore, the next step was to compare the CD8⁺ cells defined with clustering against the cells that were defined using manual gating (Figure 1.17). This was done by comparing the relative frequency of CD8⁺ cells from both methods for each sample and applying a trend-line. The plot shows a flat-to-negative correlation, whereas a correlation would be expected if the two methods were comparable. This indicates that, as the methods are not correlative, the outputs from the two methods either by clustering or with manual gating are not comparable.

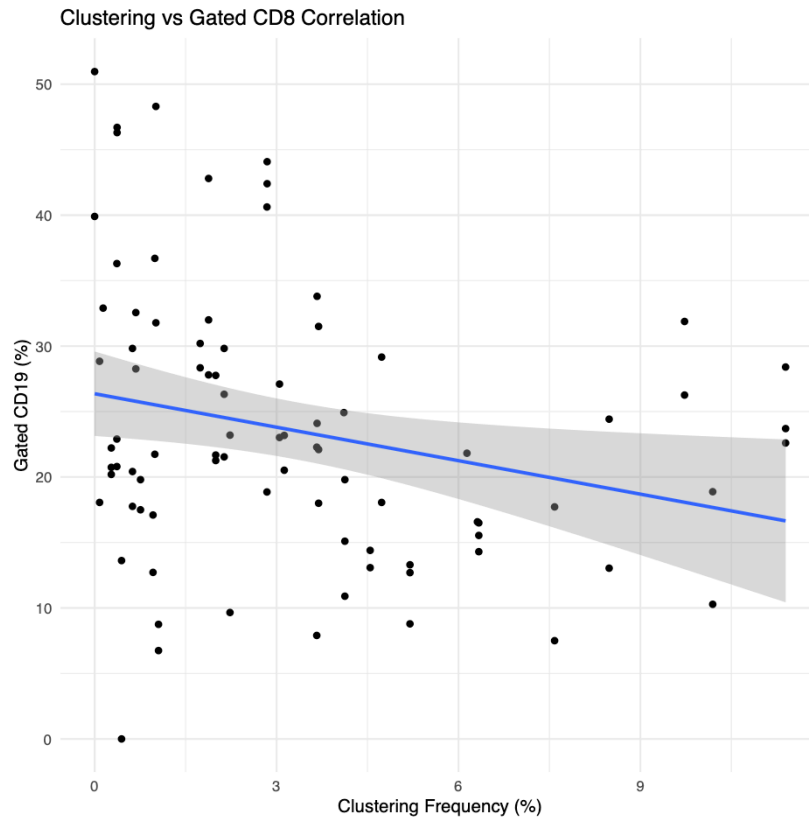


Figure 1.17: Values of clustering plotted against manual gating for identification of CD8 T cells, with a limma trend applied.

There appears to be a slight negative (-0.25) correlation between the two values. This indicates that the identification by clustering provided very different results compared to the manual gating. Further studies would be required to determine which method is most accurate for this particular study. CD8 resolved well in both manual gating and clustering individually. The scatter is widely spread around the fitted trend line, with several samples showing high manually gated CD8 frequencies but very low clustering frequencies, illustrating that the two approaches did not consistently agree on the size of this population even where each method's underlying data quality was acceptable.

1.4 Discussion

One of the aims of this chapter was to assess the use of peritoneal fluid as an input for immunophenotyping. Between both the manual gating and the unsupervised clustering, it was possible to detect immune cells. With a larger cohort and more complex panels, this could lead to an ‘immune atlas’ of cell types found in peritoneal fluid after colorectal resection. The other aim of this chapter was to determine if the relative abundance of immune cells could be calculated to ultimately differentiate between the group that develop anastomotic leak and the group with no complications. Despite using separate methods to define the cell types, it was not possible to define the anastomotic leak group using the data that had been collected. A larger cohort and improved technical quality control may provide a dataset that can differentiate between the two groups.

The analysis of peritoneal fluid in this study demonstrates that immune cell profiling is feasible even when statistical power is limited. Some differences were found within the clinical cohort, and cell subsets were extracted from the data. The robustness of the data was tested by applying multiple, independent analysis techniques to test what method provides more reliable and translatable findings. Technical effects and covariates which can confound the data could be identified and described. This is a common problem with many flow cytometry projects and being able to distinguish what is driving variance in the data is an incredibly useful quality check tool. As more sophisticated analysis become commonplace, and studies with spectral machines become more competitively priced and accessible, fully understanding the data and identifying any confounding factors is fast becoming a crucial part to the analysis workflow which must be considered.

Although these results show that technical issues were observed, this study establishes that peritoneal fluid analysis is measurable and exhibits biologically relevant variability across the cohort. Importantly, this represents the first application of peritoneal fluid immune profiling in perioperative recovery following colorectal resection. The markers identified here show promise for larger studies, particularly if data collection consistency and batch correction are improved.

To be able to discern if this effect was statistically significant between groups, if the method of data acquisition stayed the same, 246 individuals with AL would need to be recruited. In this study, there were 3 individuals that had AL. However, with improvements to methods, namely by reducing technical effect, it is hypothesised that the cohort size could be much lower while still being appropriately powered. Potentially, cell type frequency, clinical metadata and systemic measurements could be combined for a score that could be specific to AL patients. In this case, the potential cell types that could be used for this would be monocytes and CD8 T cells.

Visually, some samples looked drastically different between the groups – but this was not statistically significant, and the inability to isolate the technical and biological effect means that the true causes of the variability within groups cannot be further defined. Repeating the study with the same samples would elucidate if any improvements to the practical workflow would remove some variability within groups.

1.4.1 Flow cytometry analysis of peritoneal fluid

Phenotyping by flow cytometry is a common method applied across many clinical studies. Despite using the same methods, it is not usually possible for data to be combined across studies. This is because people use a variety of different immune

markers, different machines, and different sample preparation. Longitudinal studies are more likely to experience issues such as different operators, variation on sample preparation, machine laser drift, and different batches of fluorescent antibodies. This can mean that over the course of a project there can be measurable phenotyping differences within the data that is driven by technical effect.

There are a number of different methods that are often utilised across many studies to correct for batch effect, with most accessible through FlowJo. The introduction of these packages into FlowJo allows usage of these packages to a much broader audience of biologists who are more comfortable with a user-friendly graphical user interface, rather than using the command line interface. A popular package is FlowAI (Monaco et al. 2016), an R package and FlowJo plugin that can identify anomalous recorded events predominantly caused by changes in flow rate. The differences in expression over the course of a study in most cases is small, and implementing FlowAI is sufficient for extracting statistics by manual gating. Manual gating is the most common method of analysis for flow cytometry data – as this is often used to identify a specific, directed question, and gated flow cytometry plots can provide a visualisation of differences between groups that is easily interpretable.

If positive and negative populations are discrete, manual gating and adjusting gates can reduce some technical variation. Markers for flow cytometry can be defined as either ‘state’ or ‘type’, where type markers define a lineage and are often either present or absent from the cell surface, whereas state markers can be on a continuum of expression, and present across different cell types, such as memory related markers. The gating of state markers is much more difficult, therefore, as variation caused by technical effect along a continuum will not be possible to discern from any differences

in biological variation.

Correcting for most technical batch effect will depend on the experimental design. Many ways to reduce the influence of technical batch effect must be addressed before data acquisition. Many Cytometry Time-Of-Flight (CYTOF) and flow cytometry studies include an ‘anchor’ sample. This could be, for example, multiple aliquots of a PBMC sample from the same timepoint and individual. This can then be included in each subsequent batch with experiment samples. Any expression profile changes to the anchor samples across batches can be identified as technical batch effect, and so all other experiment sample expression can be adjusted to the same amount as required to make the expression profile of the anchor sample the same across batches.

Another method of batch correction would be to run calibration beads which have consistent fluorescent profiles within most common lasers and filters. Calibration beads have the same fluorescent intensity and relative abundance for each peak across entire studies, and so any changes to this profile over time can be defined as technical batch effect. The benefit of using calibration beads is that they are independent of antibody staining, and so can detect changes that are influenced by changes in staining method or due to different batches of antibodies from suppliers. They do not represent a typical expression profile of a cell, however, so having both calibration beads and an anchor sample of the same species would be advantageous to identify and correct for many different forms of technical batch effect. They have been used frequently across studies and have the practice is well documented (Perfetto et al. 2012).

There are ways to remove batch effect of many different types with published packages in R. Packages such as iMUBAC and CytoNorm rely on assumptions of

biological variability across batches, and can attempt to correct these expression profiles. These packages were initially benchmarked with CyTOF data but the methods can be applied to flow cytometry data (Ogishi et al. 2021). Although these methods work best with the addition of anchor samples, it can work without anchor samples. However, for the data in this analysis, each batch corresponds to a single individual which makes the normalisation much more challenging.

Clinical covariates – including ASA score, cancer stage and type, and surgical approach (laparoscopic vs. open) should be accounted for in future analyses. Systemic clinical data, such as CRP, serve as proxies for inflammation, which could be better leveraged in larger datasets.

Also shown here was an alternative to manual gating by implementing dimensionality reduction and clustering of data. This would remove the issue of gate placement of differentiating cells being subjective and difficult to reproduce. This method provides a more unsupervised analysis, being able to more confidently identify rare populations across the whole dataset.

Clustering data is not as easily interpretable as with manual gating, especially to an audience of biologists. Different dimensionality reduction techniques and clustering methods can present the data very differently, and finding the method that best represents the data is not always clear. Therefore, manual gating and unsupervised clustering have different strengths and the methods can be applied in parallel.

The usage of flow cytometry is widespread, but the application of bioinformatics techniques in this area is disparate. Most biologists analyse data with closed-source software such as FlowJo, and so there has not been a huge demand to improve the field with open-source software. A big jump in more pipelined workflows with open-source

data began in 2009 with an R package FlowCore (Hahne et al. 2009). This package became the base of most bioinformatics workflows as the initial way to manage and import flow cytometry files into R. However, not many significant changes to the workflow has occurred since then, with most packages published since then being focused on per-sample quality control metrics such as FlowAI. The idea of creating packages and workflows for importing, checking quality, producing statistics, and exporting the high quality trimmed data is not new, and has been very successful in other fields.

Flow cytometry data is difficult to standardise across projects. The main reason is that the markers measured vary so much between experiments, depending on the investigator's particular hypothesis, making it unlikely that data across different labs will be able to be combined. Another reason is that there is a significant amount of batch-effect within projects but also across projects. There are many different ways to control for batch effect, but this is not considered as standard and most projects do not include this. Considering the highly variable outputs across different experiments, there is a limit, due to experiment design, preventing much more information to be inferred.

Some improvements have been made to have flow cytometry data made more accessible for other researchers to interpret, such as FlowRepository (Spidlen et al. 2012), or more general repositories which includes some flow cytometry data, such as ImmPort (Bhattacharya et al. 2018). There have been many attempts for making flow cytometry data more accessible and collaborative, but the uptake by researchers is minimal. There needs to be a substantial change in perception towards making flow cytometry data public for these repositories to become a more popular source of

data.

In recent years, the capabilities of flow cytometry has greatly improved with spectral machines becoming more popular across institutes. It is now possible to have 40 colour panels whereas previously panels were limited to around 12 colours. It will be possible to have multiple studies in which the same panel is used; allowing very specific questions to be asked for the specific project, but also a broader picture of different immune states will be collected across projects. This provides an exciting foundation in which many different projects can be pulled together and analysed to get a broader understanding of immune states across many different conditions, and across many different projects.

1.4.2 Key findings in immune profiling

The most prominent and measurable changes in peritoneal fluid were observed in monocytes. Monocytes are present in the peritoneal fluid after acute infection because they are rapidly recruited from peripheral blood in response to locally produced chemokines and cytokines, as documented in peritoneal dialysis patients (Zhou et al. 2015). Their accumulation reflects an early immune response, as monocytes migrate to the peritoneal cavity to differentiate into effector cells that mediate pathogen clearance and tissue repair (Shi & Pamer 2011). A rise in monocytes on postoperative day 2 could provide valuable clinical information, though this requires validation. It could be possible, that with the addition of another marker such as monocyte frequency combined with CRP level, AL could be more accurately defined at early stages than is currently possible, which is around day 2 to day 3 post operation.

CD8 T cells correlated with increasing CRP levels, but results were not robust

due to inconsistencies between manual gating and clustering. Nevertheless, CD8 remains a candidate marker for future investigation. CD8 T cells may be among the earliest responders at sites of bacterial or surgical insult because they are abundant, rapidly recruited, and detectable using flow cytometry from day 1. Alternatively, their recruitment could reflect localised antigen presentation and cytotoxic activity at the infection site. Future studies could test whether CD8 levels correlate with early infection markers, abundance, or tissue-specific immune activation.

