Regulation and heterogeneity of human mucosal-associated invariant T (MAIT) cells in blood and liver

Lucy C. Garner
Merton College





This thesis is submitted for the degree of Doctor of Philosophy in Infection, Immunology and Translational Medicine, University of Oxford

Supervisors: Professor Paul Klenerman and Dr Nicholas Provine

Acknowledgements

Somehow it's the end of my time as a DPhil student – I'm not sure where the time has gone, but it does seem like forever ago that I started this journey. I won't ramble on for too long because I think, and hope, that everyone who has helped me along the way knows how much I appreciate it!

First and foremost, I want to thank my supervisors Paul Klenerman and Nick Provine – none of this would have been possible without them. Thanks to Paul for his continual support, for providing me with diverse opportunities during my DPhil and for always having confidence in me (even when I didn't have confidence in myself!). Nick, I hope that your first experience having a DPhil student has been a good (or at least bearable) one. I have really enjoyed working with you, and I am very grateful for all the time you have spent helping and advising me, and for your general encouragement and enthusiasm.

Next, I need to thank all members of the Klenerman group and the TGU, particularly Hannah, Fitz, Ali, Steph, Chantal, Helen and Sarah. I couldn't have asked for a more fun or supportive environment in which to do my DPhil. I would also like to say thanks to the IITM 2015 cohort – it was great to go through this journey together with all of you.

Finally, I need to thank my friends and family, especially my Mum and Dad, Maddy, Elisabeth, Izzy, Anna, Natalie, Catherine, Lauren, Georgina and Sam. Thanks for being patient with me and for cheering me up when required. I think the DPhil is a good test of friendships and I am pleased to report that mine have weathered the storm. I am hoping for more time to spend with you all in 2020!

Table of contents

Abstract 1

Thesis contributions 3

Abbreviations 4

1. Introduction 7

1.1. Key characteristics 8
1.1.1. Frequency and localisation ... 8
1.1.2. TCR usage ... 12
1.1.3. Phenotype ... 13
1.2. MAIT cell definition 15
1.3. Mouse models 15
1.3.1. MR1 knockout ... 15
1.3.2. TCR transgenic ... 16
1.3.3. Non-transgenic ... 17
1.4. Development 19
1.4.1. Thymic development ... 19
1.4.2. Peripheral development ... 27
1.4.3. Foetal development ... 33
1.5. Activation 35
1.5.1. Mechanisms ... 35
1.5.2. Regulation ... 41
1.5.3. Therapeutic modulation ... 42
1.6. Effector function 43
1.6.1. Cytokine production ... 43
1.6.2. Cytotoxic activity ... 45
1.6.3. Tissue repair ... 46
1.6.4. Immune interactions ... 46
1.7. Subsets 49
1.8. Role in disease 51
1.9. Avenues for future MAIT cell research 51
1.10. Conclusion 55
1.11. Aims and key questions of the thesis 55

2. Materials and methods 58

2.1. Data generation 58
2.1.1. PBMC isolation and cell enrichment ... 58
2.1.2. Media ... 58
2.1.3. Human tissue collection and processing ... 59
2.1.4. Ethics ... 60
2.1.5. MR1 tetramer staining ... 60
2.1.6. Surface and intracellular antibody staining for flow cytometry ... 60
2.1.7. Staining for fluorescence-activated cell sorting (FACS) ... 61
2.1.8. FACS ... 61

2.1.9. Staining panels ... 62
2.1.10. *In vitro* stimulation of enriched CD8$^+$ T cells for scRNA-seq and scTCR-seq ... 64
2.1.11. RNA-seq library generation ... 65
2.1.12. ATAC-seq library generation ... 65
2.1.13. 10x Genomics scRNA-seq and scTCR-seq library generation and sequencing ... 67
2.2. Data analysis 68
2.2.1. Flow cytometry ... 68
2.2.2. RNA-seq ... 68
2.2.3. ATAC-seq ... 69
2.2.4. ATAC-seq and RNA-seq ... 73
2.2.5. 10x Genomics scRNA-seq and scTCR-seq ... 75

3. Method development: optimisation of MR1 tetramer staining and the ATAC-seq protocol 83

3.1. Introduction 83
3.1.1. MAIT cell identification ... 83
3.1.2. ATAC-seq method ... 86
3.1.3. Removal of multiplets from paired scRNA-seq and scTCR-seq data ... 87
3.2. Results 90
3.2.1. MAIT cell identification ... 90
3.2.2. ATAC-seq method ... 98
3.2.3. Removal of multiplets from paired scRNA-seq and scTCR-seq data 109
3.3. Discussion 112
3.4. Summary 116

4. Comparison of the transcriptional profile and chromatin accessibility landscape of human blood CD8$^+$ MAIT cells and T_{mem} cells 117

4.1. Aim 120
4.2. Key questions 121
4.3. Experimental design 121
4.4. Results 122
4.4.1. MAIT cells are transcriptionally distinct from T_{mem} and naïve T cells 122
4.4.2. The chromatin accessibility landscape of MAIT cells is distinct from that of naïve T cells and T_{mem} cells ... 128
4.4.3. Identification of TF motifs with increased accessibility in MAIT cells compared with T_{mem} cells using HOMER *de novo* motif discovery ... 135
4.4.4. AME motif enrichment analysis gives similar results to HOMER *de novo* motif discovery ... 138
4.4.5. CENTIPEDE TF footprinting analysis ... 142
4.5. Discussion 149
4.6. Limitations 151
4.7. Future plans 154
4.8. Conclusion 154
4.9. Summary 155

5. Characterisation of the single-cell transcriptional and TCR repertoire heterogeneity of human MAIT cells in the blood and liver 167

5.1. Introduction 167
5.1.1. MAIT cell phenotypic and functional heterogeneity ... 167
5.1.2. MAIT cell TCR repertoire heterogeneity ... 170
5.1.3. MAIT cells in the liver ... 172
5.1.4. Summary and unanswered questions ... 173
5.2. Aim 173
5.3. Key questions 174
5.4. Experimental design (Experiment 1) 174
5.5. Results 174
5.5.1. Gene expression ... 174
5.5.2. TCR usage ... 191
5.5.3. Gene expression and TCR usage ... 209
5.6. Discussion 214
5.7. Limitations 224
5.8. Conclusion 227

6. Delineation of the transcriptional response of human blood MAIT cells to TCR and cytokine stimulation at a single-cell level 249

6.1. Introduction 249
6.1.1. Summary and unanswered questions ... 252
6.2. Aim 253
6.3. Key questions 253
6.4. Experimental design (Experiment 2) 253
6.5. Results 255
6.5.1. Quality control ... 255
6.5.2. Comparison of the gene expression and TCR repertoire of unstimulated MAIT cells with blood MAIT cells from Experiment 1 ... 257
6.5.3. Pathways of MAIT cell activation following TCR and cytokine stimulation ... 275
6.6. Discussion 294
6.7. Limitations 300
6.8. Conclusion 302
6.9. Summary 303

7. Discussion 311

7.1. Limitations 318
7.2. Future experiments 320
7.3. Concluding remarks 322

References 323

List of publications 348

Abstract

Mucosal-associated invariant T (MAIT) cells are an abundant population of innate-like, effector memory T cells, that secrete proinflammatory cytokines and exhibit direct cytotoxic activity. Given their semi-invariant T cell receptor (TCR) and characteristic phenotype, MAIT cells appear as a homogeneous population. However, recent studies report some heterogeneity in surface marker expression, response to stimulation and cytokine production, as well as tissue-specific functionality. Whether such diversity reflects variable activation states, or alternatively indicates the presence of multiple transcriptionally and functionally distinct MAIT cell subsets, is currently unknown. Moreover, while the semi-invariant TCR of MAIT cells is well defined, how TCR usage varies between tissues and individuals is yet to be defined. In this thesis, I have used bulk and single-cell genomic technologies to investigate the regulation and heterogeneity of human blood and liver MAIT cells, and to explore the transcriptional response of blood MAIT cells to TCR and cytokine stimulation. In the blood, $CD8^+$ MAIT cells exhibited a distinct transcriptional and chromatin accessibility landscape relative to $CD8^+$ conventional memory T cells, as well as altered regulation by known and novel TFs. Single-cell RNA-sequencing indicated that human MAIT cells essentially comprise a single subset. Previously reported phenotypic heterogeneity appears to reflect differences in activation state that depend in part on both clonal origin and tissue localisation. Following TCR or cytokine stimulation, MAIT cells responded as a single population, with observed differences in functionality between cells reflecting their relative degree of activation. Lastly, I have provided an in-depth characterisation of the MAIT cell TCR repertoire and demonstrated that TCR usage is highly shared between blood and liver, but essentially unique to each individual.

Overall, this study has provided fundamental insights into human MAIT cell biology and will serve as a useful foundation for studies investigating the role of MAIT cells in health and disease.

Thesis contributions

Chapter 1

The text and figures were modified from my published review article with permission from coauthors:

Garner, L.C., Klenerman, P., and Provine, N.M. (2018). Insights Into Mucosal-Associated Invariant T Cell Biology From Studies of Invariant Natural Killer T Cells. Front. Immunol. *9*, 1478.

Abbreviations

5-OE-RU	5-(2-oxoethylideneamino)-6-D-ribitylaminouracil
5-OP-RU	5-(2-oxopropylideneamino)-6-D-ribitylaminouracil
6-FP	6-formylpterin
AA	Amino acid
Ac-6-FP	Acetyl-6-formylpterin
ANOVA	Analysis of variance
AP-1	Activator protein-1
APC	Antigen-presenting cell
ATAC-seq	Assay for transposase-accessible chromatin using sequencing
BCG	Bacillus Calmette-Guérin
bp	Base pairs
BTK	Bruton tyrosine kinase
C. albicans	*Candida albicans*
cDNA	Complementary DNA
CPM	Counts per million
CRC	Colorectal cancer
CVID	Common variable immunodeficiency
DC	Dendritic cell
DMSO	Dimethyl sulfoxide
DN	Double negative/$CD4^-CD8^-$
DP	Double positive/$CD4^+CD8^+$
E. coli	*Escherichia coli*
F. tularensis	*Francisella tularensis*
FACS	Fluorescence-activated cell sorting
FC	Fold change
FCS	Foetal calf serum
FDR	False discovery rate
FRiP	Fraction of reads in peak regions
GEM	Gel Beads-in-emulsion
GF	Germ-free
GO	Gene ontology
GSEA	Gene set enrichment analysis

HCC	Hepatocellular carcinoma
HIV	Human immunodeficiency virus
ICAM-1	Intercellular adhesion molecule 1
IDR	Irreproducibility discovery rate
Ii	Invariant chain
IL-12R	IL-12 receptor
IL-18R	IL-18 receptor
IL-23R	IL-23 receptor
ILC	Innate lymphoid cell
iNKT	Invariant natural killer T
L. longbeachae	*Legionella longbeachae*
LFA-1	Lymphocyte function-associated antigen 1
M. smegmatis	*Mycobacterium smegmatis*
MFI	Mean fluorescence intensity
MICA	MHC class I-related chain A
MSigDb	Molecular Signatures Database
mtDNA	Mitochondrial DNA
NASH	Nonalcoholic steatohepatitis
NES	Normalised enrichment score
NK	Natural killer
NKG2D	Natural killer cell group 2 member D receptor
PBMC	Peripheral blood mononuclear cell
PCA	Principal component analysis
PFM	Position frequency matrix
PI3K	Phosphatidylinositol 3-kinase
PWM	Position weight matrix
qPCR	Quantitative PCR
RAG	Recombination activating gene
RL-6-Me-7-OH	7-hydroxy-6-methyl-8-D-ribityllumazine
RL-6,7-diMe	6,7-dimethyl-8-D-ribityllumazine
RNA-seq	RNA-sequencing
RPKM	Reads per kilobase million
RT	Room temperature
RUNX	Runt-related transcription factor
S. epidermidis	*Staphylococcus epidermidis*

***S*. Paratyphi A**	*Salmonella enterica* serovar Paratyphi A
***S*. Typhimurium**	*Salmonella enterica* serovar Typhimurium
SAP	SLAM-associated protein
scRNA-seq	Single-cell RNA-sequencing
scTCR-seq	Single-cell TCR-sequencing
SEM	Standard error of the mean
SLAM	Signalling lymphocyte activation molecule
SNN	Shared nearest neighbour
SPF	Specific pathogen-free
SPRI	Solid phase reverse immobilisation
TAP	Transporter associated with antigen processing
T_{CM}	Central memory T
TEC	Thymic epithelial cell
T_{EM}	Effector memory T
TF	Transcription factor
TLR	Toll-like receptor
T_{mem}	Conventional memory T
TMM	Trimmed mean of M-values
TRE	TPA-responsive element
TREC	T cell receptor excision circle
T_{RM}	Tissue-resident memory T
TSS	Transcription start site
UMAP	Uniform manifold approximation and projection
WT	Wild-type

1. Introduction

Mucosal-associated invariant T (MAIT) cells are an abundant population of innate-like T cells that have emerged in recent years as crucial players in the development and maintenance of effective immunity. This is exemplified by the array of infectious, inflammatory and malignant diseases in which they have been implicated and in which they play diverse roles (Toubal et al., 2019).