1.4.3 Context of peritoneal fluid analysis in other diseases

The use of peritoneal fluid has been studied in other diseases. In a clinical context, peritoneal fluid can be analysed to monitor and detect for bacterial infection of peritoneal dialysis patients, along with many other measurements (Colombo et al. 2024). Phenotyping of peritoneal fluid in the clinic is much less common. Immune cell analysis in peritoneal fluid has shown that monocytes and macrophages dominate inflammatory responses, while T cell populations vary with infection and dialysis duration. Such data suggest that peritoneal fluid can provide a window into immune activation that complements systemic measures, and has potential for providing more information in a clinical context.

Peritoneal fluid has been used to study endometriosis, providing insights into local immune responses in chronic inflammatory conditions. In endometriosis, peritoneal fluid contains elevated macrophages and altered differentiation of monocytes (Na et al. 2008). Analysis of immune cell subsets, including regulatory T cells, cytotoxic T cells, and monocytes, can inform disease mechanisms and guide potential immunomodulatory interventions.

1.4.4 Interpretation and clinical implications

Future studies could apply the same panel or broader panels to better characterise cell types, or even a cell atlas of expected immune cells found in peritoneal fluid across all patients during colorectal resection. Existing cell atlases from endometriosis patients could provide a comparative framework, or expected baselines. (Guo et al. 2020).

CRP remains a robust systemic marker but is non-specific and delayed. It may serve as a predictor when combined with peritoneal fluid markers, providing insight into local versus systemic inflammation. Expanding immune panels and including matched blood samples would strengthen baseline characterisation and interpretation.

Data robustness remains a concern. Manual gating likely removes debris more effectively than clustering but introduces subjectivity and reproducibility challenges. Future studies could optimise sample processing to detect innate immune cells such as MAIT and $\gamma\delta$ T cells and test correlations with postoperative complications.

1.4.5 Clinical cohort size and power calculations

There is limited research on the usage of peritoneal fluid in this context, especially with acute disease. This thesis analyses a cohort of 41 individuals of which 3 individuals developed AL. This pilot study can determine if it is possible to measure changes in peritoneal fluid and if the samples can be processed and analysed at scale, with reproducible results. After the pilot study, power calculations can provide a better estimate of required cohort size.

1.4.6 Sample Size Power Calculation

The current data does not allow identification of any statistically significant differences in relative abundance of immune cells between the groups. There are some markers that show more promise as a marker than others, such as monocytes. A sample size power calculation can provide an estimate for the number of individuals required to measure this difference between the groups and achieve statistical significance.

This is dependent on the quality of data collection and processing staying the same as was done in this study. The focus of this analysis is the relative frequency of monocytes at day 2, as this had the highest correlation with the clinical data suggesting it is most associated with perioperative complications. The hypothesis is that there is a measurable difference in immune cell abundance at day 2, so the power calculation was made with the day 2 samples.

A power calculation was run with an alpha (Type I error) rate of 0.05, and a power of 80% (i.e. a Type II error rate, β , of 20%), for the chance of being able to detect a difference between groups if that exists, reducing the chance of a false negative. With the frequency of monocytes as a reference the estimate is 246 patients in the anastomotic leak group to show a statistically significant difference between the groups. In practice, it is likely that fewer patients would be required if the consistency between data collection, processing and analysis could be improved.

Power calculation parameters

Table 1.5: Statistical Parameters for Sample Size Calculation

Parameter	Value
Population mean (μ_0)	3.9
Observed AL mean (μ_1)	4.4
Difference to detect ($\delta = \mu_1 - \mu_0$)	0.5
Population standard deviation (σ)	2.8
Significance level (α)	0.05 (two-sided)
Power ($1 - \beta$)	0.80

Power calculation formula

For a one-sample two-sided t -test (approximated by normal distribution):

$$n = \left(\frac{(z_{1-\alpha/2} + z_{1-\beta}) \cdot \sigma}{\delta} \right)^2$$

This calculation shows that 246 individuals would need to be recruited in the AL group to reach statistical significance, if the difference in frequency of monocytes between the groups stays consistent.

1.4.7 CRP is an accurate predictor of AL

CRP has the best correlation with complications, and the CRP value of the AL group is visually distinct at day 2. CRP is able to define the AL group (Figure 1.18), both in this cohort and more generally in clinical practice. This dataset has clinical observations associated that are already recorded in as part of standard practice. These other observations, combined with peritoneal fluid analysis, could improve identification of patients with AL earlier than just CRP on its own. This combination

score did not improve specificity in this case, and could be due to technical effect, small cohort size, or small effect size. Further investigation with a larger cohort could improve this method of classification, or with the addition of other peritoneal associated markers such as neutrophil/lymphocyte ratio. The CRP measurement cannot be used to specifically identify the AL group at day 1, but could be possible with a combined score. Day 1 postoperative classification of AL with high specificity would allow for earlier treatment and could improve recovery reduce AL associated morbidity.

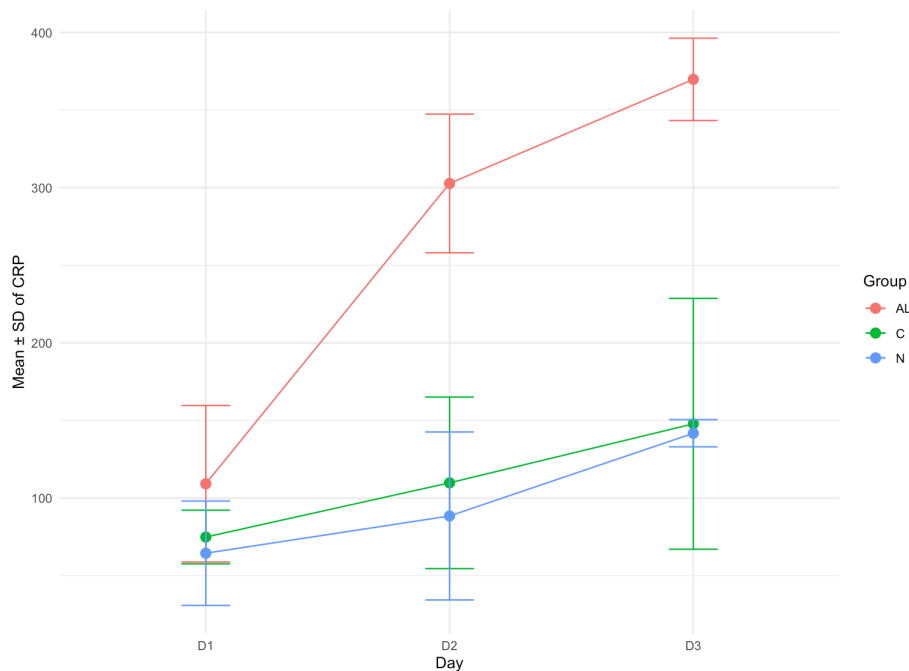


Figure 1.18: Grouped mean plot – Blood CRP

Grouped mean plot of CRP levels in peripheral blood of each complication group, split by timepoint. AL = anastomotic leak (n=3), C = complications (n=6), N = no complications (n=31). This data was collected as part of routine patient monitoring. The AL group increases CRP level rapidly after day 2 compared to the other groups. By day 3, the AL group's mean CRP had reached approximately 370, roughly double that of the complications and no-complications groups, which remained close together across all three timepoints; the wide error bars in the AL group, however, reflect the small sample size (n=3) underlying this trend.

1.4.8 Future direction

The data collected allowed for the identification of cell types that were present in peritoneal fluid, and changes of relative abundance between patients could be measured. This analysis sets a benchmark for future work, with an expected required number of patients for the study and expected immune cell subsets. Future work should focus on improving batch correction, expanding panel breadth, integrating systemic markers, and accounting for clinical covariates. Larger, optimised studies could confirm the utility of peritoneal fluid profiling for early detection of infection or postoperative complications, and help develop composite predictive scores combining local and systemic markers.

The manual gating and batch-effect correction approaches used in this study, together with consensus from recent literature, informed the proposed workflow (Figure 1.19). This is a similar workflow to Spasic et al. (2024), and Den Braanker et al. (2021), both published recently and with a focus on ~ 40 colour panel spectral data. Establishing a pipeline for spectral data is important, larger panels with as many as 40 colours cannot be easily presented in gating. Entire datasets can be presented with clustering the data, and can also identify biological and technical effect in the data. Spasic et al. (2024) conduct batch correction of the data using the **Harmony** algorithm, with no anchor samples. They then use this data to suggest that this can be used to identify cell types that are uniquely present or absent in specific diseases, such as chronic lymphoblastic leukemia. However, the reproducibility and specificity is questionable, as it would not be possible to be confident in the identification of cell types without appropriate anchor samples. An experimental design with more controls could have allowed for more direct comparison of the different conditions, and

therefore greater confidence in the results. This indicates the need for standardisation and validation of data analysis pipelines when considering analysing flow cytometry data beyond traditional gating methods.

Harmony is a very popular batch correction algorithm, initially validated in single cell RNA sequencing data (Korsunsky et al. 2019). Since then, this has been applied to CyTOF data and flow cytometry data, with packages published data processing and analysis such as **CyCONDOR** (Kröger et al. 2024). This batch correction is more appropriate with more markers in the panel, and for correcting larger datasets, to reduce overcorrecting to a small dataset of samples that do not fully represent the population.

The translation of methods from scRNA to flow cytometry indicates the potential for other techniques to be applied to flow cytometry data. For example, the python-based **scVERSE** uses numerous machine learning techniques to detect, for instance, batch effect, differential abundance, and perturbation (Virshup et al. 2023). Some of these methods could be applied to flow cytometry data, but would likely prove most informative at a much larger scale (with more biological replicates, a complex panel, and millions of cells per sample). A larger cohort of 100+ individuals and a more complex panel would enable validation of these more advanced machine learning methods in the context of flow cytometry.

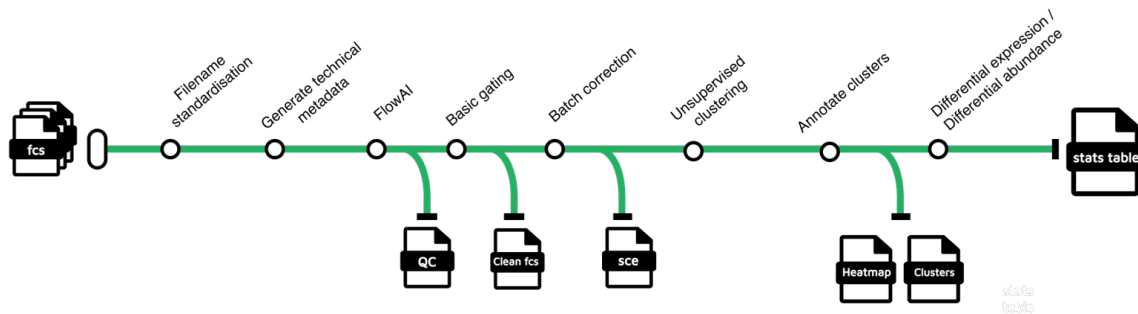


Figure 1.19: Preprocessing pipeline

Proposed pipeline for preprocessing flow cytometry files, based on consensus of recent literature (Den Braanker et al. 2021, Kröger et al. 2024, Spasic et al. 2024). The pipeline proceeds from raw .fcs files through filename standardisation and metadata generation, automated quality control with flowAI, basic gating and batch correction, before unsupervised clustering, cluster annotation and finally differential expression or abundance testing, with intermediate outputs retained at each stage for review.

CHAPTER 2

UNCONVENTIONAL T CELLS: RESPONSE TO ACUTE BACTERIAL INFECTION

2.1 Introduction

2.1.1 Unconventional T cells

The immune system is generalised into the adaptive and innate immune response. This consists of innate immune cells that are germline encoded and characterised by rapid response to antigens with pattern recognition receptors (PRRs), which are stimulated by conserved structures of bacteria, fungi and viruses. Innate immune cells are often highly specialised and localised to specific tissues, responding quickly to infection at barrier tissues. Innate immune cells include natural killer (NK) cells and neutrophils.

The adaptive immune system involves immune cells that respond to novel infections that would otherwise not stimulate an innate response. This includes T cells,

which undergo somatic recombination of their T cell receptors (TCRs) at the cell surface, allowing them to react to foreign peptides presented by antigen presenting cells (APCs) along with cytokine stimulation to aid clearance of novel pathogens.

T cells of the adaptive immune system can be further classified as ‘conventional’ and ‘unconventional’. Conventional T cells have CD4 or CD8 receptors and carry out somatic recombination of the α and β loci, producing a range of TCRs with broad reactivity to antigens. Unconventional T cells include $\gamma\delta$ T cells, which undergo somatic recombination of the γ and δ loci, producing TCR chains with a much fewer combinations and a smaller repertoire. $\gamma\delta$ T cells are often situated at tissue barriers, and respond to recognised pathogens with innate-like kinetics; much faster than conventional immune cells. This summary, and Table 2.1, provide a condensed overview of the concepts discussed in the reviews by Hu et al. (2023) and Pasman & Kasper (2017).

Table 2.1: Defining features of conventional, unconventional and innate immune cells.

Feature	Conventional lymphocytes	adaptive lymphocytes	Unconventional T cells	Innate immune cells
Representative cells	$\alpha\beta$ CD4 and CD8 T cells, B cells	$\gamma\delta$ T cells, MAIT cells, iNKT cells	NK cells, macrophages, neutrophils, dendritic cells, ILCs	
Receptor generation	Somatic V(D)J recombination	Somatic V(D)J recombination	Germline-encoded receptors	
Receptor diversity	Very high	Intermediate; often invariant	Limited	
Antigen recognition	Highly specific antigens	Conserved microbial products, lipids, metabolites, stress signals	PAMPs, DAMPs, cytokines	
Memory	Robust classical memory	None	None	
Response time to immune challenge	Relatively slower (days)	Rapid	Rapid	
Dependence on antigen presentation	Dependent	Independent	Independent	
Express TCR	Yes	Yes	No	

Summary of differences between conventional, unconventional and innate immune cells. Information summarised from Pasman & Kasper (2017).

2.1.2 $\gamma\delta$ T cells facilitate immune response to bacterial infection

A distinct, innate-like function of V γ 9/V δ 2 T cells is their ability to recognise (E)-4-hydroxy-3-methyl-but-2-enyl pyrophosphate (HMB-PP), a metabolite produced by many microbial pathogens (Carding & Egan 2002). This recognition is assisted by expression of butyrophilins (BTN) – specifically BTN3A1 and BTN2A1, which are expressed on the cell surface of human V γ 9/V δ 2 T cells along with other blood lymphocytes (Rhodes et al. 2016). Butyrophilins can be considered as regulators of $\gamma\delta$ T cell response through indirect or direct binding of the TCR. This response to HMB-PP by V γ 9/V δ 2 T cells make them potent and fast-acting in protection against most Gram-negative bacteria, and some Gram-positive bacteria including *Corynebacterium*, *Mycobacterium* and *Clostridium* (Kaminski et al. 2021).

V γ 9/V δ 2 T cells with a T follicular helper (Tfh)-like function are present in secondary lymphoid tissues and have been shown to provide B cell help. V γ 9/V δ 2 undergo HMB-PP dependent elevated IL-21R and CXCL13 expression (Vermijlen et al. 2007). This characteristic $\gamma\delta$ Tfh-like function promotes B cells recruitment to lymph node follicles assisting in B cell maturation (Bansal et al. 2012). Tissue specificity provided by the humoral environment allows $\gamma\delta$ T cells to perform their various functions at the appropriate time. $\gamma\delta$ T cells are often found to be crucial to mounting an initial immune response, particularly with bacterial infection.

2.1.3 The role of $\gamma\delta$ T cells in inflammation

$\gamma\delta$ T cells have been found to take part in many roles, including regulation of inflammation and early-stage immune response (Figure 2.1)

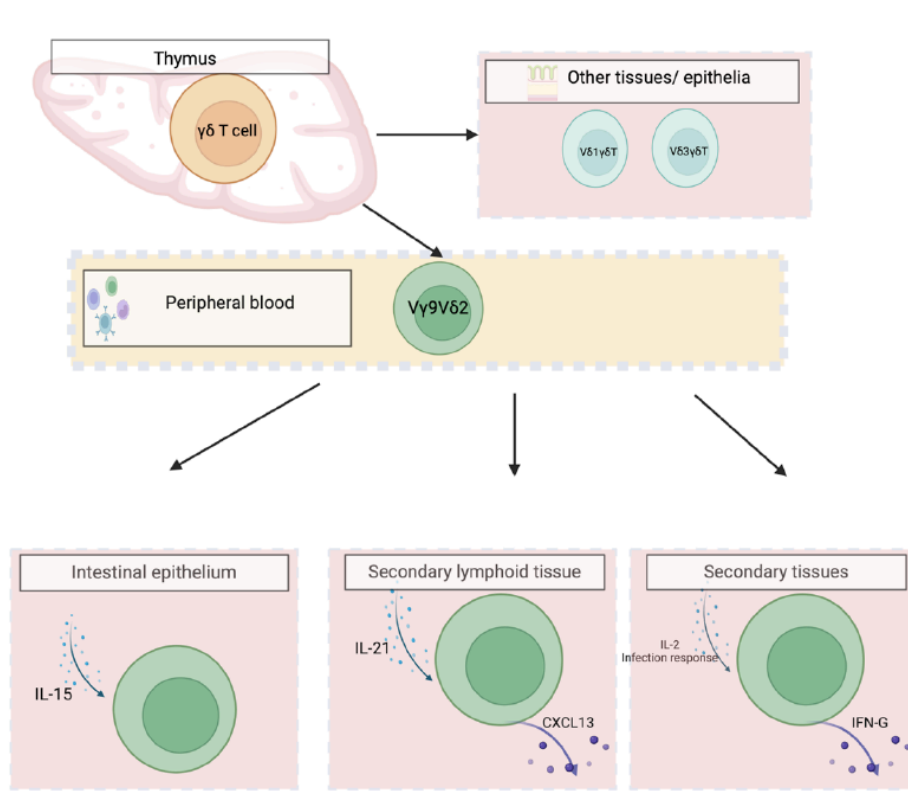


Figure 2.1: Tissue specific honing of $\gamma\delta$ T cells

The tissue and role specificity of $\gamma\delta$ T cell subsets. Schematic showing the effect of cytokines for homing of $\gamma\delta$ T cells for tissue and role specificity. $\gamma\delta$ T cells develop in the thymus before migrating to peripheral blood as V γ 9V δ 2 cells or homing directly to epithelial tissues as V δ 1 or V δ 3 subsets. From peripheral blood, cytokine cues direct $\gamma\delta$ T cells to distinct functional niches: IL-15 promotes homing to the intestinal epithelium, IL-21 supports a CXCL13-secreting phenotype in secondary lymphoid tissue, and IL-2, in the context of infection, drives an IFN- γ -secreting response in secondary tissues.

IL-2, a T cell growth factor, rescues the function of V γ 9/V δ 2 T cells in chronic inflammation (Costanzo et al. 2015, Fay et al. 2016), indicating a role of IL-2 in a pro-inflammatory environment to activate $\gamma\delta$ T cells in which IFN- γ production and macrophage recruitment enables the destruction of pathogens.

2.1.4 $\gamma\delta$ T cells at the intestinal barrier

IL-22, initially characterised as a Th1 cytokine, is central to epithelial barrier function, particularly in the intestine. Functions of IL-22 production in this environment include inducing antimicrobial peptides. CD4⁺ T cells have been documented to produce IL-22 and IL-17 when stimulated by innate lymphoid cells (ILCs), requiring an IL-6 microenvironment (Yang et al. 2020).

Previous work has sought to further understand mechanisms for IL-22 regulation. Microbe-responsive V γ 9/V δ 2 T cells were found to acquire APC function (V γ 9/V δ 2 T-APC) and stimulate CD4⁺ T cells in the absence of IL-6. This stimulation by V γ 9/V δ 2 T cells is also unique; stimulated CD4⁺ T cells did not promote IL-17 production. Whereas IL-2 and IL-15 each supported upregulation of a gut-tropic chemokine CCR9 expressing V γ 9/V δ 2 T cell phenotype, the closely related cytokines IL-7 and IL-21 failed to induce CCR9. IL-21 does induce APC markers of $\gamma\delta$ T cells, but IL-7 showed no significant APC induction. The polarisation of CD4⁺ T cells by V γ 9/V δ 2 T cells requires the presence of microbial ligand HMB-PP, which is produced by many bacteria associated with the intestine (Tyler et al. 2015).

Together, this supports the current understanding of the immunosurveillance role of $\gamma\delta$ T cells at the intestinal epithelium, and that $\gamma\delta$ T cells activated by bacterial pathogens can assume distinct effector functions depending on the microenvironment; including the ability to promote the polarisation of CD4⁺ T cell effector subsets.

2.1.5 Understanding the role of $\gamma\delta$ T cells in regulating inflammation

Research from the Eberl lab [Eberl, unpublished] suggests that IL-10 production is increased in naive and memory $CD4^+$ T cells in the presence of $\gamma\delta$ T cells in an HMB-PP dependent IL-21 microenvironment. IL-10 is prominent in the intestinal epithelium, and levels of IL-10 provide immunoregulation in a balance between tolerating intestinal commensal bacteria and inducing an inflammatory response to bacterial pathogens that cross the intestinal barrier.

With this, the aim was to further understand the gene expression profile of $V\gamma9/V\delta2$ APC T cells in an IL-21, IL-15, and IL-2 environment. This would enable inference of the potential cell surface protein expression and could inform a regulatory mechanism by which $V\gamma9/V\delta2$ APC T cells in an intestinal microenvironment encourage a $CD4^+$ effector phenotype.

2.1.6 Selecting conditions for investigation

IL-21, IL-15 and IL-2 conditions were selected as cytokine environments for this cell culture experiment. Previous research demonstrated that upon activation in the presence of IL-2 or IL-15, microbe-responsive human $V\gamma9/V\delta2$ T-cells acquire APC function and polarise $CD4^+$ T-cells towards expression of interferon (IFN)- γ and IL-22, respectively (Tyler et al. 2017).

IL-21 is related to IL-15, and high expression of IL-21 is correlated with inflammatory bowel disease (IBD) (Toskas et al. 2023). IL-21 production is associated with $CD4^+$ helper T cells, and functional outputs include the induction of NK cells, which are known to cause chronic inflammation in IBD patients (Yadav et al. 2011). In re-

lation to $\gamma\delta$ T cells, IL-21 has been shown to induce lower levels of $CD4^+$ T cell proliferation and lower levels of IFN- γ . IL-15 and IL-21 are two conditions related to modulating immune response at mucosal barrier of the intestine, and good candidates for further investigation on the characterisation of $\gamma\delta$ T cell activity during an inflammatory response in the intestine. Not only would this improve knowledge of the mechanism by which $\gamma\delta$ T cells are regulated, it could also inform what cell types would be expected during an initial acute response after anastomosis surgery.

The most appropriate assay for this investigation was transcriptomic analysis by RNA sequencing. This data can be used to identify candidate IL-21-dependent genes, including surface receptors and ligands, that could mediate regulatory crosstalk between V γ 9/V δ 2 T-APCs and $CD4^+$ T cells.

2.1.7 Aims

There are two main aims of this chapter. First, to determine the transcriptional changes induced in $\gamma\delta$ T cells by clinically relevant cytokine environments (IL-21, IL-15 and IL-2). Second, to identify candidate genes and pathways, with particular focus on the pro-inflammatory IL-21 environment, that may explain the mechanistic role of $\gamma\delta$ T cells in regulating $CD4^+$ T cell responses during acute bacterial infection. Differential gene expression analysis was used to identify up- and downregulated genes that may contribute to known immune-related pathways.

2.2 Methods

2.2.1 RNA sequencing

This work focused on the bioinformatic pipeline of sequencing data; there was no involvement in the data acquisition or experimental design of the cell culture experiment. Cell culture experiments were previously conducted by other members of the lab to investigate $\gamma\delta$ T cell mediated immune modulation. In these experiments, $\gamma\delta$ T cells were co-cultured with monocytes at a $\gamma\delta$ T cell:monocyte ratio of 1:10 and stimulated with 10 nM HMB-PP, either alone or in combination with 100 U/mL IL-2, 20 ng/mL IL-15, or 20 ng/mL IL-21. HMB-PP was used as a specific activator of $\gamma\delta$ T cells. Previous findings had shown that $\gamma\delta$ T cells activate monocytes. Based on this, the co-culture system was considered an appropriate and physiologically relevant model for studying $\gamma\delta$ interactions. After three days in culture, CD3⁺V δ 2⁺ T cells were selected by cell sorting using a FACS Aria II (BD Sciences), to purities >99.8%. RNeasy Micro Kit (Qiagen) and examined qualitatively and quantitatively using the TapeStation 2200 system (Agilent) and a Qubit fluorometer (Thermo Fisher Scientific). Samples were subjected to paired-end sequencing with 75 bp reads on a NextSeq 500 sequencer (Illumina), generating >50 million reads per sample.

2.2.2 RNA Sequencing data analysis

RNA sequencing data were processed using the Nextflow (Di Tommaso et al. 2017) nf-core (Ewels et al. 2020) RNASeq (version 3.14.0) pipeline, using the Hisat2 aligner, and removing ribosomal reads. The Binary Alignment Map (BAM) files produced from this pipeline were then used as input for featureCounts (Subread package, version 2.0.1), to obtain a count table. The reference genome used for alignment and

feature counting was the ENSEMBL (Martin et al. 2023) Homo sapiens GRCh38 (version 108). Each step in the pipeline was assessed for quality, per sample, using MultiQC (version 1.17) (Ewels et al. 2016). The count table was then imported into R version 4.3.1 (2023-06-16) using RStudio IDE version 2023.6.1.524 (Posit team 2023). Differential gene expression analysis was carried out using DESeq2 (version 1.42.0). Genes with an adjusted p-value less than 0.05 were considered differentially expressed. Differential expression visualisation was conducted using EnhancedVolcano (version 1.20.0). Volcano plots show a log2 fold change of 2.5 as a dashed reference line to highlight genes with a larger effect size, and genes exceeding this fold-change threshold, or of specific biological interest, were labelled by gene name. Fold change was therefore used to guide interpretation and highlight genes of interest, rather than as an additional statistical criterion for the differential expression results reported in the text and tables.