While MAIT cells express a T cell receptor (TCR) similar to conventional T cells of the adaptive immune system, their TCR is semi-invariant, restricted by the monomorphic MHC Class I-like molecule MR1 and specific for a limited range of bacterially-derived riboflavin metabolite ligands, the most potent of which is 5-(2-oxopropylideneamino)-6-D-ribitylaminouracil (5-OP-RU) (Corbett et al., 2014; Kjer-Nielsen et al., 2012; Tilloy et al., 1999; Treiner et al., 2003). Furthermore, MAIT cells develop an effector memory phenotype intrathymically (prior to peripheral antigen exposure), can be activated by cytokines independent of their TCR, and can rapidly exert their effector functions upon activation without the requirement for clonal expansion, properties more analogous to innate immune cell types (Dusseaux et al., 2011; Martin et al., 2009; Ussher et al., 2014). Thus, MAIT cells belong to the developmentally diverse group of innate-like T cells, which are at the bridge between innate and adaptive immunity, positioned midway along a spectrum of lymphocytes that encompasses innate lymphoid cells (ILC) and natural killer (NK) cells at one end, and conventional adaptive $CD4^+$ and $CD8^+$ T cells at the other (Fig. 1.1). Amongst these populations, MAIT cells exhibit a particularly close developmental and functional relationship with the lipid-specific CD1d-restricted population of invariant natural killer T (iNKT) cells (Garner et al., 2018). Given their

specific traits, MAIT cells are likely to play particularly important roles during the early stages of an immune response, prior to priming and recruitment of conventional effector T cells.

1.1. Key characteristics

1.1.1. Frequency and localisation

MAIT cells are the largest subset of donor-unrestricted T cells in humans. They comprise on average 3.1% of T cells in peripheral blood, but their frequency varies widely between individuals (Gherardin et al., 2018b). In mice, they are considerably less abundant, comprising 0.1% of blood T cells (C57BL/6 mice) (Rahimpour et al., 2015). MAIT cells preferentially localise to peripheral tissues, whilst their frequency is low in peripheral lymphoid organs (Dusseaux et al., 2011; Rahimpour et al., 2015; Salou et al., 2018). In humans, MAIT cells are particularly enriched in the liver (5-50% of T cells) and are also abundant in adipose tissue, in the lung, in the female genital tract, and to varying degrees in the gut (Booth et al., 2015; Carolan et al., 2015; D'Souza et al., 2018; Dusseaux et al., 2011; Fergusson et al., 2015; Gibbs et al., 2016; Gold et al., 2010; Hiejima et al., 2015; Reantragoon et al., 2013; Treiner et al., 2003). In C57BL/6 mice, MAIT cells show enrichment in similar tissues, including the liver (0.6%), the lung (3.3%) and the gut lamina propria (0.7%) (Rahimpour et al., 2015).

Evidence from parabiotic mice indicates that MAIT cells are predominantly tissue-resident in the spleen, liver, skin and lung (with the exception of the small MAIT1 population in the lung), in contrast with conventional $CD4^+$ and $CD8^+$ memory T cells

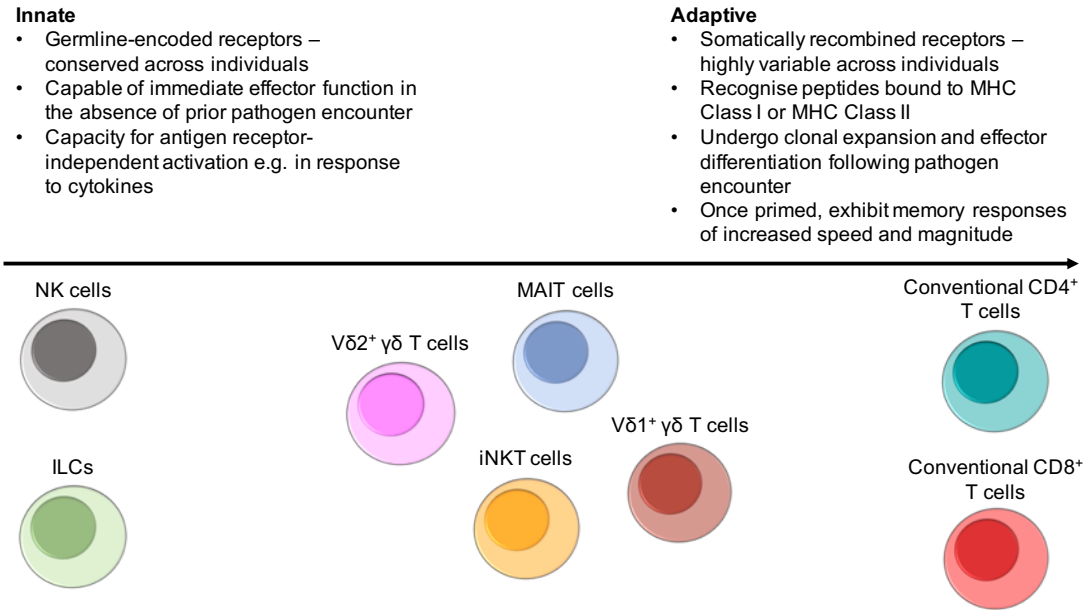


Figure 1.1. Diagram illustrating the functional relationship between lymphocyte populations. At one end of the spectrum fall the ILCs and NK cells that express germline-encoded receptors, display immediate effector function following activation in the absence of prior pathogen encounter, and can be activated in an antigen receptor-independent manner, for example by cytokines. At the other end of the spectrum fall conventional $CD4^+$ and $CD8^+$ T cells that recognise short peptide antigens presented by MHC Class II and MHC Class I, respectively. Conventional T cells of a particular antigen specificity are present at low frequency prior to antigen encounter, following which they undergo clonal expansion and differentiate into effector cells. Innate-like T cells, including MAIT cells, iNKT cells, Vδ1 γδ T cells and Vδ2 γδ T cells (Chitadze et al., 2017; Davey et al., 2018; Garner et al., 2018; Godfrey et al., 2019), show properties of both innate and adaptive immune cells, and hence fall in the middle of the spectrum, with their exact order dependent on the cell properties analysed. Similar to conventional T cells, innate-like T cells express a TCR, however their TCR repertoire typically shows limited variability and antigen specificity. In contrast, analogous to innate cells, innate-like T cells acquire effector function within the thymus prior to peripheral pathogen encounter and can be activated in the absence of specific antigen. Through semi-invariant TCRs, MAIT cells and iNKT cells recognise a limited array of ligands bound to MHC Class I-like molecules, namely riboflavin metabolites bound to MR1 and lipid antigens bound to CD1d, respectively. In contrast, γδ T cells are not MHC-restricted and their ligands are relatively poorly characterised. Vδ1 γδ T cells recognise diverse ligands including stress-induced molecules. In adult peripheral blood, the Vδ2 chain is almost exclusively paired with Vγ9 and Vδ2Vγ9 T cells recognise phosphoantigens via a semi-invariant TCR. In contrast, the TCR repertoire of Vδ1 γδ T cells is unfocussed in cord blood, with subsequent expansion and naïve-to-effector differentiation of specific clonotypes, more akin to conventional T cells. Finally, innate-like T cells can be activated in the absence of TCR stimulation, for example via cytokines for MAIT cells, iNKT cells and Vδ2 γδ T cells.

(Constantinides et al., 2019; Salou et al., 2018). Human MAIT cells are likely also tissue-resident. They lack expression of the lymph node homing receptors CD62L and CCR7 (Dusseaux et al., 2011) and express the transcription factor (TF) PLZF (Savage et al., 2008). Conventional $CD4^+$ T cells in mice acquire a tissue-resident phenotype following ectopic expression of PLZF (Thomas et al., 2011). In the liver and mucosal tissues, MAIT cells express tissue-resident memory T (T_{RM}) cell markers, including CD69, CD103 and CRTAM (Booth et al., 2015; Salou et al., 2018; Slichter et al., 2016). Additionally, human liver MAIT cells express LFA-1 (Jeffery et al., 2016), and LFA-1/ICAM-1 interactions are required in mice for the retention of iNKT1 and MAIT1 cells in the liver and spleen (but not in the lung) (Salou et al., 2018; Thomas et al., 2011). However, $CCR7^-CD103^-$ MAIT cells have been identified in human thoracic duct lymph at a similar frequency to that in peripheral blood (Voillet et al., 2018). As CCR7 is required for lymph node entry from the blood, the authors suggest that MAIT cells in the lymph must have exited from nonlymphoid tissues. In summary, the evidence suggests that MAIT cells are largely resident in tissues, but particular subsets and/or MAIT cells in certain tissues may be capable of recirculation.

MAIT cell frequency is strongly influenced by environmental factors. A recent study indicates the importance of early life exposure to riboflavin-synthesising microbes in governing the lifetime abundance of tissue MAIT cells (Constantinides et al., 2019). Genetically identical specific pathogen-free (SPF) mice housed in different cages show highly variable MAIT cell frequencies in tissues such as the skin and lung, while frequencies are comparable in mice from the same cage. Moreover, neonatal germ-free (GF) mice colonised with riboflavin-synthesising bacteria or treated with 5-A-RU + methylglyoxal (which react to form 5-OP-RU) develop a MAIT cell frequency comparable to that of SPF mice, while exposure in adulthood fails to

promote MAIT cell development. In humans, MAIT cell frequency is positively correlated with age up until 20-30 years (although some estimates suggest that adult frequencies are reached much earlier e.g. by 6 years of age), presumably due to microbial-driven expansion, and negatively correlated with age thereafter (Ben Youssef et al., 2018; Dusseaux et al., 2011; Gherardin et al., 2018b; Koay et al., 2016; Novak et al., 2014; Walker et al., 2014). MAIT cell frequency declines in the blood in numerous diseases (Toubal et al., 2019), whilst they expand in certain tissues upon infection or microbial exposure (Chen et al., 2017; Meierovics et al., 2013; Wang et al., 2018). While tissue MAIT cell expansion appears to be long-lasting (> 280 days post-infection) (Chen et al., 2017; Wang et al., 2018), the frequency of MAIT cells before and after microbial exposure is highly correlated (Constantinides et al., 2019), suggesting that individuals with a low baseline MAIT cell frequency will also exhibit a lower frequency post-expansion.

In mice, MAIT cell frequency is under considerable genetic control. MAIT cells show differential abundance in different strains of mice (Rahimpour et al., 2015), and increased MAIT cell numbers in CAST/EiJ mice can be mapped to a single genetic locus (Cui et al., 2015). In contrast, the role of genetics in governing human MAIT cell frequency is unclear. The frequency of Vα7.2^{+}CD161^{hi} T cells (a proxy for MAIT cell frequency) shows no correlation in human mothers and neonates, and the correlation in Vα7.2^{+}CD161^{hi} T cell frequency at birth is equally high in monozygotic and dizygotic twins (Ben Youssef et al., 2018). This suggests that environmental factors may dominate over genetic factors in regulating MAIT cell frequency in humans. However, these findings need to be confirmed using the MR1/5-OP-RU tetramer for MAIT cell identification, as MR1/5-OP-RU tetramer^{+} (hereafter MR1/5-OP-RU^{+}) MAIT cells comprise only a small fraction (2-15%) of Vα7.2^{+}CD161^{hi} T cells at birth, in contrast with adults, where Vα7.2^{+}CD161^{hi} T cells are > 92% MR1/5-OP-

RU$^+$ (Ben Youssef et al., 2018). Therefore, further research is required to establish the relative role of genetic and environmental factors in regulating MAIT cell frequency.

1.1.2. TCR usage

Conventional T cells recognise short peptide antigens presented by highly polymorphic MHC Class I or MHC Class II molecules. In contrast, MAIT cells express a semi-invariant TCR that is specific for non-peptide riboflavin metabolite antigens, the most potent of which is 5-OP-RU, bound to the monomorphic MHC Class I-like molecule MR1 (Corbett et al., 2014; Kjer-Nielsen et al., 2012; Treiner et al., 2003) (Table 1).

The semi-invariant αβ TCR of MAIT cells comprises a largely invariant TCRα chain paired with a biased repertoire of Vβ chains (Table 1). In humans, MAIT cells predominantly use a *TRAV1-2*/*TRAJ33* (Vα7.2-Jα33) TCRα chain, with smaller subsets using *TRAV1-2*/*TRAJ12* (Vα7.2-Jα12) or *TRAV1-2*/*TRAJ20* (Vα7.2-Jα20) (Lepore et al., 2014; Reantragoon et al., 2013; Tilloy et al., 1999). The CDR3α region is highly restricted in length (12 amino acids) and sequence, with minor junctional variability (Reantragoon et al., 2013; Tilloy et al., 1999). A conserved tyrosine residue at position 95 of the CDR3α region directly contacts MAIT cell stimulatory ligands, forming a hydrogen bond with the ribityl moiety of riboflavin metabolites, and this interaction is essential for MAIT cell activation (Corbett et al., 2014; Eckle et al., 2014; Patel et al., 2013; Reantragoon et al., 2012). *TRBV* usage is biased towards *TRBV6* and *TRBV20* (Vβ13 and Vβ2) in humans, although a wide range of *TRBV* segments are incorporated into the MAIT cell TCR (Reantragoon et al., 2013; Tilloy et al., 1999). *TRAJ12* and *TRAJ20* TCRα chains may preferentially

pair with *TRBV6* TCRβ chains, although a limited number of cells have been examined so far (Lepore et al., 2014; Reantragoon et al., 2013).