2.2.3 Data availability

The config files and settings for processing the raw RNA sequencing files to produce a count table is found here: <https://doi.org/10.5281/zenodo.12790718>. The analysis of the count table, including the differential analysis and plots, can be found here: <https://doi.org/10.5281/zenodo.16903807>. The raw fastq files can be found here: <https://www.ebi.ac.uk/biostudies/arrayexpress/studies/E-MTAB-13875>, and was used as part of this research article: <https://doi.org/10.1093/discim/kyae008>.

2.3 Results

2.3.1 Experiment design

The goal of this experiment was to determine if there were any genes differentially expressed in the selected cytokine environments. The conditions of interest were IL-2, IL-15, and IL-21, with an unstimulated condition as the control group. Three biological replicates were included for each condition (Table 2.2); there were no technical replicates. The samples went through library preparation and these samples were split across two independent sequencing runs.

Table 2.2: RNA sequencing sample metadata

Sample ID	Healthy donor ID	Condition
MERNA1	D1	IL-2
MERNA2	D1	IL-15
MERNA3	D1	IL-21
MERNA4	D1	IL-7
MERNA5	D1	Unstimulated
MERNA6	D2	IL-2
MERNA7	D2	IL-15
MERNA8	D2	IL-21
MERNA9	D2	IL-7
MERNA10	D2	Unstimulated
MERNA11	D3	IL-2
MERNA12	D3	IL-15
MERNA13	D3	IL-21
MERNA14	D3	IL-7
MERNA15	D3	Unstimulated

Metadata of sequencing samples: healthy donor ID and cytokine condition. D1, D2 and D3 corresponding with biological replicate IDs. Each condition in triplicate, for a total of 5 conditions and 3 biological replicates.

2.3.2 Quality control

Before analysis, the data was processed from raw sequencing files to a count table, which was then used for analysis. This includes the use of well-documented, standardised workflows that include quality checking measures to identify any potential issues with the data. The files directly from the sequencer are **fastq** files, which are just base-called sequences. Then, the data underwent transcript trimming, ribosomal read removal, and gene counting to create a count table. The series of packages

used are documented in the nf-core Nextflow pipeline that was used for this data (version 3.14.0), including BBSplit for ribosomal sequence removal and HiSAT2 for sequence alignment. A schematic of this pipeline can be found on the nf-core website (<https://nf-co.re/rnaseq/3.14.0>). This stage of data processing was done on the Cardiff University High-Performance Computer (HPC) cluster (Supercomputing Wales), as these steps are computationally intensive and this workflow can be conducted non-interactively. The pipeline generates reports which can be reviewed retrospectively to ensure all steps were completed successfully.

At each step, **MultiQC** was performed, to identify any issues with the files during the processing steps. MultiQC assesses the quality of the output files for each of the processing steps. Of note, MultiQC shows a warning across most samples for high GC content, high per-base sequence content, high duplication and high overrepresentation of sequences. The high GC bias is a common problem with sequencing data, as GC rich regions are more stable. This may indicate that some sequences are lost during library preparation, that were low in GC content. As this effect is across nearly all samples, this data is still comparable between these samples, as they were prepared in the same batch and the technical effect of that is likely to be uniform across the samples. Other potential issues with data include overrepresentation of sequences, which is further investigated in section 2.3.3.

After the processing steps, the genes were mapped to a reference human genome and the genes were counted using **FeatureCounts** package, to produce a count table. The count table is the final output from the pipeline, and can be loaded in R, in which analysis of group-wise differential expression was performed.

2.3.3 Sequencing saturation

The overrepresented sequences found during the MultiQC check can indicate that there was sequence bias during library preparation. To further investigate this, a sequence saturation estimate of the data was conducted (Figure 2.2). This was used on the fully mapped and feature counts table to run an estimate of how deep the sequencing was. This is done by random sampling of the data, to determine, of each random sample, how many genes are ‘new’. A new random sampling of genes was then generated for the remaining data and the ‘new’ genes were recorded. This was repeated for the entire dataset. The result of this was how many random sampling events before no more genes are recorded. This provides some certainty that all of the genes that were available in that sample have been found. In this data, a plateau was reached for all samples – indicating that the samples have been deeply sequenced. The sequencing was run twice, and both reached a plateau individually. Therefore, it is unlikely to change the results if the analysis is of just one sequencing run, or both sequencing runs combined. For the sake of completeness, both count tables for the sequencing runs were combined. As they were of the same library preparation, of the same samples, run on the same sequencer, and with very similar MultiQC reports, there was no problems detected by combining this data in the way of introducing any technical effect.

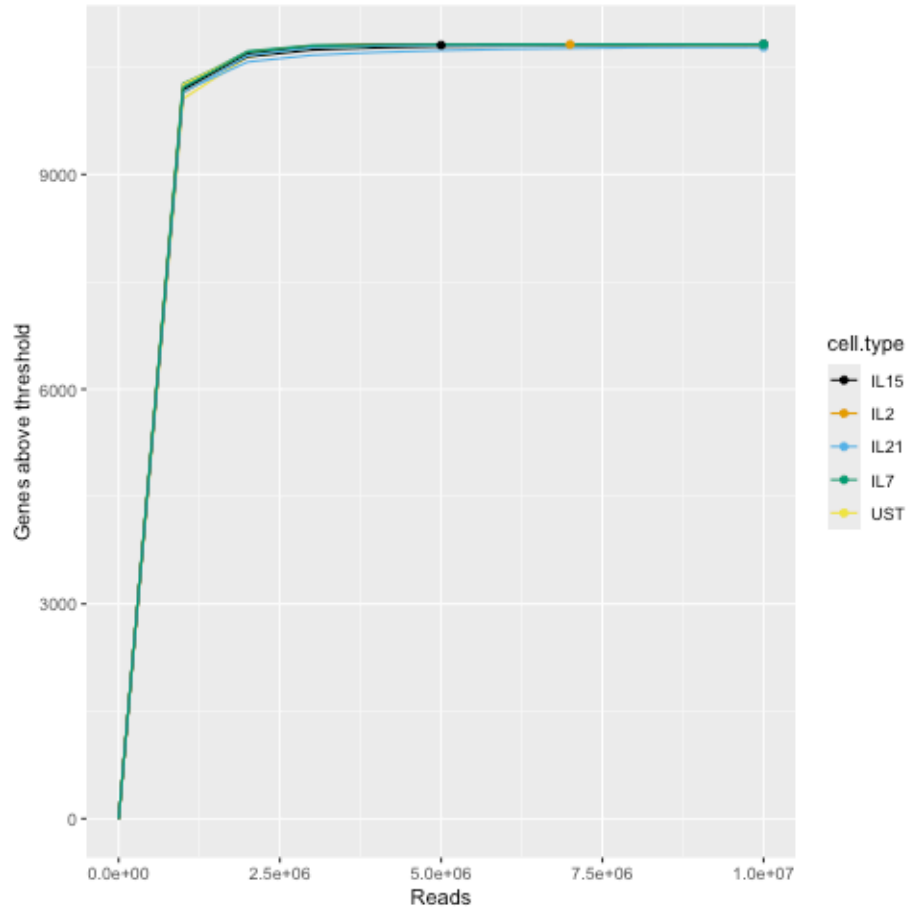


Figure 2.2: Representative example of sequencing saturation.

Two sequencing runs were performed on the same samples. This sequencing saturation plot visualises random sampling and discovery of genes across the dataset. This shows that the samples all reached sequencing saturation. All five conditions (IL-2, IL-7, IL-15, IL-21 and unstimulated) show near-identical saturation curves, plateauing at just under 12,000 genes above threshold once read depth exceeds approximately 2.5 million reads, indicating that sequencing depth was sufficient and comparable across conditions.

2.3.4 DESeq2 Differential expression workflow

The DESeq2 package and proposed workflow is well established and very popular (Love et al. 2014). The first step was to transform the data. This allows for more relevant comparisons between samples. The variance stabilised (VST) object can

then be visualised (Figure 2.3). Most of the genes would be expected to be of low variance, with some genes being highly variable across the samples.

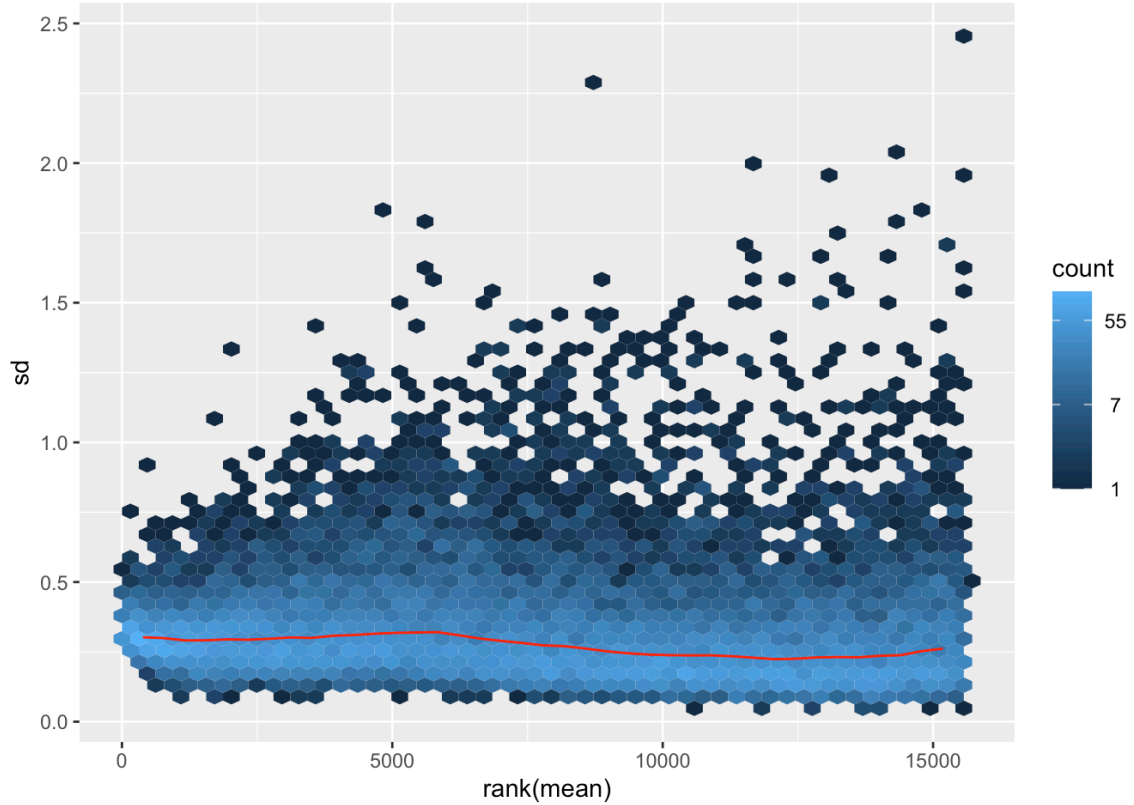


Figure 2.3: VST transformed RNA sequencing data

Genes from all samples ranked by variance, of variance stabilised transformed data. This shows that most datapoints (genes) are around zero in terms of variance, with some genes being highly variable with an increase in standard deviation. The red trend line remains relatively flat across the full range of mean expression rank, indicating that the variance stabilising transformation was largely successful in removing the dependence of variance on mean expression that is characteristic of raw count data, although a number of high-variance outlier genes remain visible above the main band.

2.3.5 Identifying highly variable genes

Highly variable genes were assessed compared to all conditions, to see if the variance is described by any of the conditions (Figure 2.4). This allows assessment of whether the variance in the most highly variable genes (HVGs) also correlates with the different

conditions or biological replicates. Some HVGs were related to collagen, including the most highly variable gene, COL5A3.

Although there was variability between biological samples, it is still possible to conduct other comparisons to see what other genes have expression profiles that may correlate with conditions regardless of biological replicate or potential technical effect.