Few studies have explored the MAIT cell TCR repertoire in tissues, although preliminary evidence suggests that TCR chain usage might vary across tissues. For example, Lepore and colleagues identified an increased use of *TRAJ12* in various tissues relative to the blood, including the kidney and intestine (Lepore et al., 2014). Furthermore, while use of *TRAJ33* was predominant in the blood of all donors, the fraction of MAIT cells using *TRAJ12*, *TRAJ20* and *TRAJ33* was variable across donors in buccal mucosa, with dominant use of *TRAJ12* and *TRAJ20* in some cases (Sobkowiak et al., 2018). Finally, the frequency of MAIT cells expressing Vβ2 and Vβ13.2 was reduced in breast tissue compared with blood (Zumwalde et al., 2018).

In the blood, MAIT cells comprise oligoclonal expansions, with the degree of oligoclonality largely comparable to that of conventional memory T (T_{mem}) cells (Ben Youssef et al., 2018; Lepore et al., 2014). The extent of sharing of paired TCRαβ chains between different tissues and between individuals is currently unknown, although a high degree of CDR3β sharing has been shown between matched blood and lymph, in contrast with low sharing between donors (Lepore et al., 2014; Voillet et al., 2018).

1.1.3. Phenotype

In humans, MAIT cells are primarily $CD8^+$ (70-90%), with some $CD4^-CD8^-$ (DN) (10-20%), and a minor population of $CD4^+$ cells (Gherardin et al., 2018b; Martin et al., 2009; Reantragoon et al., 2013) (Table 1). Within the $CD8^+$ subset, MAIT cells predominantly express CD8αα homodimers or are $CD8\alpha^+\beta^{low}$ (Reantragoon et al.,

2013; Walker et al., 2012), in contrast with conventional T cell populations that are mainly CD8αβ^{+} (> 90%). CD8αα homodimers might function to inhibit T cell activation, although their physiological role remains poorly defined (Cheroutre and Lambolez, 2008).

Human MAIT cells display an effector memory phenotype and characteristic expression of several surface molecules (CD161, CD26), cytokine and chemokine receptors (IL-18 receptor [IL-18R], CCR5, CCR6), and TFs (PLZF, RORγt, T-bet) (Table 1) (Kurioka et al., 2016). As their phenotype is largely homogeneous in humans and MR1 tetramers have only recently been developed, human MAIT cells are routinely identified using surrogate markers, most commonly as Vα7.2^{+}CD161^{hi} T cells (Martin et al., 2009), but also using Vα7.2 combined with IL-18Rα (Le Bourhis et al., 2010) or CD26 (Sharma et al., 2015). However, given that it is now readily available through the NIH Tetramer Core Facility, use of MR1/5-OP-RU is becoming more routine. While human MAIT cells coexpress the TFs PLZF, T-bet, and RORγt, their expression is subset specific for mouse MAIT cells, with cells expressing PLZF and either T-bet (MAIT1 cells) or RORγt (MAIT17 cells; Table 1) (Koay et al., 2016; Leeansyah et al., 2015; Rahimpour et al., 2015).

In summary, MAIT cells express a semi-invariant TCR, recognise non-peptide riboflavin metabolite antigens and exhibit an innate-like effector memory phenotype. Their frequency is strongly influenced by environmental factors, while the role of genetic factors in humans remains unclear. MAIT cells predominantly localise to peripheral tissues under homeostatic conditions, especially the liver and mucosal tissues, and are therefore optimally positioned to act as a first line of defence at the site of microbial infection.

1.2. MAIT cell definition

For the purposes of this thesis, peripheral blood and tissue MAIT cells will be defined as T cells: (1) expressing a semi-invariant Vα7.2-Jα33/12/20 TCR (Vα19-Jα33 in mice) with a tyrosine 95 residue in the CDR3α region, (2) restricted by MR1 bound to microbial riboflavin metabolites, (3) selected on $CD4^+CD8^+$ (DP) thymocytes, and (4) exhibiting an innate-like and mixed type 1/type 17 phenotype and function in humans, or comprising distinct MAIT1 and MAIT17 subsets in mice.

1.3. Mouse models

1.3.1. MR1 knockout

$MR1^{-/-}$ mice lack MAIT cells (Treiner et al., 2003), and comparison of $MR1^{-/-}$ with $MR1^{+/+}$ mice is commonly employed to investigate the function of MAIT cells, often in disease settings. However, as well as lacking MAIT cells, $MR1^{-/-}$ mice will also lack any other MR1-restricted T cells. A population of functionally diverse MR1-restricted non-MAIT T cells are present in humans (Lepore et al., 2017), while a population of mature naïve $CD24^-CD44^-CD62L^+$ MR1/5-OP-RU^+ T cells (predominantly $CD4^+$) have been identified in mouse thymus and spleen (~4% of splenic MR1/5-OP-RU^+ T cells) (Legoux et al., 2019a). This should be taken into consideration when using $MR1^{-/-}$ mice, as any identified phenotypes may not be entirely attributable to MAIT cells. Additionally, the microbiota of $MR1^{-/-}$ mice is distinct from that of wild-type (WT) mice (Smith et al., 2019; Varelias et al., 2018) and therefore cohousing experiments should be performed to address the impact of the altered microbiota in any given experimental model.

1.3.2. TCR transgenic

Transgenic mouse models have been widely used to study the phenotype and function of MAIT cells, as well as their role in different disease settings. While use of these models has provided major contributions to our understanding of MAIT cell biology, it is also important to be aware of their limitations.

Mice that constitutively express the MAIT cell Vα19-Jα33 invariant TCRα chain (termed Vα19i/iVα19 in several studies) have been generated on a $C\alpha^{-/-}$ background (Croxford et al., 2006; Kawachi et al., 2006; Martin et al., 2009; Okamoto et al., 2005). As intended, these mice have an increased frequency of MAIT cells, however normal T cell development is dysregulated. TCR diversity is greatly reduced, T cell numbers are significantly decreased in the thymus and many peripheral tissues, and the mice have an expanded population of DN T cells. As mice overexpressing the MAIT cell TCRα chain also harbour other T cell populations, comparison of mice deficient and sufficient in MR1 is necessary in order to identify features specific to MR1-restricted MAIT cells (Kawachi et al., 2006; Martin et al., 2009). Along with global changes in T cell development, MAIT cells in TCRα transgenic mice exhibit certain differences in their phenotype, function, subset distribution and tissue localisation compared with their WT counterparts. MAIT cells from iVα19 TCRα transgenic mice display a naïve phenotype, lack expression of PLZF, and secrete considerable amounts of IL-10 and Th2 cytokines (Croxford et al., 2006; Cui et al., 2015; Kawachi et al., 2006; Martin et al., 2009), in contrast with MAIT cells from WT mice (Rahimpour et al., 2015).

Vβ transgenic mice, for example Vβ6 and Vβ8 transgenic mice, can be studied as an alternative to TCRα transgenics, or can be crossed with TCRα transgenics to

further increase MAIT cell frequency and to decrease unwanted TCR specificities (Martin et al., 2009). An important limitation of these models is that, as MAIT cells utilise multiple TCRβ chains, the forced use of a single Vβ chain may bias the antigen specificity and thereby the functionality of the generated population. In addition to the use of double transgenics, MAIT cell frequency can be increased by studying transgenic mice on a $RAG^{-/-}$ or $TAP^{-/-}Ii^{-/-}$ background (Martin et al., 2009). However, in these mice, interactions between MAIT cells and other conventional T cells (and B cells in $RAG^{-/-}$ mice), which might influence their phenotype and development in a WT setting, are completely absent.

1.3.3. Non-transgenic

Given the scarcity of MAIT cells in WT mice and the limitations of TCR transgenic models, alternative models with increased MAIT cell frequency have been developed. A mouse strain (CAST/EiJ) with a 20-fold greater frequency of MAIT cells than C57BL/6 mice was identified, and crossing this strain onto a C57BL/6 background generated a B6-$MAIT^{Cast}$ strain with an increased frequency of MAIT cells (Cui et al., 2015). These $MAIT^{Cast}$ cells display a phenotype more consistent with MAIT cells from WT animals, including expression of PLZF, but some phenotypic and functional abnormalities may remain. B6-$MAIT^{Cast}$ mice also have the added advantage of expressing a RORγt-GFP reporter, making it straightforward to study mouse MAIT1 ($ROR\gamma t^{-}$) and MAIT17 ($ROR\gamma t^{+}$) MR1/5-OP-RU^{+} T cells (Cui et al., 2015; Salou et al., 2018).

An alternate, non-genetic approach to increase the frequency of MAIT cells in mice is through the intranasal administration of MR1 ligand (5-OP-RU) combined with a Toll-like receptor (TLR) agonist, which increases their frequency to approximately

50% of lung αβ T cells (Chen et al., 2017). MAIT cell frequency in the lung can also be artificially increased through intranasal infection with the live-attenuated vaccine strain *Salmonella enterica* serovar Typhimurium (*S*. Typhimurium) BRD509 (Chen et al., 2017; van Wilgenburg et al., 2018) or with *Legionella longbeachae* (*L*. *longbeachae*) (Wang et al., 2018). Further work is required to understand how this "boosting" may impact on the phenotype and function of MAIT cells, and thereby to establish the robustness of this experimental approach. Regardless of potential shortcomings, efforts to develop mouse models with increased MAIT cell frequencies, while avoiding the limitations of TCR transgenic systems, appear promising.

Overall, whilst transgenic mouse models enable the role of MAIT cells to be interrogated *in vivo* in health and disease, caution is necessary when using these models. Results from any mouse model should be validated in other models to avoid findings that result from peculiarities of the chosen experimental system. Study of $MR1^{+/+}$ and $MR1^{-/-}$ $MAIT^{Cast}$ mice provides the best current model for the mechanistic analysis of MAIT cell function, given the increased frequency of MAIT cells in these mice and their similar phenotype/function to MAIT cells from C57BL/6 mice. Lastly, it is important to bear in mind that discoveries from mouse studies may not directly translate to humans, given the vastly different frequencies of MAIT cells in these species, in addition to other differences, for example the presence of functional subsets in mice but not in humans.

1.4. Development

1.4.1. Thymic development

Selection

As with conventional T cells, the semi-invariant TCR of MAIT cells is generated via random recombination (Greenaway et al., 2013; Tilloy et al., 1999). However its formation requires an extended DP thymocyte lifespan. Initial rearrangement of the TCRα locus utilises 3' Vα and 5' Jα segments, with later rearrangements using progressively more 5' Vα segments and more 3' Jα segments (termed proximal to distal rearrangement) (Guo et al., 2002; Huang and Kanagawa, 2001). Thus, formation of the MAIT cell semi-invariant TCR, which incorporates the 5' most Vα segment (*TRAV1-2*), occurs late in the lifespan of DP thymocytes. RORγt enhances the survival of DP thymocytes by increasing the expression of the anti-apoptotic protein Bcl-xL (Sun et al., 2000). As a result, DP thymocyte lifespan is reduced in *Rorc*$^{-/-}$ (RORγt-deficient) mice. In peripheral blood T cells from *RORC*$^{-/-}$ patients, TCRs incorporating 5' Vα and 3' Jα TCRα segments are absent, and hence the patients lack MAIT cells (Okada et al., 2015), presumably due to lack of rearrangement of their characteristic TCRα chain at the DP thymocyte stage.

Following TCR expression, conventional T cells undergo positive selection on cortical thymic epithelial cells (TEC) that present self-peptides on MHC Class I and MHC Class II. In contrast, MAIT cells are selected by MR1-expressing DP thymocytes (Constantinides et al., 2019; Gold et al., 2013; Seach et al., 2013; Wang et al., 2019) (Fig. 1.2A). A recent study suggests that exogenous riboflavin metabolites such as 5-OP-RU, derived from commensal bacteria, are the MAIT cell-

selecting antigens presented by MR1 (Legoux et al., 2019a). The frequency of MR1/5-OP-RU^{+} T cells, including newly selected CD24^{+}CD44^{-} cells, is greatly decreased in the thymus of GF mice and CD24^{+}CD44^{-} cells show lower Nur77 expression, indicative of reduced TCR signalling (Koay et al., 2016; Legoux et al., 2019a). Monocolonisation of GF mice with *Escherichia coli* (*E. coli*) induces MAIT cell development, but only if the *E. coli* strain is capable of 5-A-RU (undergoes non-enzymatic condensation with methylglyoxal to form 5-OP-RU) production (Legoux et al., 2019a). Moreover, 5-OP-RU injected intraperitoneally or painted onto the skin of mice can be captured and presented by a variety of thymic antigen-presenting cells (APC), including DP thymocytes. Overall the data is strongly suggestive of a role of 5-OP-RU in MAIT cell selection, although it remains possible that 5-OP-RU contributes instead to the expansion of previously selected immature thymocytes.

Conventional T cells are positively selected in the thymus when their TCR exhibits moderate affinity for MHC/self-peptide, whilst thymocytes expressing high affinity TCRs are removed from the repertoire. Selection of MAIT cells by the potent antigen 5-OP-RU would suggest that MAIT cells undergo agonist selection, consistent with a number of unconventional T cell lineages, including iNKT cells, regulatory T cells, and CD8αα gut intraepithelial T cells (Stritesky et al., 2012). Additional evidence supports MAIT cell agonist selection, as outlined below. Compared with conventional thymocytes, the high avidity interaction between the iNKT TCR and selecting glycolipids results in strong TCR signalling (Moran et al., 2011) and therefore prolonged upregulation of the TCR-induced TFs Egr1 and Egr2 (Seiler et al., 2012). The TF PLZF, encoded by *Zbtb16*, is a direct downstream target of Egr2 (Seiler et al., 2012), and MAIT cells upregulate PLZF expression during thymic development (Koay et al., 2016), contrasting with conventional T cells. Moreover, mouse MAIT cells upregulate CD44 expression, and mouse and human

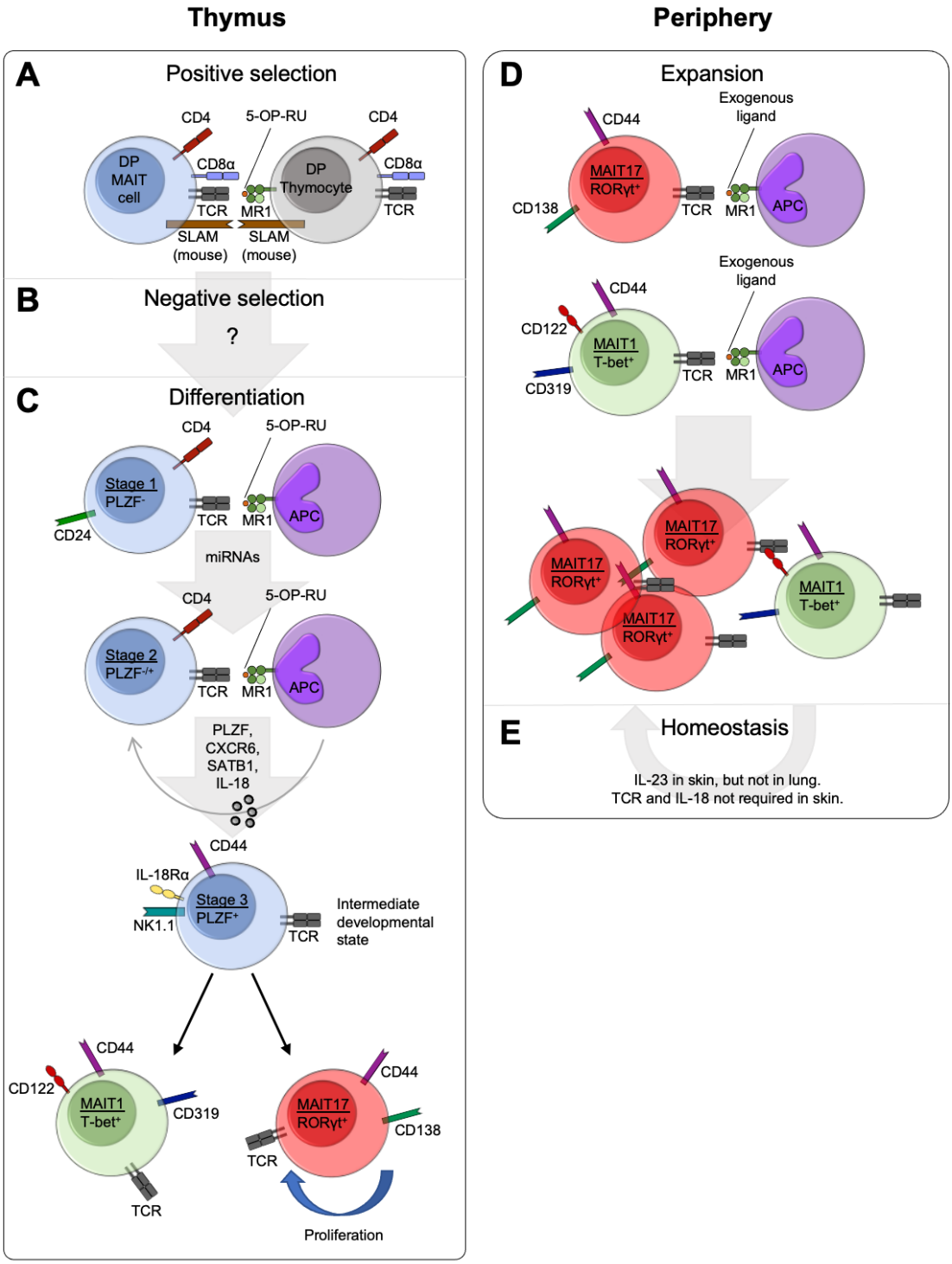


Figure 1.2. MAIT cell thymic and peripheral development in mice. (A) MAIT cells are positively selected by MR1/5-OP-RU on DP (double positive; $CD4^+CD8^+$) thymocytes. Concomitant with TCR-MR1/5-OP-RU binding, homotypic interactions between signalling lymphocyte activation molecule (SLAM) family receptors are required for the selection of effector MAIT cells. **(B)** The role of negative selection in MAIT cell development is currently unknown. **(C)** Following selection, MAIT cells differentiate through three stages of development defined by the expression of CD24 and CD44 (stage 1: $CD24^+CD44^-$, stage 2: $CD24^-CD44^-$, stage 3: $CD24^-CD44^+$), with effector capacity acquired at stage 3. A number of factors are required for differentiation including miRNAs, PLZF and CXCR6, as well as MR1.

Recent transcriptional data suggests that MAIT1 and MAIT17 cells arise from a shared precursor population – the exact nature of this population remains to be defined. MAIT17 cells are highly proliferative in the thymus, in contrast with MAIT1 cells, likely contributing to the predominance of MAIT17 compared with MAIT1 cells in peripheral tissues. **(D)** MAIT cells exit the thymus with a $CD24^{-}CD44^{+}$ memory phenotype and undergo expansion in the periphery, most likely driven by the presentation of exogenous ligands from commensal bacteria. **(E)** Recent evidence suggests a role for IL-23 in MAIT cell homeostasis in the skin, while IL-18 and TCR stimulation are not required. In contrast to the skin, IL-23 is not essential for MAIT cell homeostasis in the lung. Thus, the signals mediating MAIT cell homeostasis may be tissue-specific.

MAIT cells can acquire effector function within the thymus (Koay et al., 2016), properties of antigen-experienced conventional T cells.

In addition to the TCR-MR1/ligand interaction, development of effector MAIT cells in mice requires SLAM-associated protein (SAP) (Koay et al., 2019a; Legoux et al., 2019b), an intracellular adaptor that is activated downstream of homotypic interactions between SLAM family receptors expressed on DP thymocytes (Griewank et al., 2007). Mechanistically, SAP signalling increases the expression of Egr2 (downstream of TCR signalling) in thymocytes undergoing selection, thus enabling the expression of *Zbtb16*, a direct target of the Egr2 TF (Dutta et al., 2013; Legoux et al., 2019b; Seiler et al., 2012). SAP-deficient mice have a significantly reduced frequency of MAIT cells in the thymus (5-12-fold) and spleen, while MAIT cells are absent in nonlymphoid tissues (Koay et al., 2019a; Legoux et al., 2019b). MAIT cell frequency is normal in SAP-deficient patients and the cells express normal levels of *ZBTB16* (Martin et al., 2009), suggesting the role of SAP may differ in mice and humans.

Differentiation

Following selection, MAIT cells transition through three stages of development in mice (Fig. 1.2C) and humans. In mice, these stages are defined by the expression of CD24 and CD44 (stage 1: $CD24^{+}CD44^{-}$, stage 2: $CD24^{-}CD44^{-}$, stage 3: $CD24^{-}CD44^{+}$) (Koay et al., 2016). In human thymus, stage 1 cells are defined as $CD161^{-}CD27^{-}$, stage 2 as $CD161^{-}CD27^{+}$ and stage 3 as $CD161^{+}CD27^{pos\text{-}lo}$. Many of the genes that are modulated in expression from stage 1 to stage 3 are consistent in humans and mice, including upregulation of *ZBTB16*, *IL18R1*, *CXCR6* and *IL7R*, and downregulation of *CD4*, *LEF1*, *TCF7* and *BACH2* (Koay et al., 2019a). However, the two developmental pathways diverge at stage 3.

In humans, stage 3 cells coexpress T-bet and RORγt (Koay et al., 2016) and are not thought to comprise multiple subsets. In contrast, mature thymic MAIT cells in mice comprise two transcriptionally distinct differentiated populations, namely T-bet-expressing MAIT1 cells and RORγt-expressing MAIT17 cells (Fig. 1.2C) (Koay et al., 2016; Legoux et al., 2019b). Single-cell RNA-seq (scRNA-seq) data suggests that MAIT1 and MAIT17 cells arise from a shared intermediate developmental state. In one recent study, this population (named clusters 7a and 7b) was marked by the expression of *Zbtb16*, *S1pr1*, *Klf2* and *Ccr9* (Legoux et al., 2019b), and in another these cells were identified as $CD4^{+}LEF1^{+}CD319^{-}CD138^{-}CD24^{-}CD44^{+}$ thymocytes (Koay et al., 2019a). The relationship between these populations and the specific signals required for commitment to the MAIT1 or MAIT17 lineage remain to be determined. While MAIT17 development is more dependent on CCR7 (Koay et al., 2019a) and 5-OP-RU derived from commensal bacteria (Legoux et al., 2019a), whether these signals affect lineage commitment is currently unknown.

Development of MAIT cells along the linear developmental pathway requires a number of different factors (Fig. 1.2C). MR1 is essential for all stages of MAIT cell development *in vitro*, and thymic and peripheral MAIT cells are almost absent in MR1-deficient mice (Koay et al., 2016; Legoux et al., 2019b; Martin et al., 2009; Treiner et al., 2003). The frequency of MAIT cells at all stages of development is significantly reduced in GF mice, with only a few cells reaching stage 3 (Koay et al., 2016; Legoux et al., 2019a; Treiner et al., 2003). MAIT cell development beyond stage 1 requires microRNAs, as the abundance of stage 2 and stage 3 MAIT cells is decreased in *Drosha*-deficient mice (Koay et al., 2016). Evidence suggests a key role for miR-181a/b-1, a single pair of microRNAs known to augment TCR signal strength, as miR-181a/b-$1^{-/-}$ mice exhibit a phenotype similar to *Drosha*-deficient mice (Li et al., 2007; Winter et al., 2019). PLZF is necessary for MAIT cell

maturation from stage 2 to stage 3 and for their acquisition of effector function (Koay et al., 2016; Rahimpour et al., 2015). Other factors important for development from stage 1 to stage 3 include CXCR6, SATB1, ZAP70 and IL-18 (Koay et al., 2019a; 2016). The exact developmental stage at which some of these factors function may need revisiting given the recent identification of a mature naïve MR1/5-OP-RU^{+} T cell population in the thymus that is selected on TECs (Legoux et al., 2019b). These cells lack expression of CD24 and CD44, and thus could be incorrectly identified as stage 2 cells.

Mouse MAIT cells accumulate in the thymus with age, accompanied by an increase in the proportion of stage 3 cells (Koay et al., 2016). Expression of Ki67, which indicates actively proliferating cells, is largely absent in stage 1, while it is expressed by a large fraction of stage 2 cells and by almost all stage 3 cells (Winter et al., 2019). Thymic MAIT17 cells express Ki67 and gene signatures associated with proliferation, receive strong TCR signals in the thymus, and proliferate in a manner dependent on phosphatidylinositol 3-kinase (Legoux et al., 2019a; 2019b). In contrast, MAIT1 cells predominantly survive, rather than proliferate, in the thymus. In humans, thymic MAIT cells are present at a low frequency irrespective of age (Gold et al., 2013; Koay et al., 2016; Martin et al., 2009), and T cell receptor excision circle (TREC) analysis of MAIT cells in human thymus and cord blood identified no differences in TREC concentration compared with conventional T cells (Gold et al., 2013). This suggests that human MAIT cells may not undergo intrathymic proliferation, but this requires further study.

Acquisition of innate-like effector function

MAIT cells can acquire innate-like effector function in the thymus and secrete cytokines upon activation (Koay et al., 2016). Expression of the TF PLZF is

necessary and sufficient to drive innate-like effector differentiation (Koay et al., 2016; Raberger et al., 2008; Savage et al., 2008). During thymic MAIT cell development, PLZF expression begins at stage 2 (mouse: $CD24^{-}CD44^{-}$, human: $CD161^{-}CD27^{+}$) and is highest at stage 3 (mouse: $CD24^{-}CD44^{+}$, human: $CD161^{+}CD27^{pos\text{-}lo}$) (Fig. 1.2C) (Koay et al., 2016). Therefore, MAIT cells acquire effector function at stage 3. In PLZF-deficient mice, MAIT cell development is almost completely abrogated and residual cells exhibit a $CD44^{-}$ phenotype, reduced expression of characteristic phenotypic markers and impaired cytokine secretion (Koay et al., 2016; Rahimpour et al., 2015).