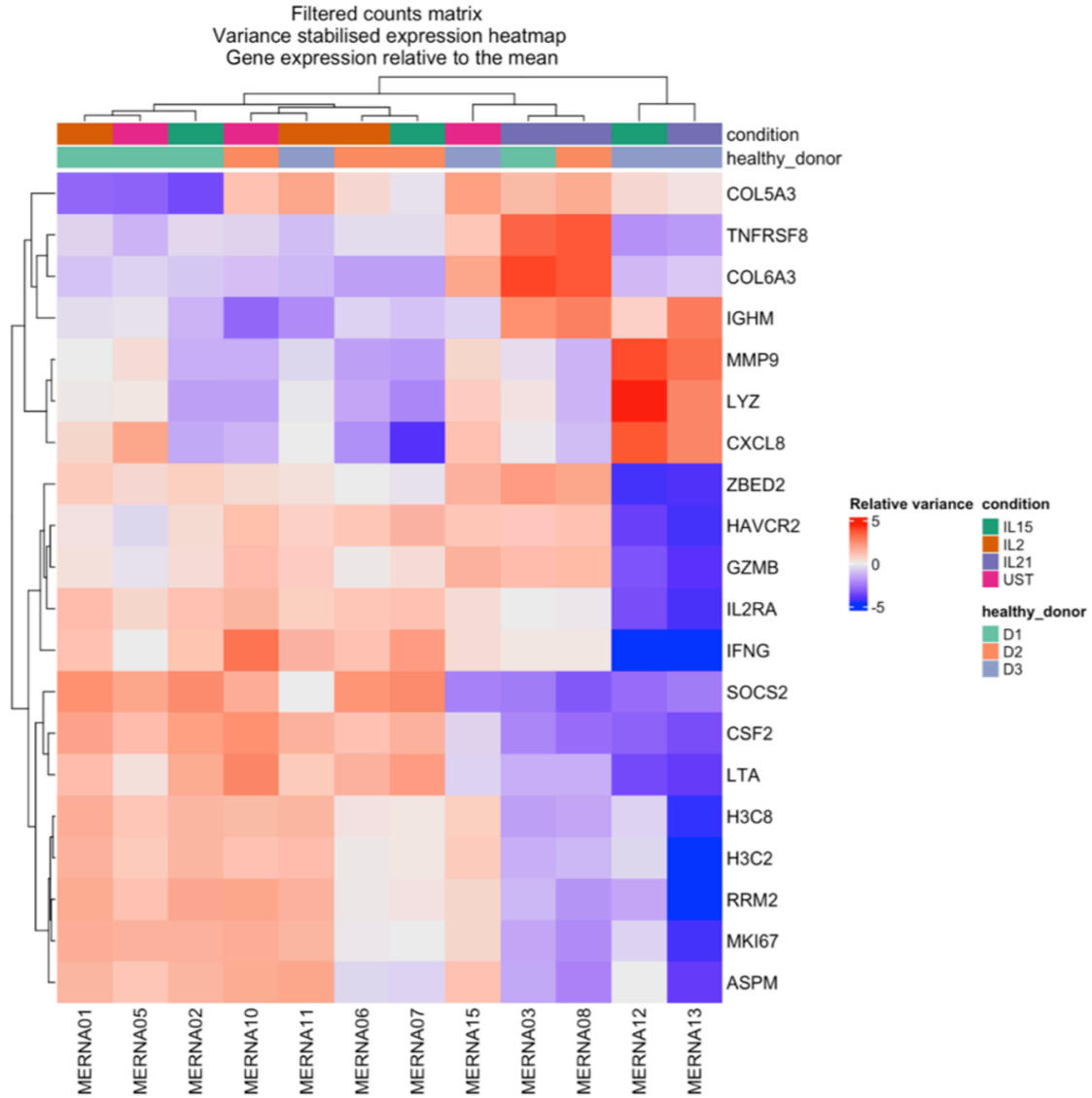


Figure 2.4: Heatmap showing variance across all samples

Heatmap of the genes with the highest variance across all samples. Of the samples that were treated with IL-21, one donor has very difference variance compared to the other samples. This may be due to biological or technical effect. The two IL-21 samples from donor D3 form a distinct cluster with strongly reduced expression of genes such as ZBED2, IL2RA, IFNG, SOCS2 and the proliferation markers MKI67 and ASPM relative to all other samples, while genes including TNFRSF8, IGHM, MMP9, LYZ and CXCL8 are correspondingly elevated in these two samples, underscoring how much this single donor's IL-21 response diverges from the rest of the cohort.

2.3.6 Dimensionality reduction of RNA sequencing data

A common method to further characterise the RNA sequencing data is to conduct a PCA plot (Figure 2.5). This attempts to fit as much of the variance in the whole dataset in 2 dimensions, at a sample-wise level. PC1 and PC2 explains 61 and 17 percent variance respectively. This enables visualisation of the similarity of the samples with respect to each other. The datapoints were colour-coded to the associated metadata to see what correlates with the variance in the data. Shown here are the samples coloured by condition, and this plot shows that the IL-21 condition is most similar to each other, and different from the rest of the samples. This is mostly noticeable in PC1, which explains 61% variance. This makes the IL-21 condition the best candidate for further investigation to understand what differential gene expression drives the cohort-wide variance. The samples were also colour-coded by technical metadata, and there was no correlation with PC1 and PC2. There is one IL-21 condition sample that is further apart from the other samples of the condition – as with the previous figure (Figure 2.4), healthy donor 3, condition IL-21 (MERNA13) sits apart from the rest of the data indicating it has a unique expression profile that is very different to the rest of the samples. However, this does not cause the greatest variance, and MERNA13 is still closer in similarity to the rest of the IL-21 condition samples compared to the rest of the samples.

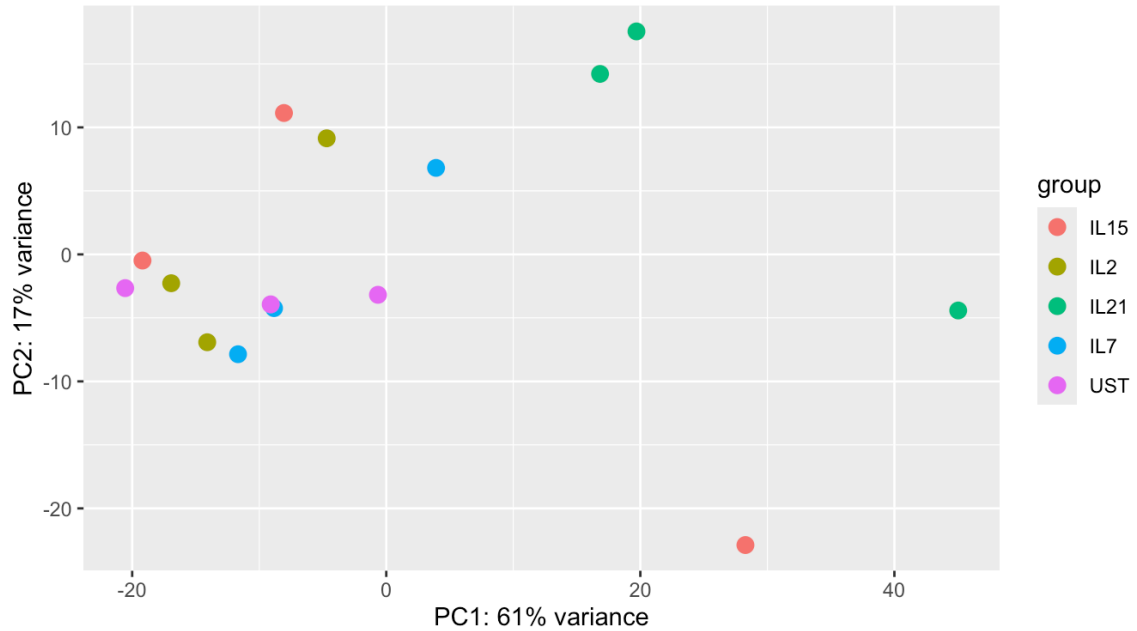


Figure 2.5: MDS plot of RNA sequencing dataset

Multidimensional scaling (MDS) plot showing variance in the dataset for each sample. Samples are colour-coded by condition. IL-21 is clustered together across biological replicates and share similarity within the group, different to the rest of the samples. The other conditions do not show obvious difference based on sample variance. The three IL-21 samples span the widest range along PC1, which explains 61% of total variance, from the most tightly grouped condition on one axis to the most extreme sample overall, again reflecting the disproportionate influence of donor D3 on this condition's expression profile.

2.3.7 Identifying differentially expressed genes

The next comparison is the IL-21 condition group against the IL-2 and IL-15 conditions, to identify differentially expressed genes that are unique to IL-21 stimulation. This can indicate genes and differentiation pathways that $\gamma\delta$ T cells undergo during specific inflammatory conditions.

When comparing to unstimulated, many housekeeping genes such as collagen related genes were identified as differentially expressed. To further understand the more subtle differences between the different cytokine stimulations, IL-21 was compared to the other two cytokine stimulation conditions. This comparison can be seen as more informative because this compares against samples that have some cytokine stimulation, rather than comparing to the unstimulated sample. Therefore, volcano plots for IL-21 vs IL-2 and IL-21 vs IL-15 were produced (Figure 2.6).

Of note, CXCL13 and TNFRSF8 (CD30) was up-regulated in IL-21 compared to both IL-2 and IL-15 conditions. CSF2, SOCS2, LTA, DUSP6, CEACAM1, IFNG, OSM pro-inflammatory genes are downregulated in IL-21 compared to IL-2 and IL-15. There are some non-immune-related genes that were differentially expressed in the IL-21 group. COL6A3 is enriched in IL-21 condition. This could be due to contamination of other cells, potentially with stromal cells being present during the cell culture experiment.

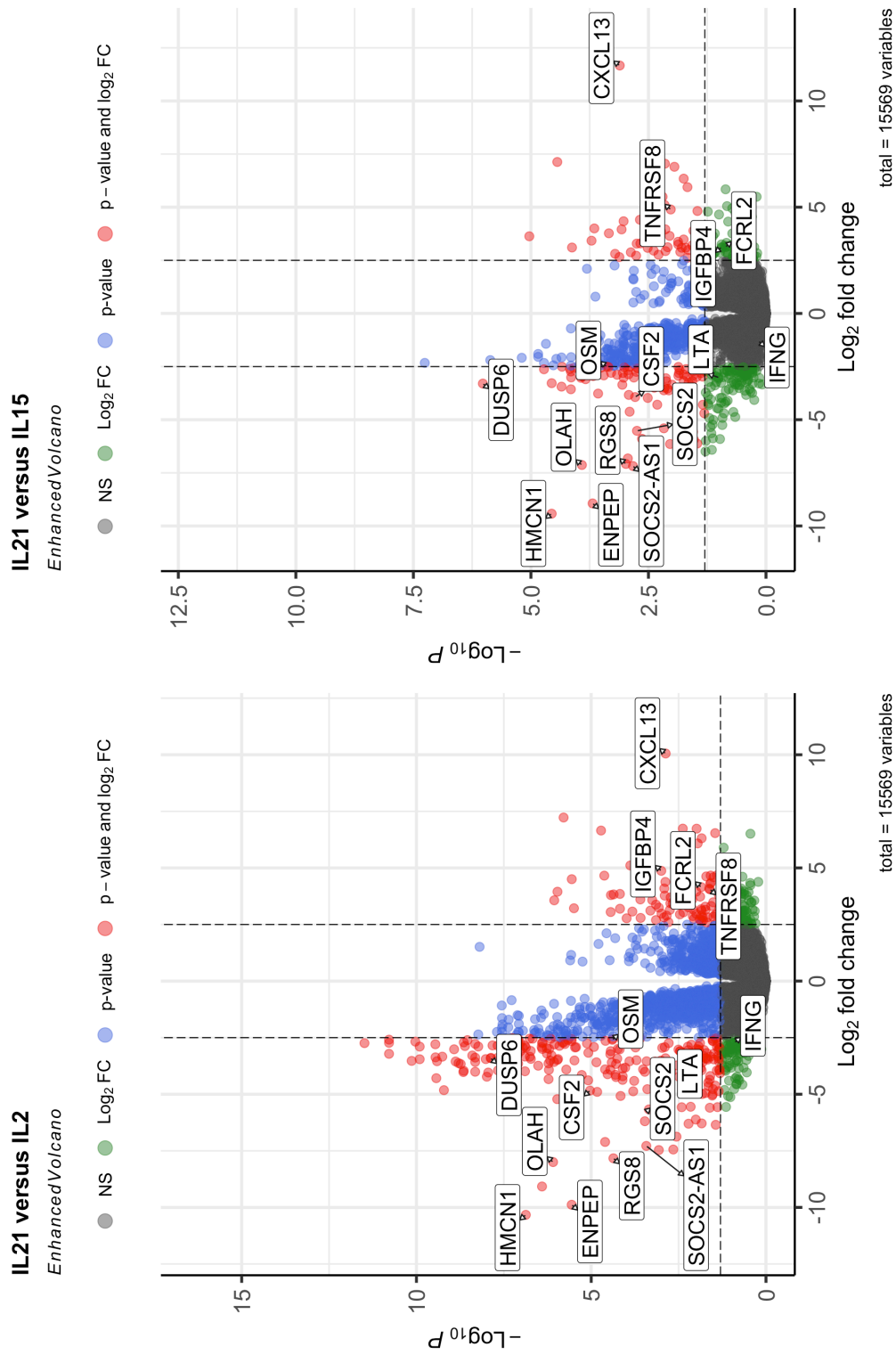


Figure 2.6: Volcano plot IL-21 condition against IL-2 and IL-15

IL-21, with the highest variance, against the other two conditions tested. In both the IL-15 and IL-2 conditions, there are similarities in the genes that are differentially expressed compared to the IL-21 condition, including CXCL13 and TNFRSF8. Dashed lines indicate a log₂ fold change of 2.5 and an adjusted p-value of 0.05; labelled genes were selected by a stricter fold-change threshold or prior biological interest, rather than by significance (adjusted p-value < 0.05) alone. Point colour indicates which threshold(s) each gene meets: red for both, blue for fold change alone, and green for non-significant.