Although they can acquire effector function within the thymus, it is unclear whether the majority of human MAIT cells exit the thymus at stage 2 or stage 3. Naïve stage 2 MAIT cells predominate in human thymus and are found to a lesser degree in cord blood (~20%) and young blood (~10%; donors aged 5 days-14 years) (Koay et al., 2016), indicating that at least some MAIT cells exit the thymus at stage 2. Moreover, peripheral blood MR1/5-OP-RU^{+} T cells are naïve at birth, but rapidly acquire a memory phenotype, with > 80% of MR1/5-OP-RU^{+} T cells expressing CD45RO by one month of age (Ben Youssef et al., 2018). Thus, further studies are required to fully define exactly when and where human MAIT cells acquire their effector memory phenotype and function. In contrast with humans, stage 3 mature MAIT cells are the main population in mouse thymus (Koay et al., 2016), suggesting that the majority of mouse MAIT cells exit the thymus as $CD44^{+}$ memory cells.

In conclusion, MAIT cells are selected on DP thymocytes, develop through analogous stages in mice and humans, and can acquire innate-like effector function within the thymus. While mouse MAIT cells likely exit the thymus as $CD44^{+}$ effector memory cells, further studies are required to clarify the exact developmental stage

of thymic exit in humans. Stage 3 MAIT cells in mouse thymus are comprised of differentiated T-bet^{+} MAIT1 cells and RORγt^{+} MAIT17 cells, while subsets of human MAIT cells have not been identified. The reason for this dichotomy between mouse and human MAIT cell development is currently unclear.

1.4.2. Peripheral development

Changes in abundance

Whilst their frequency is relatively constant in human thymus, MAIT cells undergo a large population expansion in the periphery, increasing over 100-fold from < 0.01% of T cells at birth to 1-10% of T cells by adulthood (Ben Youssef et al., 2018; Dusseaux et al., 2011; Gherardin et al., 2018b; Koay et al., 2016; Novak et al., 2014; Walker et al., 2014). The increase in MAIT cell numbers occurs over a number of years, although estimates for the age at which adult frequencies are reached vary between studies (6-30 years of age). In mice, MAIT cells are absent in barrier tissues at two weeks of age but by three weeks of age they are present in all barrier tissues, and in some tissues such as the lung they have already reached adult frequencies (Constantinides et al., 2019).

Peripheral population expansion in mice is dependent on early life exposure to riboflavin metabolites from commensal bacteria (Constantinides et al., 2019). MAIT cell frequency is markedly reduced in the thymus and peripheral tissues of GF mice (Koay et al., 2016; Legoux et al., 2019a; Treiner et al., 2003). Following colonisation of neonatal GF mice with an early-life microbial community, MAIT cells accumulate in tissues to a frequency comparable to SPF mice (Constantinides et al., 2019). A similar expansion is achieved following monocolonisation with a

riboflavin-synthesising bacterial species, but not with a species incapable of 5-OP-RU production. In contrast, colonisation of adult GF mice with early-life riboflavin-synthesising bacteria does not induce the development of tissue MAIT cells (although stage 3 thymic MAIT cells are induced). While, early life exposure is required for the initial development of tissue MAIT cells, local exposure to riboflavin-synthesising bacteria can induce further expansion. For example, upon intranasal infection of C57BL/6 mice with *S*. Typhimurium, MAIT cells expand from 0.5-2% of lung αβ T cells in uninfected mice to 25-50% at one week post-infection (Chen et al., 2017). Local expansion of MAIT cells in the skin can be induced by topical application of the skin commensal bacteria *Staphylococcus epidermidis* (*S. epidermidis*) and is independent of lymph node priming (Constantinides et al., 2019). Interestingly, 5-OP-RU alone is sufficient to induce expansion in the skin, while lung expansion requires 5-OP-RU plus a costimulus such as a TLR ligand, suggesting that the requirements for local MAIT cell expansion might differ between tissues.

Peripheral MAIT cell expansion in humans is also thought to be driven by the presentation of microbial riboflavin metabolite antigens on MR1, derived from commensal and/or pathogenic bacteria (Fig. 1.2D). Along with the mouse data discussed above, a variety of evidence supports this hypothesis. MAIT cells are naïve in human thymus, cord blood, and in newborns, but rapidly acquire a memory phenotype in the blood during the first month of life (Ben Youssef et al., 2018; Dusseaux et al., 2011; Koay et al., 2016; Martin et al., 2009), concomitant with exposure to bacteria. Human MAIT cells proliferate *in vitro* in response to riboflavin-synthesising bacteria (Kurioka et al., 2015), similar to the local expansion of MAIT cells in mouse tissues. Moreover, studies of the MAIT cell TCR repertoire support microbe-mediated expansion. MAIT cells with distinct TCRs are activated *in vitro*

following stimulation with different bacteria (Gold et al., 2014), and in a human *in vivo Salmonella enterica* serovar Paratyphi A (*S*. Paratyphi A) challenge setting, the relative abundance of different MAIT cell TCR clonotypes changes in response to infection (Howson et al., 2018). Vα7.2^{+}CD161^{hi} T cells exhibit a polyclonal TCR repertoire in cord blood compared with an oligoclonal repertoire in adult blood (Ben Youssef et al., 2018; Lepore et al., 2014; Tilloy et al., 1999; Walker et al., 2012), consistent with the hypothesised expansion of specific clones in response to particular bacteria. However, the findings from cord blood need to be replicated using the MR1/5-OP-RU tetramer, as only 2-15% of cord blood Vα7.2^{+}CD161^{hi} T cells are MR1/5-OP-RU^{+}, compared with > 92% of Vα7.2^{+}CD161^{hi} T cells in adult blood (Ben Youssef et al., 2018). An initial comparison of MR1/5-OP-RU^{+} compared with MR1/5-OP-RU^{-} Vα7.2^{+}CD161^{hi} T cells from the cord blood of four donors indicates some differences in the TCR repertoire of these populations, but further studies are required to validate these findings.

B cells were thought to be necessary for peripheral MAIT cell expansion in mice, due to a significant reduction in *TRAV1*/*TRAJ33* (mouse MAIT cell invariant TCRα chain) transcripts in DN TCRαβ^{+} cells from the mesenteric lymph nodes and gut lamina propria of B cell-deficient ($J_H^{-/-}$) mice, along with the expansion of MAIT cells in iVα19/Vβ6 RAG$^{-/-}$ mice following B cell transfer (Martin et al., 2009; Treiner et al., 2003). However, recent studies showed WT numbers of MR1/5-OP-RU^{+} MAIT cells in the thymus, spleen and lung of $J_H^{-/-}$ B6-MAIT^{Cast} mice (Legoux et al., 2019a) and in the skin of mice lacking MR1 expression on B cells (Constantinides et al., 2019). The role of B cells in human MAIT cell development remains uncertain. Treiner and colleagues observed a reduced frequency of MAIT cells (as measured by the presence of *TRAV1-2/TRAJ33* TCRα transcripts) in the blood of patients that lack B cells due to a mutation in Bruton tyrosine kinase (*BTK*) (Treiner et al., 2003).

However, only four patients were analysed, one of which had a normal number of MAIT cells. A study of common variable immunodeficiency (CVID) provides indirect evidence against a role for B cells in regulating human MAIT cell frequency. Although the abundance of B cells and MAIT cells was variably decreased in CVID patients, the frequency of MAIT cells showed no correlation with that of B cells (Paquin-Proulx et al., 2017). A major confounding factor in these human studies is the increased occurrence of infections in patients with *BTK* deficiency and CVID, which can independently modulate MAIT cell frequency. Thus, the role of B cells, if any, in mouse and human MAIT cell development, requires clarification.

Phenotypic and functional maturation

In addition to expansion, MAIT cells undergo maturation in the periphery, as indicated by marked phenotypic changes. Approximately half of MAIT cells in the thymus are either DP or $CD4^+$, whereas MAIT cells in adult blood are predominantly DN or $CD8^+$ (Gherardin et al., 2018b; Koay et al., 2016). Furthermore, $CD8^+$ MAIT cells in the thymus and cord blood are $CD8\alpha\beta^+$, whereas roughly half of $CD8\alpha^+$ MAIT cells in adult blood express CD8αα homodimers (Koay et al., 2016; Walker et al., 2012). $CD8\alpha\alpha^+$ MAIT cells are thought to arise from $CD161^{hi}CD8\alpha\beta^+$ cells in the periphery in the first few months of life (Ben Youssef et al., 2018; Walker et al., 2012). Following thymic exit, MAIT cells acquire a memory $CD45RO^+$ phenotype and upregulate the expression of characteristic phenotypic markers, such as CD161 and IL-18Rα, whilst downregulating the expression of lymph node homing receptors, including CD62L and CCR7 (Ben Youssef et al., 2018; Dusseaux et al., 2011; Gold et al., 2013; Koay et al., 2016; Leeansyah et al., 2014; Walker et al., 2012). These changes are gradual, as MAIT cells in cord blood, young blood, and adult blood exhibit an increasingly mature phenotype.

As well as phenotypic changes, MAIT cells undergo further functional differentiation following thymic exit. Although stage 3 MAIT cells in human thymus can produce IFNγ and TNFα following PMA and ionomycin stimulation, their capacity to do so is significantly decreased compared with peripheral blood MAIT cells (Koay et al., 2016). Moreover, some human MAIT cells may exit the thymus at stage 2 before they acquire effector capacity. In contrast with adult blood Vα7.2^{+}CD161^{hi} T cells, cord blood Vα7.2^{+}CD161^{hi} T cells are unable to produce IFNγ or granzyme B in response to overnight stimulation with *E. coli*-infected THP-1 cells (Ben Youssef et al., 2018). However, as MAIT cells comprise the majority of Vα7.2^{+}CD161^{hi} T cells in adult blood, but only a small fraction of Vα7.2^{+}CD161^{hi} T cells in cord blood (Ben Youssef et al., 2018), whether this finding applies to MAIT cells needs to be confirmed using the MR1 tetramer. In contrast with human MAIT cells, mouse thymic MAIT cells strongly produce cytokines (Koay et al., 2016), suggesting possible species-specific differences in when cytokine-producing capacity is acquired.

A recent transcriptional analysis suggests that human MAIT cells may undergo more extensive peripheral maturation than in mice (Koay et al., 2019a). Human stage 3 MAIT cells from thymus and young blood showed considerable transcriptional differences, with genes upregulated in the blood including *TGFB1*, *IRF1* and *CXCR4*, and downregulated genes including *LEF1*, *CCR7* and *EOMES*. In contrast, only 19 genes were differentially expressed between stage 3 MAIT cells from mouse thymus and spleen. The reason for this difference between mouse and human MAIT cells is unclear, but is consistent with the potent cytokine-producing capacity of mouse thymic MAIT cells.

Homeostasis

The requirements for MAIT cell proliferation and survival in the periphery are poorly understood (Fig. 1.2E). T_{mem} cells depend predominantly on cytokines for peripheral maintenance (Surh and Sprent, 2008), suggesting cytokines rather than TCR stimulation, might be key for MAIT cell homeostasis. Consistent with this hypothesis, tonic TCR signalling is not required for the maintenance of skin MAIT cells, as their frequency remains unchanged three weeks following ubiquitous Cre-mediated deletion of MR1 (Constantinides et al., 2019). While MR1 is widely expressed by haematopoietic and non-haematopoietic cells (Hashimoto et al., 1995; Riegert et al., 1998), it is largely retained in the endoplasmic reticulum prior to ligand exposure (McWilliam et al., 2016), further suggesting its role in MAIT cell homeostasis may be limited.

MAIT cells show high sensitivity to cytokine stimulation (Leeansyah et al., 2015; Sattler et al., 2015; Slichter et al., 2016; Tang et al., 2013; Ussher et al., 2014; van Wilgenburg et al., 2016) and can proliferate in response to IL-2, IL-15 and to a lesser degree IL-12 (Kurioka et al., 2015). Cytokines that signal via STAT3 are required for human MAIT cell development and/or homeostasis, as indicated by the four-fold reduction in their frequency in patients with heterozygous loss-of-function mutations in *STAT3* (Wilson et al., 2015). The central role of STAT3 appears to be downstream of the IL-23 receptor (IL-23R) and possibly the IL-21 receptor. Consistent with a role for IL-23 in MAIT cell development or homeostasis, IL-23R-deficient mice exhibit significantly reduced MAIT cell frequency in the skin (Constantinides et al., 2019). However, mice deficient in either subunit of IL-23 (*Il-12p40*$^{-/-}$ and *Il-23p19*$^{-/-}$) have normal numbers of MAIT cells in the lung under steady state conditions (Wang et al., 2019). IL-18-deficient mice exhibit reduced thymic

and peripheral MAIT cell frequencies (Koay et al., 2016), suggesting IL-18 is necessary for MAIT cell development and/or survival. However, steady-state MAIT cell frequency in mouse skin was unaffected by anti-IL-18 treatment (Constantinides et al., 2019) and MAIT cell development is normal in IL-18Rα-deficient mice (Koay et al., 2016). Therefore, further work is necessary to determine the role of IL-23 and IL-18 in regulating MAIT cell frequency, and to establish the requirement for other cytokines in MAIT cell homeostasis. Following infection with *S*. Typhimurium, MAIT cells are skewed towards a MAIT17 phenotype in *Il-23p35*$^{-/-}$ and *Il-12p40*$^{-/-}$ mice, and a MAIT1 phenotype in *Il-23p19*$^{-/-}$ mice (Wang et al., 2019). However, whether MAIT cell subsets are differentially regulated by cytokines under homeostatic conditions, remains to be determined.

In summary, MAIT cells undergo a large population expansion following thymic exit that, at least in mice, is dependent on exposure of neonates to riboflavin-synthesising bacteria. As well as increasing in abundance, MAIT cells undergo further phenotypic and functional maturation in the periphery that appears to be more extensive in humans compared with mice.

1.4.3. Foetal development

Human MAIT cells develop in foetal thymus, and can be identified in both lymphoid and non-lymphoid peripheral tissues in the second trimester of foetal development (Leeansyah et al., 2014). Given that the developmentally related iNKT cell population develops postnatally in mouse thymus (Bendelac et al., 1994; Pellicci et al., 2002), it is likely that MAIT cells also undergo postnatal development in mice.

Before discussing human foetal MAIT cell development, it is important to note that foetal MAIT cell development has so far only been investigated using the surrogate MAIT cell markers Vα7.2 and CD161. As previously mentioned, the MAIT cell populations defined as MR1/5-OP-RU^{+} or Vα7.2^{+}CD161^{hi} are essentially the same in adult blood, while thymic MR1/5-OP-RU^{+} MAIT cells are predominantly CD161^{-} (Koay et al., 2016) and MR1/5-OP-RU^{+} MAIT cells comprise < 20% of blood Vα7.2^{+}CD161^{hi} T cells at birth (Ben Youssef et al., 2018). Therefore, findings from the study of Vα7.2^{+}CD161^{hi} T cell development in foetal tissues may not accurately reflect the developmental pathway of MAIT cells.

As in adults, Vα7.2^{+}CD161^{hi} T cells are enriched in foetal peripheral tissues, including the lung, liver, and small intestine, with lower frequencies in the thymus and secondary lymphoid organs (Leeansyah et al., 2014). The frequency of Vα7.2^{+}CD161^{hi} T cells in foetal tissues is low compared with the corresponding adult tissues, particularly in the liver where they are ~100-fold less frequent in the foetus (Dusseaux et al., 2011; Leeansyah et al., 2014). In all foetal tissues, Vα7.2^{+}CD161^{hi} T cells are less differentiated than in adult blood (Leeansyah et al., 2014). Nevertheless, Vα7.2^{+}CD161^{hi} T cells in foetal peripheral tissues, particularly the small intestine, show a more mature phenotype than those in the thymus and secondary lymphoid organs, with increased expression of IL-18Rα and CD45RO, and reduced expression of CD62L. In addition, peripheral tissue Vα7.2^{+}CD161^{hi} T cells are functionally more mature than their counterparts in lymphoid organs, producing increased IFNγ *in vitro* following *E. coli* stimulation, although the proportion of Vα7.2^{+}CD161^{hi} T cells that produce IFNγ is significantly reduced in foetal peripheral tissues compared with adult blood (Leeansyah et al., 2014).

The discovery of mature CD45RO$^+$ Vα7.2$^+$CD161^{hi} T cells in foetal peripheral tissues appears at odds with the requirement for commensal bacteria for the development and maturation of MAIT cells in mice (Constantinides et al., 2019; Koay et al., 2016; Le Bourhis et al., 2013; Legoux et al., 2019a; Treiner et al., 2003). Maturation could be understood if the foetal environment was not entirely sterile, conflicting with the "sterile womb hypothesis". In support of this suggestion, a number of recent papers provide evidence for the presence of microbes during foetal development, although these findings remain highly controversial (reviewed in (Perez-Muñoz et al., 2017)). Alternatively, riboflavin metabolites may cross the placenta from the mother to the foetus. This is plausible given that 5-OP-RU painted on the intact skin of mice can reach the thymus and activate MAIT cells (Legoux et al., 2019a).

In summary, we have a very limited understanding of foetal MAIT cell development as the only study that has addressed this so far did not use the MR1 tetramer. Thus, it will be necessary to establish whether MR1/5-OP-RU$^+$ MAIT cells show similar foetal maturation to Vα7.2$^+$CD161^{hi} T cells and if so, to explore whether this is driven by microbial metabolites.

1.5. Activation

1.5.1. Mechanisms

MAIT cells can be activated by TCR signals, cytokine signals independent of the TCR, or by combined TCR and cytokine signals (Fig. 1.3).

The MAIT cell semi-invariant TCR recognises bacterial and yeast riboflavin metabolite ligands in the context of MR1, with the most potent ligands being 5-OP-RU and 5-(2-oxoethylideneamino)-6-D-ribitylaminouracil (5-OE-RU; Fig. 1.3A; Table 1) (Corbett et al., 2014; Kjer-Nielsen et al., 2012). Alternative weaker riboflavin metabolite ligands include a number of ribityl lumazines, including 7-hydroxy-6-methyl-8-D-ribityllumazine (RL-6-Me-7-OH) and its precursor 6,7-dimethyl-8-D-ribityllumazine (RL-6,7-diMe) (Kjer-Nielsen et al., 2012). As the riboflavin synthesis pathway is present in diverse pathogenic and commensal bacteria, as well as in yeast, but absent in mammals, recognition of riboflavin metabolites enables MAIT cells to effectively discriminate self from non-self. No endogenous ligands have been identified for MAIT cells, and evidence suggests the absence of a role for endogenous ligands in MAIT cell activation in response to bacteria and viruses. Namely, the ability of bacteria to activate MAIT cells *in vitro* strongly correlates with the presence of the riboflavin metabolic pathway (Kjer-Nielsen et al., 2012; Tastan et al., 2018). Moreover, viruses do not express riboflavin metabolites and activation of MAIT cells by a number of viruses, including dengue, influenza A, and hepatitis C, is MR1 independent (Loh et al., 2016; Paquin-Proulx et al., 2018; Provine et al., 2019; van Wilgenburg et al., 2016; 2018). Nevertheless, endogenous ligands may have a role *in vivo* in inflammation and cancer, or in specific infectious settings. This is plausible, as MR1 can bind endogenous ligands to activate non-MAIT T cells (Lepore et al., 2017).

In the absence of riboflavin metabolites, MAIT cells can be activated by cytokines independent of their TCR (Fig. 1.3B). They are potently activated by various combinations of IL-12, IL-15, IL-18, type I interferons, TL1A and TNFα (Leng et al., 2019; Provine et al., 2019; Sattler et al., 2015; 2019; Slichter et al., 2016; Ussher et al., 2014; van Wilgenburg et al., 2016). MAIT cells express high levels of the IL-18R

and the IL-12 receptor (IL-12R), and in synergy with other cytokines, IL-12 and IL-18 appear dominant for their TCR-independent activation, at least in viral infections (Cui et al., 2015; Le Bourhis et al., 2010; Loh et al., 2016; Paquin-Proulx et al., 2018; van Wilgenburg et al., 2016; 2018). For example, MAIT cells are activated and accumulate in the lungs of mice infected with influenza A (van Wilgenburg et al., 2018). Pulmonary accumulation is significantly reduced in IL-18$^{-/-}$ mice, but not in IL-12$^{-/-}$, IL-15$^{-/-}$ or IFNαR$^{-/-}$ mice. In contrast, IL-12 is key for the activation of MAIT cells in the lung, with a significant role for IL-18, IL-15 and type I interferons. In general, a single cytokine is insufficient to induce significant activation, but can prime MAIT cells for enhanced responsiveness to subsequent stimulation. For example, MAIT cells stimulated with IL-7 for 48 hours upregulate the expression of cytotoxic molecules (e.g. granzyme B, perforin) and TFs (e.g. PLZF, RORγt) in the absence of cytokine production (Leeansyah et al., 2015). Following overnight stimulation with different doses of *E. coli*, IL-7-primed MAIT cells show significantly increased production of cytokines and enhanced degranulation compared with untreated MAIT cells.

MAIT cells can integrate both TCR- and cytokine-dependent signals to augment functional capacity (Fig. 1.3C) (Bánki et al., 2019; Jo et al., 2014; Lamichhane et al., 2019a; Leeansyah et al., 2014; Sattler et al., 2015; Slichter et al., 2016; Tang et al., 2013; Ussher et al., 2014). Many of the cytokines that can drive TCR-independent activation have been shown to costimulate TCR signalling for MAIT cells, including IL-12, IL-15, IL-18 and/or type I IFNs. These are typically produced by APCs downstream of TLR activation. MAIT cell activation following *E. coli* stimulation of THP-1 cells is mediated by TLR4-induced IL-12 and IL-18, combined with MR1-dependent TCR activation by microbial ligand(s) (Ussher et al., 2014). In this model, early MAIT cell activation depends predominantly on TCR signals, while both

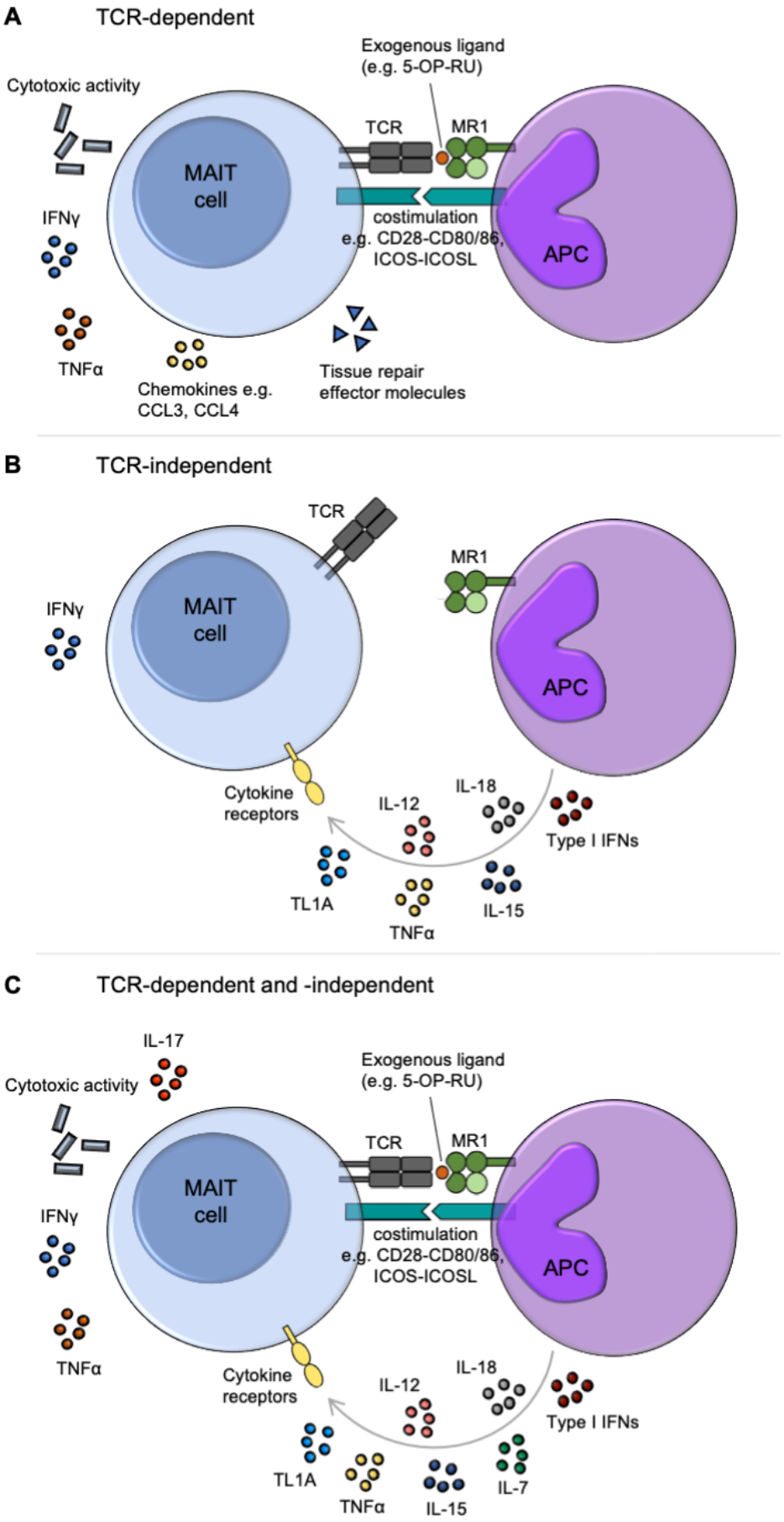

A
TCR-dependent
Exogenous ligand
(e.g. 5-OP-RU)
Cytotoxic activity
TCR
MR1
MAIT
cell
IFNγ
costimulation
e.g. CD28-CD80/86,
ICOS-ICOSL
APC
TNFα
Chemokines e.g.
CCL3, CCL4
Tissue repair
effector molecules
B
TCR-independent
TCR
MR1
IFNγ
MAIT
cell
APC
Cytokine
receptors
IL-12
IL-18
Type I IFNs
TL1A
TNFα
IL-15
C
TCR-dependent and -independent
IL-17
Exogenous ligand
(e.g. 5-OP-RU)
Cytotoxic activity
TCR
MR1
IFNγ
MAIT
cell
costimulation
e.g. CD28-CD80/86,
ICOS-ICOSL
APC
TNFα
Cytokine
receptors
IL-12
IL-18
Type I IFNs
TL1A
TNFα
IL-15
IL-7

Figure 1.3. Mechanisms of MAIT cell activation. MAIT cells are activated via three main pathways: TCR-dependent (A), TCR-independent (predominantly cytokine-driven) (B), and combined TCR-dependent and -independent (C). **(A)** In response to TCR signalling, MAIT cells produce cytokines and chemokines, and exhibit cytotoxic activity. Moreover, recent studies have identified the upregulation of tissue repair pathways and *in vivo* tissue repair functionality following TCR-driven, but not cytokine-driven, MAIT cell activation. **(B)** TCR-independent activation is largely cytokine-mediated, with various combinations of cytokines capable of activating MAIT cells, for example IL-12 + IL-18 and IL-12 + IL-15. The MAIT cell response to cytokine stimulation appears less polyfunctional than their response to TCR stimulation, with dominant production of IFNγ. **(C)** MAIT cells can also be activated through a combination of TCR and cytokine signalling. The cytokines involved in this mode of activation are largely overlapping with those that mediate TCR-independent activation.