2.4 Discussion

The aim of this chapter was to characterise $\gamma\delta$ T cells during acute inflammation at the intestinal barrier with an *in vitro* experiment. FACS sorted $\gamma\delta$ T cells were RNA sequenced to measure gene expression changes between the different condition groups. This data has shown upregulation of inflammatory genes that attract immune cells in the IL-21 pro-inflammatory environment.

2.4.1 Cohort size

The cohort size of this experiment is small. There are three biological replicates (donors) for each condition, and no technical replicates. Donor D3 shows a similar gene expression profile for both IL-21 (MERNA13) and IL-15 (MERNA12), indicating that the highest variance in donor D3 is explained by donor origin, not condition. In the plot showing the overall highest variance in the data (Figure 2.4), this means that the IL-15 condition of donor D3 is clustered with the rest of the IL-21 condition samples. Unlike donor D3, the gene expression profiles of donor D1 and D2 were similar within each condition, and comparing IL-21 to IL-15 condition, show relative upregulation of TNFRSF8, COL6A3 and IGHM, with downregulation of SOCS2, CSF2 and LTA. Donor D3, however, shows downregulation of TNFRSF8, seemingly showing the opposite relative expression as found in donor D1 and D2 for some genes. This then affects the overall analysis downstream. For example, when showing a condition-wise comparison of all the data, TNFRSF8 is not one of the most differentially expressed genes (Figure 2.6).

It is not possible to determine if donor D3 is representative of the population, if they are a non-responder, or if a technical error occurred during the cell culture

experiment. All the data passed quality checking, and so there appeared to be no technical issues with the library preparation and sequencing. There is no reason to remove donor D3 from analysis, as there is no evidence of any technical problems during data acquisition. More biological replicates would determine if this a biological or technical effect. With only three donors in this experiment, it is not possible to determine what causes this gene expression difference between donor D3 and donors D1 and D2, and so was included in subsequent analysis.

The most obvious difference between conditions can be shown with an MSD plot (Figure 2.5). The MSD plot shows that the IL-21 condition clusters exclusively, which is not visually apparent in the other conditions. The other conditions are not as obviously clustered, and so the variance appears to be explained by donor ID as much and condition. It could be that there is a small effect size in these conditions, which cannot be measured with this cohort size, or that there is no change between unstimulated condition and the IL-15 and IL-7 conditions. As IL-21 drives the most unique change in the data, this then became the focus of the analysis, and was used as the comparator against IL-2 and IL-15 to characterise what genes are differentially expressed in IL-21.

Pathway analysis is another method of inferring a potential function response. This is commonly done using gene set enrichment analysis. In this case, however, with the small cohort and with one individual with a different gene expression profile to the other two, the data was not high-powered enough to look at gene sets to known pathways. Considering that this is a cell culture experiment, it may not provide physiologically relevant results. Therefore, in this case, the most differentially expressed genes were identified and described, and further studies would be required

for understanding functional output.

A likelihood-ratio test was also used as a multivariate approach, testing for differential expression across the four conditions of Unstimulated, IL-2, IL-15 and IL-21 simultaneously rather than in pairwise contrasts, and this identified a substantial number of significant genes. IL-7 was not included in this model, as it was used as a negative control condition: despite being a closely related γ -chain cytokine, IL-7 does not drive the same functional response in $\gamma\delta$ T cells as IL-2, IL-15 and IL-21, so it was excluded from this comparison of active cytokine environments. However, which genes reached significance differed from the pairwise comparisons above, reflecting the variability between donors D1, D2 and D3: because D3's expression profile is so different from D1 and D2, the IL-21-associated genes highlighted above were repressed in this joint model. The clearest way to distinguish differences between conditions remained the pairwise Wald test, as shown in the volcano plot (Figure 2.6). With only three biological replicates per condition and no technical replicates, this remains a limited cohort for any multivariate model, and general guidance for RNA sequencing experiments recommends substantially larger replicate numbers to reliably detect differential expression (Schurch et al. 2016).

2.4.2 IL-21 condition – Differentially expressed genes

In an IL-21-rich environment with CD4 T cells, $\gamma\delta$ T cells show a transcriptional shift that appears to be an immunoregulatory phenotype rather than cytotoxic.

Upregulation of CXCL13 suggests involvement in B-cell recruitment and tertiary lymphoid structure (TLS)-like organisation. Increased TNFRSF8 (CD30) is associated with T cell activation and co-stimulatory signalling with CD4⁺ T cells (Cui

et al. 2024). FCRL2 (statistically significant in only one comparison, IL-21 vs IL-2) is associated with regulating B cell function (Jackson et al. 2010).

Concurrent downregulation of DUSP6 and SOCS2 indicates relief of negative feedback on cytokine signalling. Reduced CSF2 and LTA expression suggests that the functional role of $\gamma\delta$ T cells are not pro-inflammatory in this IL-21 environment.

Collectively, these changes suggest $\gamma\delta$ T cells may participate in shaping adaptive immune architecture and regulate CD4⁺ T cell response, likely promoting localised immune organisation and regulation rather than direct effector inflammation.

2.4.3 Tertiary lymphoid structure

Both CD30 and CXCL13 expression increased in the IL-21 environment compared to IL-15 and may suggest that $\gamma\delta$ T cells are also interacting with, or regulating, B cell recruitment and a humoral response.

Multiple studies have found that CD30 deficient mice lose sustained germinal centre formation, indicating a role of CD30 in facilitating an adaptive immune response – though this has been contradicted by an older study from one lab group (Van Der Weyden et al. 2017). CD30 upregulation is also associated with certain lymphomas and is a potential target for immunotherapy.

IL-10 upregulates CXCL13 in lung macrophages and monocyte-derived macrophages through activation of the JAK/STAT pathway (Zhao et al. 2024). Previous work by the Eberl lab has shown that $\gamma\delta$ T cells, co-cultured with CD4 T cells, show an increase in IL-10 expression.

Taken together, this indicates that $\gamma\delta$ T cells may be associated with tertiary lymphoid structure (TLS) assembly in gut-associated lymphoid tissues, and corroborates

with published literature concerning the role of CXCL13 in driving tertiary lymphoid structure assembly and recruitment of cells related to the humoral immune response. Further investigation in JAK/STAT pathway would further bolster this.

This study shows the utility in applying robust workflows to the data. The adoption of standardised workflows like those in **NF-Core** ensures analytical consistency and enhances the reproducibility. This analysis shows the upregulation of CXCL13 and CD30 in the context of IL-21 stimulation. This indicates the role of $\gamma\delta$ T cells in orchestrating tertiary TLS formation through recruitment.

These findings support a broader model in which $\gamma\delta$ T cells contribute to immune response to both the tumour microenvironment and acute infection.

2.4.4 Future direction

As described above, further work would include more biological replicates to determine if donor D3 is representative of a broader population. Repeating the experiment with technical replicates would also improve confidence in the data.

This analysis has indicated a role of JAK/STAT pathway modulation in the IL-21 condition, and so experiments with JAK/STAT reporter cell lines would show if this pathway is being activated in the IL-21 environment compared to the other cytokine conditions.

Co-culture with B cells and measuring germinal centre-like differentiation (such as BCL6) would also indicate if the CXCL13 upregulation found in the RNA sequencing data translates to a functional TLS response.

DISCUSSION

$\gamma\delta$ T cells are tissue specific, innate immune cells that can launch an early immune response against bacterial infection by releasing inflammatory related markers at the site of infection to recruit other circulatory immune cells, regulating that response, and clearing infection. Previous work has found that $\gamma\delta$ T cells can modulate $CD4^+$ T cell response to infection, causing an increase in IL-10 production of $CD4^+$ T cells. This indicates that $\gamma\delta$ T cells have a regulatory role. These findings have shown that the response of $\gamma\delta$ T cells to infection is dependent on the environment surrounding them, as shown by the RNA sequencing data in the presence of IL-21.

The largest variance between the groups was described in the difference between IL-21 and the other conditions. Directed by the data, IL-21 samples were then characterised relative to all other groups. The unstimulated condition did not have the most variance compared to the other groups. This enabled a model for comparison at the different mucosal barriers, and what made IL-21 so unique in comparison. A number of genes upregulated in the IL-21 condition that are also potential targets for further investigation and are associated with immune cell recruitment and initiation

of immune response, such as CXCL13 and CD30.

Improving the quality of flow cytometry data

When expecting small changes in population frequencies in flow cytometry data, the traditional manual gating of flow cytometry can be highly subjective, which can cause a lack of reproducibility and may prevent the detection of subtler populations with lower overall frequency. As discussed further in Chapter 1, an alternative is to implement dimensionality reduction and clustering of the data; this removes the subjectivity of gate placement and can more confidently identify rare populations across the dataset, but is more susceptible to technical effects and less readily interpretable to a biology audience, so a dataset with stringent quality control remains vital.

Regardless of downstream analysis, the initial steps of preprocessing are the same. Removing debris, selecting for live cells, removing doublets, and selecting CD45⁺ leukocytes is best done via manual gating. The initial removal of debris then allows for greater confidence in visualisation of data with manual gating, and this gated data can be exported and analysed by clustering without debris and doublets creating technical noise.

Finding biomarkers associated with complications of anastomosis surgery would be valuable. Current methods include patient observations and an increase in blood CRP. Although measuring blood CRP level is sensitive, it is not specific. Further, determination of AL using blood CRP levels are often only statistically significant 3 days after operation. The initial hypothesis was that cell types can be identified and changes can be measured in peritoneal fluid 1 and 2 days after operation. With further investigation, it could be possible to further characterise the peritoneal fluid

and find biomarkers associated with postoperative complications that are identifiable before elevated blood CRP levels. This is shown by the fact that immune cells were detected from day one post surgery. Being able to identify a large number of immune cells at an early stage could provide a better picture of the recovery of the patient, rather than relying on the delayed, systemic and non-AL-specific CRP marker.

Workflows and pipelines

There is a benefit of the use of pipelines with RNA sequencing data; for citation, reproducibility, and accessibility. Either informal (list of packages) or formal (**Nextflow**) pipelines can be used in these cases and have helped standardise expected output and results that can be compared across labs with confidence. In these cases, input material causes very few changes to the pipeline. This only becomes more niche and experiment-specific if using species other than human and mouse, where there may not be a robust reference genome.

This was standardised further as sequencing became more common and affordable. This would be considered unsupervised data analysis, as all available genes are sequenced and mapped, with the highly variable genes being selected after.

Initially, microarray data was a popular method of sequencing, and this would be considered supervised data. Over time, the methods of sequencing have moved toward total gene expression, and the packages and workflows have evolved with this trend.

With flow cytometry data, there is less of a standardised pipeline. The input and measurable markers are incredibly varied, and there is less standardisation in machine function, and quality and content of output. These methods of directed, supervised

analysis will be difficult to pipeline for everyone, like was done for RNA sequencing data.

An advantage to flow cytometry is the accessibility, cost-effectiveness, scalability, and customisability. Flow cytometry is ideal for longitudinal studies, and the preparation and analysis of samples can be standardised and end-users can be trained with relative ease.

A substantial issue with flow cytometry data is the technical effect. This can happen within a single session, over the course of a project, and most notably between machines. Historically, the way to overcome this is to manually gate on populations, and make comparisons of change *within* a cohort, not across different cohorts. This cross-machine and cross-lab technical effect creates a lot of variability in the quality of the data produced.

With the advent of spectral machines, the output has become more consistent, without the need for complex batch correction to make the data comparable across experiments. Further, spectral machines are able to measure many more markers compared to traditional flow cytometry – panels can now contain up to 40 markers, whereas traditional flow cytometry typically had a maximum of 15 markers. Some large panels have already been optimised and would be a potential improvement to the current experimental design (Barros-Martins et al. 2022). This allows for more consistent panels across studies. The benchmarking of clustering methods in this thesis is the proof of concept of preprocessing and presenting flow cytometry data, in readiness of flow cytometry data that will become more complex in the future.