TCR and cytokine signals are crucial at later time points. The mechanism(s) underlying TCR and cytokine synergy in MAIT cells remains to be established.

Synergy between TCR and cytokine signalling likely plays a key role in the robust activation of MAIT cell antimicrobial functionality *in vivo*. This is supported by the requirement for both the riboflavin biosynthesis pathway and costimulatory signals for MAIT cell accumulation *in vivo* following bacterial lung infection (Chen et al., 2017). Furthermore, while production of riboflavin metabolites is required for the expansion of MAIT cells in mouse skin following *S. epidermidis* association, expansion is significantly reduced upon IL-18 blockade (Constantinides et al., 2019). In certain settings, cytokines may have a dominant effect on MAIT cell activation, even in the presence of appropriate TCR stimuli (Jo et al., 2014; Kurioka et al., 2018). The relative role of TCR- and cytokine-mediated activation in any given setting will be influenced by many factors, including the nature of the APC. MAIT cell activation in response to *Streptococcus pneumoniae in vitro* is driven purely by cytokines when THP-1 cells are used as the APC, whereas in the presence of monocyte-derived macrophages, activation is driven by both MR1 and cytokines (Kurioka et al., 2018).

In addition to TCR- and cytokine-dependent activation, MAIT cells could potentially be directly activated via NK cell receptors, as has been shown for iNKT cells (Arase et al., 1996; Kuylenstierna et al., 2011). While direct stimulation has yet to be demonstrated for MAIT cells, reports have suggested both costimulatory and coinhibitory roles for the NK cell receptor CD161 (Fergusson et al., 2014; Le Bourhis et al., 2013). CD94 dimerises with different NKG2 isoforms to form activating and inhibitory NK cell receptors (Martinet and Smyth, 2015). MAIT cells with increased CD94 expression show enhanced IFNγ production following IL-12 and IL-18

stimulation (Dias et al., 2017), but the reason for this remains unknown. Despite high expression of NKG2D (Dusseaux et al., 2011), the cytotoxic activity of MAIT cells against *E. coli*-infected HeLa cells (overexpressing MR1) *in vitro* was unaffected by the presence of anti-NKG2D antibody (Le Bourhis et al., 2013). Thus, the role of NK cell receptors in mediating MAIT cell activation (or inhibition) requires further study.

1.5.2. Regulation

MAIT cells express a variety of costimulatory and coinhibitory molecules (Dias et al., 2017), but we have limited understanding of their functional role. A few have been shown to modulate MAIT cell effector function *in vitro* (Jiang et al., 2018; 2014; Shaler et al., 2017a; Turtle et al., 2011). For example, costimulation with αCD28 augments MAIT cell cytokine production and proliferation upon αCD3 stimulation (Turtle et al., 2011). In contrast, the coinhibitory molecule PD-1 is upregulated on MAIT cells in several bacterial and viral infections, including hepatitis C (Hengst et al., 2016) and tuberculosis (Jiang et al., 2014), and PD-1 blockade leads to enhanced IFNγ production by MAIT cells from active tuberculosis patients in response to *in vitro* stimulation with live bacillus Calmette-Guérin (BCG) (Jiang et al., 2014). The role of costimulatory and coinhibitory molecules in modulating MAIT cell activation *in vivo*, and their capacity to differentially skew the MAIT cell effector response, has yet to be investigated.

In addition to the expression of coinhibitory molecules, two additional mechanisms may function to negatively regulate MAIT cell activation and/or to switch off MAIT cell effector functions upon resolution of infection or inflammation. Firstly, MR1-binding antagonist ligands, including 6-formylpterin (6-FP) and acetyl-6-formylpterin

(Ac-6-FP), competitively inhibit MAIT cell activation *in vitro* in response to synthetic agonist ligand (Table 1) (Eckle et al., 2014; Keller et al., 2017; Patel et al., 2013). Furthermore, intranasal administration of Ac-6-FP can inhibit MAIT cell accumulation in a dose-dependent manner in the lungs of mice following intranasal administration of 5-OP-RU and a TLR agonist (riboflavin-deficient *S*. Typhimurium) (Keller et al., 2017). Secondly, MAIT cells exhibit a proapoptotic phenotype driven by PLZF, akin to iNKT cells (Gérart et al., 2013). This sensitivity to activation-induced cell death could function to restrain the MAIT cell effector response in order to minimise immunopathology.

1.5.3. Therapeutic modulation

Although microbe-driven TCR-dependent MAIT cell activation requires expression of the riboflavin biosynthesis pathway (Kjer-Nielsen et al., 2012), non-riboflavin activating and inhibitory MR1 ligands have been identified. Keller and colleagues used a variety of *in silico* approaches to screen libraries of small organic molecules and drugs for potential MR1 ligands (Keller et al., 2017). Identified targets were then tested for their ability to modulate MR1 expression and MAIT cell activation. Metabolites of the drug diclofenac were found to activate certain Jurkat MAIT cell lines *in vitro*, depending on their TCRβ chain usage, whereas 3-formylsalicylic acid could inhibit 5-OP-RU-dependent MAIT cell activation *in vitro* and *in vivo* (Table 1). While these findings imply that common therapeutics might inadvertently affect human MAIT cell activity *in vivo*, they also indicate the potential to design drugs to modulate MAIT cell function.

To summarise, diverse activation mechanisms are available for MAIT cells. Due to their capacity for cytokine-mediated activation, MAIT cells can play key roles in

diverse infectious, as well as inflammatory and malignant diseases, even in the absence of their cognate microbial antigens. The relative importance of TCR- and cytokine-dependent activation *in vivo* is likely to be context-dependent and influenced by the nature of the pathogen and its TCR/TLR ligands, the type of activated APC(s), the availability of costimulatory/coinhibitory molecules, and the stage of infection or inflammation. Nevertheless, the role for cytokines appears more important for the activation of MAIT cells compared with conventional T cells, and may dominate in MAIT cell activation in some settings even in the presence of microbial TCR ligands.

1.6. Effector function

1.6.1. Cytokine production

Upon activation, MAIT cells rapidly produce cytokines such as IFNγ, TNFα, and IL-17 (Table 1) (Dusseaux et al., 2011). While MAIT cells typically produce proinflammatory type 1 and type 17 cytokines, a recent study showed that human MAIT cells can produce type 2 cytokines following prolonged stimulation (Kelly et al., 2019).

The factors that govern which cytokines MAIT cells produce under different stimulatory conditions remain poorly characterised. Human MAIT cells secrete IFNγ following both TCR-dependent and cytokine-dependent activation, whereas TNFα production is more contingent on TCR signalling (Dusseaux et al., 2011; Slichter et al., 2016). Extending this finding, a recent study suggests that the MAIT cell response to TCR stimulation is more polyfunctional than their response to cytokine

stimulation, particularly regarding the production of proinflammatory cytokines and chemokines (Lamichhane et al., 2019b). Though all human MAIT cells express RORγt, in addition to other type 17-associated molecules, such as CCR6 and the IL-23R, IL-17 production *ex vivo* is usually only detected following PMA and ionomycin stimulation, and not upon TCR or cytokine stimulation alone (Billerbeck et al., 2010; Dusseaux et al., 2011; Slichter et al., 2016; Tang et al., 2013; Turtle et al., 2011). Certain cytokines such as IL-7 or IL-23 + IL-1β, can induce IL-17 production when combined with a TCR stimulus (Tang et al., 2013), although the enhanced cytokine production by these MAIT cells is not specific to IL-17. In IL-1 receptor-deficient mice, cutaneous MAIT cells show significantly impaired IL-17A production in response to *S. epidermidis* topical association (Constantinides et al., 2019), suggesting that IL-1 licenses MAIT cells to produce IL-17. Whether IL-1 has a similarly important role in other mouse tissues, is currently unknown. Following stimulation of human MAIT cells *in vitro* with αCD3/αCD28, phytohaemagglutinin, IL-2 and IL-7, a fraction of MAIT cells acquire the capacity to produce type 2 cytokines, particularly IL-5 and IL-13 (Kelly et al., 2019). However, type 2 cytokine production is only detected following several days of stimulation. Further work is required to identify the specific signals that induce type 2 cytokine production by MAIT cells, as well as the potential role of these cells *in vivo*. Whether these type 2 cytokine-secreting MAIT cells comprise a distinct MAIT2 subset is unknown, although since the majority of the IL-13-producing MAIT cells cosecrete TNFα, this seems unlikely.

Human MAIT cells in different tissues exhibit differential cytokine production. MAIT cells in the female genital mucosa appear skewed towards type 17 functions, secreting increased IL-17 and IL-22, and decreased IFNγ and TNFα, compared with blood MAIT cells (Gibbs et al., 2016). Similarly, MAIT cells from the oral mucosa show enriched IL-17 production, and reduced IFNγ and TNFα production, compared

with matched blood MAIT cells (Sobkowiak et al., 2018), while IL-22-secreting cells are enriched in foetal small intestine (Leeansyah et al., 2014). Adipose tissue MAIT cells exhibit the unique capacity to secrete IL-10 (Carolan et al., 2015). Whether variation in MAIT cell cytokine production across tissues is explained by the tissue-specific enrichment of different MAIT cell subsets, or altered functional differentiation in tissues, is currently unknown. NKT10 cells, a specific IL-10-producing subset of iNKT cells preferentially localise to adipose tissue (Lynch et al., 2015; Sag et al., 2014), suggesting that the adipose tissue-enriched IL-10-producing MAIT cells in humans could comprise a distinct subset (Carolan et al., 2015).

1.6.2. Cytotoxic activity

In addition to cytokine secretion, MAIT cells display direct cytotoxic activity. MAIT cell killing is mediated via the perforin/granzyme pathway (Kurioka et al., 2015; Le Bourhis et al., 2013; Leeansyah et al., 2015). While MAIT cells can upregulate FasL expression upon TCR and cytokine stimulation (Lamichhane et al., 2019b), killing appears independent of Fas/FasL (Kurioka et al., 2015), as well as NKG2D (Le Bourhis et al., 2013). Whilst their cytotoxic capacity (i.e. perforin and granzyme B expression) is enhanced upon activation via the TCR and/or with cytokines, target cell killing is MR1 dependent (Kurioka et al., 2015; Le Bourhis et al., 2013; Leeansyah et al., 2015). Compared with DN MAIT cells, $CD8^+$ MAIT cells show increased expression of granzyme B, granulysin and perforin, although the differences are minor (Dias et al., 2018). Whether this differential expression translates into altered cytotoxic capacity, is currently unknown.

1.6.3. Tissue repair

Several recent papers have uncovered a role for MAIT cells in tissue repair. In mice, a population of commensal-specific H2-M3-restricted $CD8^+$ Tc17 cells that accumulate in the skin in response to *S. epidermidis* association, exhibit a tissue repair gene signature and can accelerate skin wound healing *in vivo* (Linehan et al., 2018). This same tissue repair gene signature is enriched in human MAIT cells following TCR-, but not cytokine-dependent activation *in vitro*, as well as in MAIT cells from the lungs of mice acutely infected with *L. longbeachae* (Hinks et al., 2019; Lamichhane et al., 2019b; Leng et al., 2019). In an *in vitro* wound-healing assay, supernatants from *E. coli*-stimulated human MAIT cells enhance wound closure (Leng et al., 2019). More strikingly, skin wound healing *in vivo* is significantly reduced in $Mr1^{-/-}Tcrd^{-/-}$ mice compared with $Tcrd^{-/-}$ mice, while direct topical application of the MAIT cell activating ligand 5-OP-RU enhances wound closure (Constantinides et al., 2019). Thus, along with their proinflammatory functions, MAIT cells may have important roles in tissue homeostasis or in tissue repair following infection or injury. The mechanisms by which MAIT cells mediate their tissue repair function, is currently unknown.

1.6.4. Immune interactions

Whilst knowledge of MAIT cell crosstalk with other immune cell subsets remains relatively limited, studies have identified important interactions with a number of immune cell types, including dendritic cells (DC), monocytes, B cells and NK cells. MAIT cells have also been shown to interact with fibroblasts, which could contribute to their newly described tissue repair functionality.

Through contact with myeloid cells, MAIT cells appear to have important immune regulatory functions. For example, upon antigen-specific activation, MAIT cells upregulate CD40 ligand and induce CD40-dependent DC maturation, and in synergy with TLR ligands, promote the secretion of IL-12 (Salio et al., 2017). DC-derived IL-12 can subsequently enhance MAIT cell activation. In addition, MAIT cells can influence monocyte differentiation. $MR1^{-/-}$ mice show enhanced susceptibility to pulmonary infection with *Francisella tularensis* (*F. tularensis*) live vaccine strain and delayed bacterial clearance (Meierovics et al., 2013). In this setting, MAIT cells promote early GM-CSF production in the lungs, resulting in the differentiation of inflammatory monocytes into DCs and the recruitment of activated $CD4^{+}$ T cells into the lungs (Meierovics et al., 2013; Meierovics and Cowley, 2016).