Conclusion

These two studies show the highly context-dependent role of $\gamma\delta$ T cells as regulators of mucosal immunity, particularly in an IL-21 environment. The identification of key mediators such as CXCL13 and CD30 indicates the role of $\gamma\delta$ T cells for immune cell recruitment and, potentially, a contributory role for the generation of tertiary lymphoid organs, linking innate response and adaptive immunity. These insights open avenues for targeted modulation of $\gamma\delta$ T cells in chronic inflammatory diseases like Crohn's disease, where aberrant IL-21 signalling may drive inflammation. Improving the reproducibility of flow cytometry analyses will be critical to advancing understanding of these complex cellular interactions, especially when applied to larger studies in the future. Combining knowledge from both studies provides more information of potential markers to study. For example, as CXCL13 appears to be influential in the immune response of $\gamma\delta$ T cells, and therefore could be measured in perioperative anastomosis patients via ELISA of peritoneal fluid. CD30 is associated with T cell activation, and so could be a potential marker in peritoneal fluid during acute infection; however, this could be obfuscated by CD30 also associating with tumour development. With the identification of new potential markers, future work would include applying markers identified in cell culture to patient samples. This could elucidate both the nuanced regulatory functions of $\gamma\delta$ T cells and offer a therapeutic potential in bowel-related infection and inflammation.

REFERENCES

- Andreu-Ballester, J. C., Tormo-Calandín, C., Garcia-Ballesteros, C., Pérez-Griera, J., Amigó, V., Almela-Quilis, A., Ruiz Del Castillo, J., Peñarroja-Otero, C. & Ballester, F. (2013), ‘Association of $\gamma\delta$ T Cells with Disease Severity and Mortality in Septic Patients’, *Clinical and Vaccine Immunology* **20**(5), 738–746. doi: 10.1128/CVI.00752-12.
- Aravani, A., Samy, E. F., Thomas, J. D., Quirke, P., Morris, E. J. & Finan, P. J. (2016), ‘A retrospective observational study of length of stay in hospital after colorectal cancer surgery in England (1998–2010)’, *Medicine* **95**(47), e5064. doi: 10.1097/MD.00000000000005064.
- Azad, A., Rajwa, B. & Pothén, A. (2016), ‘flowVS: Channel-specific variance stabilization in flow cytometry’, *BMC Bioinformatics* **17**(1), 291. doi: 10.1186/s12859-016-1083-9.
- Bansal, R. R., Mackay, C. R., Moser, B. & Eberl, M. (2012), ‘IL-21 enhances the potential of human $\gamma\delta$ T cells to provide B-cell help’, *European Journal of Immunology* **42**(1), 110–119. doi: 10.1002/eji.201142017.
- Barros-Martins, J., Bruni, E., Fichtner, A. S., Cornberg, M. & Prinz, I. (2022), ‘OMIP-084: 28-color full spectrum flow cytometry panel for the comprehensive analysis of human $\gamma\delta$ T cells’, *Cytometry Part A* **101**(10), 856–861. doi: 10.1002/cyto.a.24564.
- Becht, E., McInnes, L., Healy, J., Dutertre, C.-A., Kwok, I. W. H., Ng, L. G., Ginhoux, F. & Newell, E. W. (2019), ‘Dimensionality reduction for visualizing single-cell data using UMAP’, *Nature Biotechnology* **37**(1), 38–44. doi: 10.1038/nbt.4314.
- Bhattacharya, S., Dunn, P., Thomas, C. G., Smith, B., Schaefer, H., Chen, J., Hu, Z., Zalusky, K. A., Shankar, R. D., Shen-Orr, S. S., Thomson, E., Wiser, J. & Butte, A. J. (2018), ‘ImmPort, toward repurposing of open access immunological assay data for translational and clinical research’, *Scientific Data* **5**(1), 180015. doi: 10.1038/sdata.2018.15.

- Carding, S. R. & Egan, P. J. (2002), ‘ $\gamma\delta$ T cells: Functional plasticity and heterogeneity’, *Nature Reviews Immunology* **2**(5), 336–345. doi: 10.1038/nri797.
- Colombo, G., Aloisio, E. & Panteghini, M. (2024), ‘Laboratory investigation of peritoneal fluids: An updated practical approach based on the available evidence’, *Journal of Clinical Pathology* **77**(9), 579–585. doi: 10.1136/jcp-2023-209282.
- Costanzo, A. E., Taylor, K. R., Dutt, S., Han, P. P., Fujioka, K. & Jameson, J. M. (2015), ‘Obesity Impairs $\gamma\delta$ T Cell Homeostasis and Antiviral Function in Humans’, *PLOS ONE* **10**(3), e0120918. doi: 10.1371/journal.pone.0120918.
- Cuff, S. M., Reeves, N., Lewis, E., Jones, E., Baker, S., Karategos, A., Morris, R., Torkington, J. & Eberl, M. (2023), ‘Inflammatory biomarker signatures in post-surgical drain fluid may detect anastomotic leaks within 48 hours of colorectal resection’, *Techniques in Coloproctology* **27**(12), 1297–1305. doi: 10.1007/s10151-023-02841-y.
- Cui, D., Zhang, Y., Chen, L., Du, H., Zheng, B., Huang, M., Li, X., Wei, J. & Chen, Q. (2024), ‘CD30 plays a role in T-dependent immune response and T cell proliferation’, *The FASEB Journal* **38**(1), e23365. doi: 10.1096/fj.202301747RR.
- Davey, M. S., Willcox, C. R., Joyce, S. P., Ladell, K., Kasatskaya, S. A., McLaren, J. E., Hunter, S., Salim, M., Mohammed, F., Price, D. A., Chudakov, D. M. & Willcox, B. E. (2017), ‘Clonal selection in the human $V\delta 1$ T cell repertoire indicates $\gamma\delta$ TCR-dependent adaptive immune surveillance’, *Nature Communications* **8**(1), 14760. doi: 10.1038/ncomms14760.
- Den Braanker, H., Bongenaar, M. & Lubberts, E. (2021), ‘How to Prepare Spectral Flow Cytometry Datasets for High Dimensional Data Analysis: A Practical Workflow’, *Frontiers in Immunology* **12**, 768113. doi: 10.3389/fimmu.2021.768113.
- den Dulk, M., Witvliet, M. J., Kortram, K., Neijenhuis, P. A., de Hingh, I. H., Engel, A. F., van de Velde, C. J. H., de Brauw, L. M., Putter, H., Brouwers, M. a. M. & Steup, W.-H. (2013), ‘The DULK (Dutch leakage) and modified DULK score compared: Actively seek the leak’, *Colorectal Disease* **15**(9), e528–e533. doi: 10.1111/codi.12379.
- Di Lorenzo, B., Déchanet-Merville, J. & Silva-Santos, B. (2017), ‘Peripheral clonal selection shapes the human $\gamma\delta$ T-cell repertoire’, *Cellular & Molecular Immunology* **14**(9), 733–735. doi: 10.1038/cmi.2017.51.
- Di Tommaso, P., Chatzou, M., Floden, E. W., Barja, P. P., Palumbo, E. & Notredame, C. (2017), ‘Nextflow enables reproducible computational workflows’, *Nature Biotechnology* **35**(4), 316–319. doi: 10.1038/nbt.3820.
- Emile, S. H. & Abd El-Hamed, T. M. (2017), ‘Routine Drainage of Colorectal Anastomoses: An Evidence-Based Review of the Current Literature’, *Gastroenterology Research and Practice* **2017**, 1–7. doi: 10.1155/2017/6253898.

- Emmanuel, A., Chohda, E., Botfield, C. & Ellul, J. (2018), 'Accelerated discharge within 72 hours of colorectal cancer resection using simple discharge criteria', *The Annals of The Royal College of Surgeons of England* **100**(1), 52–56. doi: 10.1308/rcsann.2017.0149.
- Ewels, P. A., Peltzer, A., Fillinger, S., Patel, H., Alneberg, J., Wilm, A., Garcia, M. U., Di Tommaso, P. & Nahnsen, S. (2020), 'The nf-core framework for community-curated bioinformatics pipelines', *Nature Biotechnology* **38**(3), 276–278. doi: 10.1038/s41587-020-0439-x.
- Ewels, P., Magnusson, M., Lundin, S. & Käller, M. (2016), 'MultiQC: Summarize analysis results for multiple tools and samples in a single report', *Bioinformatics (Oxford, England)* **32**(19), 3047–3048. doi: 10.1093/bioinformatics/btw354.
- Fay, N. S., Larson, E. C. & Jameson, J. M. (2016), 'Chronic Inflammation and $\gamma\delta$ T Cells', *Frontiers in Immunology* **7**. doi: 10.3389/fimmu.2016.00210.
- Fearon, K., Ljungqvist, O., Von Meyenfeldt, M., Revhaug, A., Dejong, C., Lassen, K., Nygren, J., Hausel, J., Soop, M., Andersen, J. & Kehlet, H. (2005), 'Enhanced recovery after surgery: A consensus review of clinical care for patients undergoing colonic resection', *Clinical Nutrition* **24**(3), 466–477. doi: 10.1016/j.clnu.2005.02.002.
- Fichtner, A. S., Ravens, S. & Prinz, I. (2020), 'Human $\gamma\delta$ TCR Repertoires in Health and Disease', *Cells* **9**(4), 800. doi: 10.3390/cells9040800.
- Fung, W. W.-S. & Li, P. K.-T. (2022), 'Recent advances in novel diagnostic testing for peritoneal dialysis-related peritonitis', *Kidney Research and Clinical Practice* **41**(2), 156–164. doi: 10.23876/j.krcp.21.204.
- Godfrey, D. I., Koay, H.-F., McCluskey, J. & Gherardin, N. A. (2019), 'The biology and functional importance of MAIT cells', *Nature Immunology* **20**(9), 1110–1128. doi: 10.1038/s41590-019-0444-8.
- Guadagni, S., Catarci, M., Masedu, F., Karim, M. E., Clementi, M., Ruffo, G., Viola, M. G., Borghi, F., Baldazzi, G., Scatizzi, M., Pirozzi, F., Delrio, P., Garulli, G., Marini, P., Campagnacci, R., De Luca, R., Ficari, F., Sica, G., Scabini, S., Liverani, A., Caricato, M., Patriti, A. & the Italian ColoRectal Anastomotic Leakage (iCral) study group (2024), 'Abdominal drainage after elective colorectal surgery: Propensity score-matched retrospective analysis of an Italian cohort', *BJS Open* **8**(1), zrad107. doi: 10.1093/bjsopen/zrad107.
- Guo, M., Bafligil, C., Tapmeier, T., Hubbard, C., Manek, S., Shang, C., Martinez, F. O., Schmidt, N., Obendorf, M., Hess-Stumpp, H., Zollner, T. M., Kennedy, S., Becker, C. M., Zondervan, K. T., Cribbs, A. P. & Oppermann, U. (2020), 'Mass cytometry analysis reveals a distinct immune environment in peritoneal fluid in endometriosis: A characterisation study', *BMC Medicine* **18**(1), 3. doi: 10.1186/s12916-019-1470-y.

- Gustafsson, U. O., Scott, M. J., Hubner, M., Nygren, J., Demartines, N., Francis, N., Rockall, T. A., Young-Fadok, T. M., Hill, A. G., Soop, M., De Boer, H. D., Urman, R. D., Chang, G. J., Fichera, A., Kessler, H., Grass, F., Whang, E. E., Fawcett, W. J., Carli, F., Lobo, D. N., Rollins, K. E., Balfour, A., Baldini, G., Riedel, B. & Ljungqvist, O. (2019), ‘Guidelines for Perioperative Care in Elective Colorectal Surgery: Enhanced Recovery After Surgery (ERAS[®]) Society Recommendations: 2018’, *World Journal of Surgery* **43**(3), 659–695. doi: 10.1007/s00268-018-4844-y.
- Hahne, F., LeMeur, N., Brinkman, R. R., Ellis, B., Haaland, P., Sarkar, D., Spidlen, J., Strain, E. & Gentleman, R. (2009), ‘flowCore: A Bioconductor package for high throughput flow cytometry’, *BMC Bioinformatics* **10**(1), 106. doi: 10.1186/1471-2105-10-106.
- Hammond, J., Lim, S., Wan, Y., Gao, X. & Patkar, A. (2014), ‘The Burden of Gastrointestinal Anastomotic Leaks: An Evaluation of Clinical and Economic Outcomes’, *Journal of Gastrointestinal Surgery* **18**(6), 1176–1185. doi: 10.1007/s11605-014-2506-4.
- Hu, Y., Hu, Q., Li, Y., Lu, L., Xiang, Z., Yin, Z., Kabelitz, D. & Wu, Y. (2023), ‘ $\gamma\delta$ T cells: Origin and fate, subsets, diseases and immunotherapy’, *Signal Transduction and Targeted Therapy* **8**(1), 434. doi: 10.1038/s41392-023-01653-8.
- Jackson, T. A., Haga, C. L., Ehrhardt, G. R. A., Davis, R. S. & Cooper, M. D. (2010), ‘FcR-Like 2 Inhibition of B Cell Receptor-Mediated Activation of B Cells’, *The Journal of Immunology* **185**(12), 7405–7412. doi: 10.4049/jimmunol.1002305.
- Kaminski, H., Couzi, L. & Eberl, M. (2021), ‘Unconventional T cells and kidney disease’, *Nature Reviews Nephrology* **17**(12), 795–813. doi: 10.1038/s41581-021-00466-8.
- Kannan, V., Ullah, N., Geddada, S., Ibrahim, A., Munaf Shakir Al-Qassab, Z., Ahmed, O. & Malasevskaja, I. (2025), ‘Impact of “Enhanced Recovery After Surgery” (ERAS) protocols vs. traditional perioperative care on patient outcomes after colorectal surgery: A systematic review’, *Patient Safety in Surgery* **19**(1), 4. doi: 10.1186/s13037-024-00425-9.
- Kehlet, H. (1997), ‘Multimodal approach to control postoperative pathophysiology and rehabilitation’, *British Journal of Anaesthesia* **78**(5), 606–617. doi: 10.1093/bja/78.5.606.
- Korsunsky, I., Millard, N., Fan, J., Slowikowski, K., Zhang, F., Wei, K., Baglaenko, Y., Brenner, M., Loh, P.-r. & Raychaudhuri, S. (2019), ‘Fast, sensitive and accurate integration of single-cell data with Harmony’, *Nature Methods* **16**(12), 1289–1296. doi: 10.1038/s41592-019-0619-0.
- Kröger, C., Müller, S., Leidner, J., Kröber, T., Warnat-Herresthal, S., Spintge, J. B., Zajac, T., Neubauer, A., Frolov, A., Carraro, C., Jessen, F., Puccio, S., Aschenbrenner, A. C., Schultze, J. L., Pecht, T., Beyer, M. D. & Bonaguro, L. (2024), ‘Unveiling the power of high-dimensional cytometry data with cyCONDOR’, *Nature Communications* **15**(1), 10702. doi: 10.1038/s41467-024-55179-w.

- Li, Y., Li, G., Zhang, J., Wu, X. & Chen, X. (2021), ‘The Dual Roles of Human $\gamma\delta$ T Cells: Anti-Tumor or Tumor-Promoting’, *Frontiers in Immunology* **11**, 619954. doi: 10.3389/fimmu.2020.619954.
- Liu, Y., Han, B., Xu, W., Wang, Y., Huo, M., Wang, J., Wang, H. & Li, Z. (2026), ‘Neutrophil-to-lymphocyte ratio as an early predictor of anastomotic leakage after rectal cancer surgery’, *Surgery* **190**, 109829. doi: 10.1016/j.surg.2025.109829.
- Liuzzi, A. R., Kift-Morgan, A., Lopez-Anton, M., Friberg, I. M., Zhang, J., Brook, A. C., Roberts, G. W., Donovan, K. L., Colmont, C. S., Toleman, M. A., Bowen, T., Johnson, D. W., Topley, N., Moser, B., Fraser, D. J. & Eberl, M. (2016), ‘Unconventional Human T Cells Accumulate at the Site of Infection in Response to Microbial Ligands and Induce Local Tissue Remodeling’, *The Journal of Immunology* **197**(6), 2195–2207. doi: 10.4049/jimmunol.1600990.
- Love, M. I., Huber, W. & Anders, S. (2014), ‘Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2’, *Genome Biology* **15**(12), 550. doi: 10.1186/s13059-014-0550-8.
- Monaco, G., Chen, H., Poidinger, M., Chen, J., De Magalhães, J. P. & Larbi, A. (2016), ‘flowAI: Automatic and interactive anomaly discerning tools for flow cytometry data’, *Bioinformatics (Oxford, England)* **32**(16), 2473–2480. doi: 10.1093/bioinformatics/btw191.
- Na, Y.-J., Jin, J.-O., Lee, M.-S., Song, M.-G., Lee, K.-S. & Kwak, J.-Y. (2008), ‘Peritoneal fluid from endometriosis patients switches differentiation of monocytes from dendritic cells to macrophages’, *Journal of Reproductive Immunology* **77**(1), 63–74. doi: 10.1016/j.jri.2007.03.013.
- NICE TA105 (2006), ‘Laparoscopic Surgery for Colorectal Cancer’. doi: <https://www.nice.org.uk/guidance/ta105>.
- Ogishi, M., Yang, R., Gruber, C., Zhang, P., Pelham, S. J., Spaan, A. N., Rosain, J., Chbihi, M., Han, J. E., Rao, V. K., Kainulainen, L., Bustamante, J., Boisson, B., Bogunovic, D., Boisson-Dupuis, S. & Casanova, J.-L. (2021), ‘Multibatch Cytometry Data Integration for Optimal Immunophenotyping’, *The Journal of Immunology* **206**(1), 206–213. doi: 10.4049/jimmunol.2000854.
- Pasman, L. & Kasper, D. L. (2017), ‘Building conventions for unconventional lymphocytes’, *Immunological Reviews* **279**(1), 52–62. doi: 10.1111/imr.12576.
- Perfetto, S. P., Ambrozak, D., Nguyen, R., Chattopadhyay, P. K. & Roederer, M. (2012), ‘Quality assurance for polychromatic flow cytometry using a suite of calibration beads’, *Nature Protocols* **7**(12), 2067–2079. doi: 10.1038/nprot.2012.126.

- Rajabaleyan, P., Jootun, R., Möller, S., Deding, U., Ellebæk, M. B., al-Najami, I. & Lindsey, I. (2024), ‘Early warning model to detect anastomotic leakage following colon surgery: A clinical observational study’, *Annals of Coloproctology* **40**(5), 431–439. doi: 10.3393/ac.2023.00745.0106.
- Rhodes, D. A., Reith, W. & Trowsdale, J. (2016), ‘Regulation of Immunity by Butyrophilins’, *Annual Review of Immunology* **34**(1), 151–172. doi: 10.1146/annurev-immunol-041015-055435.
- Schurch, N. J., Schofield, P., Gierliński, M., Cole, C., Sherstnev, A., Singh, V., Wrobel, N., Gharbi, K., Simpson, G. G., Owen-Hughes, T., Blaxter, M. & Barton, G. J. (2016), ‘How many biological replicates are needed in an RNA-seq experiment and which differential expression tool should you use?’, *RNA* **22**(6), 839–851. doi: 10.1261/rna.053959.115.
- Sherwood, A. M., Desmarais, C., Livingston, R. J., Andriesen, J., Haussler, M., Carlson, C. S. & Robins, H. (2011), ‘Deep Sequencing of the Human TCR γ and TCR β Repertoires Suggests that TCR β Rearranges After $\alpha\beta$ and $\gamma\delta$ T Cell Commitment’, *Science Translational Medicine* **3**(90). doi: 10.1126/scitranslmed.3002536.
- Shi, C. & Pamer, E. G. (2011), ‘Monocyte recruitment during infection and inflammation’, *Nature Reviews Immunology* **11**(11), 762–774. doi: 10.1038/nri3070.
- Spasic, M., Ogayo, E. R., Parsons, A. M., Mittendorf, E. A., van Galen, P. & McAllister, S. S. (2024), ‘Spectral Flow Cytometry Methods and Pipelines for Comprehensive Immunoprofiling of Human Peripheral Blood and Bone Marrow’, *Cancer Research Communications* **4**(3), 895–910. doi: 10.1158/2767-9764.CRC-23-0357.
- Spidlen, J., Breuer, K., Rosenberg, C., Kotecha, N. & Brinkman, R. R. (2012), ‘FlowRepository: A resource of annotated flow cytometry datasets associated with peer-reviewed publications’, *Cytometry Part A* **81A**(9), 727–731. doi: 10.1002/cyto.a.22106.
- Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A. & Bray, F. (2021), ‘Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries’, *CA: A Cancer Journal for Clinicians* **71**(3), 209–249. doi: 10.3322/caac.21660.
- Tidadini, F., Bonne, A., Trilling, B., Quesada, J.-L., Sage, P.-Y., Foote, A., Arvieux, C. & Faucheron, J.-L. (2022), ‘Effect of implementation of enhanced recovery after surgery (ERAS) protocol and risk factors on 3-year survival after colorectal surgery for cancer—a retrospective cohort of 1001 patients’, *International Journal of Colorectal Disease* **37**(5), 1151–1159. doi: 10.1007/s00384-022-04155-1.
- Toskas, A., Miliadis, S., Delis, G., Meditskou, S., Sioga, A. & Papamitsou, T. (2023), ‘Expression of IL-21 and IL-33 in Intestinal Mucosa of Inflammatory Bowel Disease: An Immunohistochemical Study’, *Diagnostics* **13**(13), 2185. doi: 10.3390/diagnostics13132185.

- Tyler, C. J., Doherty, D. G., Moser, B. & Eberl, M. (2015), 'Human V γ 9/V δ 2 T cells: Innate adaptors of the immune system', *Cellular Immunology* **296**(1), 10–21. doi: 10.1016/j.cellimm.2015.01.008.
- Tyler, C. J., McCarthy, N. E., Lindsay, J. O., Stagg, A. J., Moser, B. & Eberl, M. (2017), 'Antigen-Presenting Human $\gamma\delta$ T Cells Promote Intestinal CD4⁺ T Cell Expression of IL-22 and Mucosal Release of Calprotectin', *The Journal of Immunology* **198**(9), 3417–3425. doi: 10.4049/jimmunol.1700003.
- Van Der Weyden, C. A., Pileri, S. A., Feldman, A. L., Whisstock, J. & Prince, H. M. (2017), 'Understanding CD30 biology and therapeutic targeting: A historical perspective providing insight into future directions', *Blood Cancer Journal* **7**(9), e603–e603. doi: 10.1038/bcj.2017.85.
- Van Gassen, S., Gaudilliere, B., Angst, M. S., Saeys, Y. & Aghaeepour, N. (2020), 'CytoNorm: A Normalization Algorithm for Cytometry Data', *Cytometry Part A* **97**(3), 268–278. doi: 10.1002/cyto.a.23904.
- Vermijlen, D., Ellis, P., Langford, C., Klein, A., Engel, R., Willimann, K., Jomaa, H., Hayday, A. C. & Eberl, M. (2007), 'Distinct Cytokine-Driven Responses of Activated Blood $\gamma\delta$ T Cells: Insights into Unconventional T Cell Pleiotropy', *The Journal of Immunology* **178**(7), 4304–4314. doi: 10.4049/jimmunol.178.7.4304.
- Virshup, I., Bredikhin, D., Heumos, L., Palla, G., Sturm, G., Gayoso, A., Kats, I., Koutrouli, M., Scverse Community, Berger, B., Pe'er, D., Regev, A., Teichmann, S. A., Finotello, F., Wolf, F. A., Yosef, N., Stegle, O. & Theis, F. J. (2023), 'The scverse project provides a computational ecosystem for single-cell omics data analysis', *Nature Biotechnology* **41**(5), 604–606. doi: 10.1038/s41587-023-01733-8.
- Wako, G., Teshome, H. & Abebe, E. (1970), 'Colorectal Anastomosis Leak: Rate , Risk Factors and Outcome in a Tertiary Teaching Hospital, Addis Ababa Ethiopia, a Five Year Retrospective Study', *Ethiopian Journal of Health Sciences* **29**(6). doi: 10.4314/ejhs.v29i6.14.
- Wozniak, K. L., Kolls, J. K. & Wormley, F. L. (2012), 'Depletion of neutrophils in a protective model of pulmonary cryptococcosis results in increased IL-17A production by gamma/delta T cells', *BMC Immunology* **13**(1), 65. doi: 10.1186/1471-2172-13-65.
- Yadav, P. K., Chen, C. & Liu, Z. (2011), 'Potential Role of NK Cells in the Pathogenesis of Inflammatory Bowel Disease', *Journal of Biomedicine and Biotechnology* **2011**, 348530. doi: 10.1155/2011/348530.
- Yang, W., Yu, T., Huang, X., Bilotta, A. J., Xu, L., Lu, Y., Sun, J., Pan, F., Zhou, J., Zhang, W., Yao, S., Maynard, C. L., Singh, N., Dann, S. M., Liu, Z. & Cong, Y. (2020), 'Intestinal microbiota-derived short-chain fatty acids regulation of immune cell IL-22 production and gut immunity', *Nature Communications* **11**(1), 4457. doi: 10.1038/s41467-020-18262-6.

- Zarnescu, E. C., Zarnescu, N. O. & Costea, R. (2021), 'Updates of Risk Factors for Anastomotic Leakage after Colorectal Surgery', *Diagnostics* **11**(12), 2382. doi: 10.3390/diagnostics11122382.
- Zhao, L., Jin, S., Wang, S., Zhang, Z., Wang, X., Chen, Z., Wang, X., Huang, S., Zhang, D. & Wu, H. (2024), 'Tertiary lymphoid structures in diseases: Immune mechanisms and therapeutic advances', *Signal Transduction and Targeted Therapy* **9**(1), 225. doi: 10.1038/s41392-024-01947-5.
- Zhou, L., Wen, F., Chen, G., Liu, J., Liu, H., Peng, Y., Cheng, M., Liu, Y., Zhang, H. & Liu, Y. (2015), 'Cytokine profiles in peritoneal dialysis effluent predicts the peritoneal solute transport rate in continuous ambulatory peritoneal dialysis patients', *International Journal of Clinical and Experimental Medicine* **8**(11), 20424–20433.