Factors secreted by TCR-activated human MAIT cells can influence the differentiation of monocytes to macrophages. Namely, macrophages differentiated in the presence of M-CSF + TCR-activated MAIT cells show an increased ratio of anti-inflammatory M2 to proinflammatory M1 macrophages compared with monocytes differentiated in the presence of M-CSF + nonactivated MAIT cells (Li et al., 2018). Seemingly contradictory with this finding, MAIT cells cultured with monocyte-derived macrophages induce the secretion of the proinflammatory cytokines IL-6 and IL-8, with secretion enhanced in the presence of activated MAIT cells (Hegde et al., 2018).

As well as their effects on myeloid populations, MAIT cells can influence the function of lymphocytes. For example, MAIT cells provide non-cognate B cell help *in vitro* (Bennett et al., 2017). In response to TCR-dependent, or TCR- and cytokine-dependent activation, MAIT cells secrete factors that act on B cells to promote the differentiation of memory cells into plasmablasts and to increase antibody

production. In these experiments, TCR stimulation was essential for the capacity of MAIT cells to provide B cell help. Additionally, in the context of whole blood, activated MAIT cells promote NK cell transactivation in an MR1- and IL-18-dependent manner (Salio et al., 2017).

Finally, MAIT cells can alter the phenotype and function of fibroblasts. Human hepatic myofibroblasts display significantly increased proliferation when cultured with activated MAIT cells compared with non-activated MAIT cells, and this effect is MR1 dependent (Hegde et al., 2018). Moreover, similar to their effect on monocyte-derived macrophages, activated MAIT cells promote IL-6 and IL-8 production by hepatic myofibroblasts, in part through secretion of TNFα. In an *in vivo* model of liver fibrogenesis, $MR1^{-/-}$ mice were resistant to fibrosis, while iVα19 transgenic mice showed enhanced fibrosis (Hegde et al., 2018), suggesting that the functions identified *in vitro* could also be relevant *in vivo*.

In summary, MAIT cells rapidly produce cytokines, predominantly proinflammatory, and exhibit cytotoxic activity. The relative importance of cytokine secretion versus cytotoxic activity for MAIT cell function is currently unknown, although protection against *L. longbeachae* and influenza infection *in vivo* is at least partly dependent on their production of IFNγ (van Wilgenburg et al., 2018; Wang et al., 2018). Additionally, MAIT cells can promote tissue repair, and influence the function of both immune and non-immune cell populations. Most interactions between MAIT cells and other cell subsets have so far only been demonstrated *in vitro*, thus their relevance *in vivo* and exactly where such interactions would take place, remains to be determined. However, given their abundance and wide tissue distribution, MAIT cells have the potential to profoundly influence the immune response through their crosstalk with other cell populations.

1.7. Subsets

Although human MAIT cells are thought to be largely homogeneous in phenotype and function, recent findings suggest that they may be more diverse than currently appreciated. Lepore and colleagues identified subpopulations of human MAIT cells with distinct cytokine secretion profiles (Lepore et al., 2014), although whether these represent separate lineages is currently unknown. Human MAIT cells also show heterogeneous expression of certain NK cell-associated receptors, including CD56 and CD84, the expression of which correlates with their functional response to cytokine stimulation (Dias et al., 2017). Finally, MAIT cells in certain tissues exhibit altered cytokine production. For example, MAIT cells are skewed towards a Th17-like phenotype in female genital tract, and express lower levels of the TFs PLZF and Eomes compared with peripheral blood MAIT cells (Gibbs et al., 2016), suggesting that they may comprise a distinct MAIT cell subset.

MAIT cells in humans can be $CD8^+$, $CD4^+$, or DN (Table 1) (Martin et al., 2009). While it remains unclear whether these populations reflect truly distinct subsets, some differences have been identified. Initial studies using the MR1/5-OP-RU tetramer indicate minor distinctions between $CD4^+$ and $CD8^+$/DN MAIT cells, including differential expression of certain TFs (PLZF, Eomes), chemokine receptors (CCR4, CXCR6), adhesion molecules (CD56) and NK cell receptors (NKG2A, NKG2D) (Gherardin et al., 2018b; Kurioka et al., 2017). The comparative phenotype and function of $CD8^+$ and DN MAIT cells has been characterised in detail (Dias et al., 2018). While differences are generally minor, DN MAIT cells show a strikingly enhanced propensity for activation-induced apoptosis relative to $CD8^+$ MAIT cells. Moreover, DN MAIT cells show reduced TCR diversity compared with

$CD8^+$ MAIT cells and can be derived from TCR-stimulated $CD8^+$ MAIT cells *in vitro*, implying that DN MAIT cells might represent a late stage of MAIT cell maturation. However, a recent scRNA-seq study indicates that $CD8^+$, $CD4^+$ and DN MAIT cells in mice are not transcriptionally distinct (Constantinides et al., 2019). Thus, further work is required to determine whether human $CD8^+$, $CD4^+$ and DN MAIT cells are fundamentally distinct subsets.

In contrast with humans, MAIT cells in mice comprise at least two subsets, including a major $ROR\gamma t^+IL\text{-}17^+$ MAIT17 population, and a smaller $T\text{-}bet^+IFN\gamma^+$ MAIT1 subset (Fig. 1.4) (Koay et al., 2019a; 2016; Legoux et al., 2019b; Rahimpour et al., 2015). The majority of human MAIT cells express IFNγ and a smaller fraction produce IL-17, whilst some secrete both cytokines, highlighting important differences between mouse and human MAIT cells (Dusseaux et al., 2011). scRNA-seq data indicates that mouse MAIT1 and MAIT17 subsets are transcriptionally distinct and derive from a shared precursor population (Koay et al., 2019a; Legoux et al., 2019b), the exact nature of which requires further study. Mouse MAIT cell subsets can show plasticity *in vivo*. For example, following intranasal infection with *S*. Typhimurium, $ROR\gamma t^+$ MAIT cells in the lung upregulate T-bet expression to become $ROR\gamma t^+T\text{-}bet^+$ cells that can secrete either IL-17 or IFNγ, though the majority produce IL-17 (Chen et al., 2017).

Overall, human MAIT cells appear relatively homogenous and exhibit a mixed type 1/type 17 phenotype, while mouse MAIT cells comprise two transcriptionally and functionally distinct subsets. The diversity of MAIT cell subsets identified so far is considerably reduced compared with the related iNKT cell population (Fig. 1.4). Whether minor subsets of MAIT cells are present in specific human or mouse tissues, or under certain conditions, remains to be established.

1.8. Role in disease

MAIT cells have been shown to exhibit altered frequency, phenotype and/or function in a huge range of diseases (Toubal et al., 2019). However, a mechanistic role for MAIT cells has been demonstrated in relatively few diseases, largely using $MR1^{-/-}$ mice. While several studies have demonstrated a protective role in bacterial infection, including in *F. tularensis* (Meierovics et al., 2013; Meierovics and Cowley, 2016), *Klebsiella pneumoniae* (Georgel et al., 2011) and *L. longbeachae* (Wang et al., 2018), MAIT cells are pathogenic in *Helicobacter pylori*-associated gastritis (D'Souza et al., 2018). In experimental influenza A infection, MAIT cells are activated, and weight loss and mortality are increased in $MR1^{-/-}$ mice, suggesting a protective role for MAIT cells (van Wilgenburg et al., 2018). Their role in other viral diseases has yet to be mechanistically determined. A protective role for MAIT cells has also been identified in type 1 diabetes (Rouxel et al., 2017) and graft-versus-host disease (Varelias et al., 2018), with pathogenic roles described in arthritis (Chiba et al., 2012), liver fibrosis (Hegde et al., 2018) and cancer (Yan et al., 2019). Based on the limited available data, MAIT cells are predominantly protective in infectious disease, while their role is more variable in other disease settings.

1.9. Avenues for future MAIT cell research

With the relatively recent generation of MR1 tetramers (Corbett et al., 2014; Reantragoon et al., 2013), it is now possible to detect MAIT cells in WT mice (Rahimpour et al., 2015), as well as in human settings where MAIT cell frequency is low, for example in the thymus (Koay et al., 2016). Consequently, MAIT cells are now being studied in an increasingly wide variety of settings, including in numerous

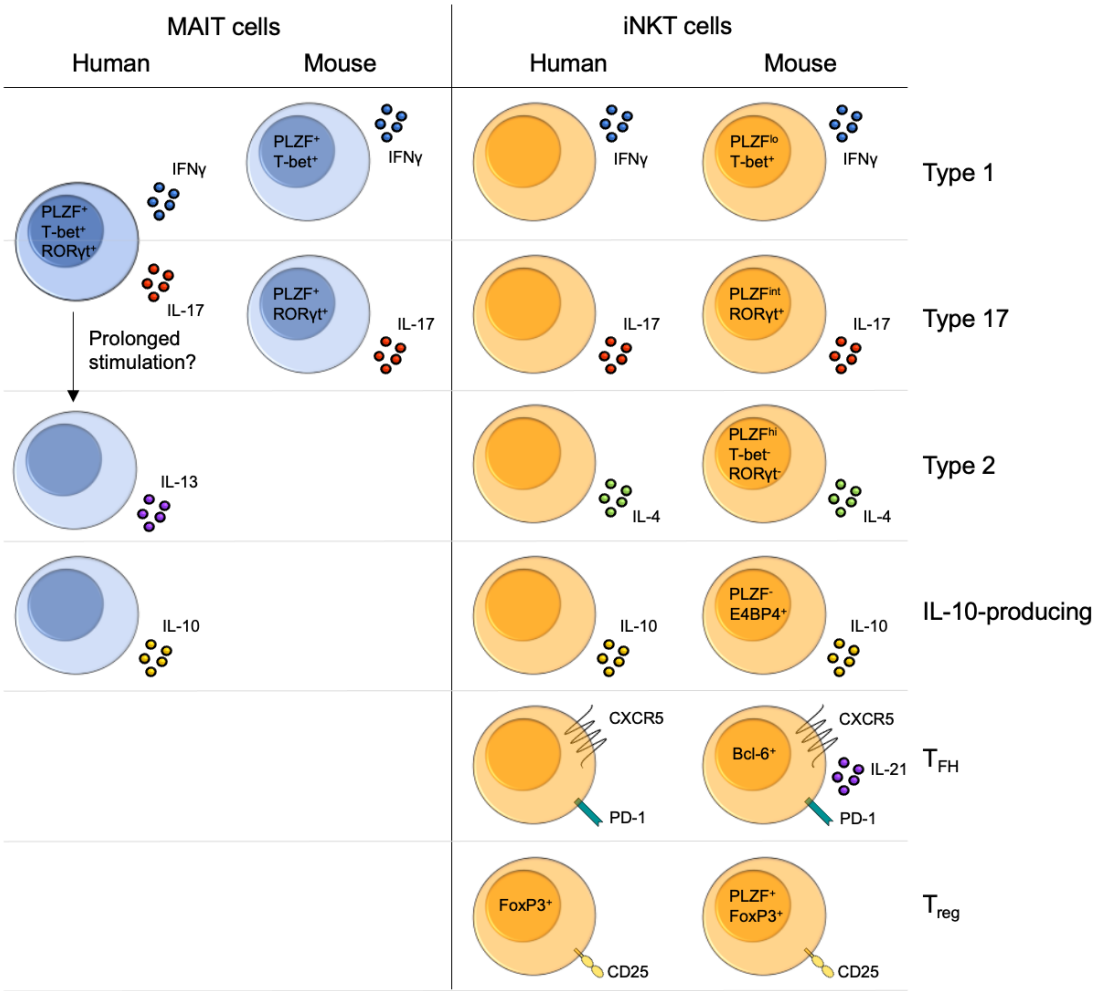


Figure 1.4. Functional capacity of MAIT cells compared with iNKT cells. Where subsets of MAIT cells or iNKT cells have been defined, characteristic cytokines, TFs, and/or surface markers, are illustrated. MAIT cells and iNKT cells exhibit overlapping functions, although a wider range of functions have been described for iNKT cells. In mice, distinct type 1 and type 17 MAIT and iNKT cell subsets have been identified. In contrast, human MAIT cells exhibit a mixed type 1/type 17 phenotype. While iNKT cells comprise a transcriptionally distinct type 2 subset, at least in mice, human MAIT cells only produce type 2 cytokines under conditions of prolonged stimulation. Given the coexpression of TNFα by the majority of IL-13-producing MAIT cells, it is likely that these cells reflect a particular state of activation, rather than a distinct subset. In mice, IL-10-producing iNKT cells comprise a distinct subset with altered TF expression. Human MAIT cells and iNKT cells can produce IL-10, and human IL-10-secreting MAIT cells are enriched in adipose tissue, similar to mouse NKT10 cells. However, whether these IL-10-producing populations comprise distinct subsets in humans is currently unknown. Finally, multiple specialised subsets of iNKT cells have been identified in mice, including NKT_{FH} cells and $iNKT_{reg}$ cells. Human iNKT cells with similar phenotypes and/or functions have also been identified (NKT_{reg} only *in vitro*), but analogous populations have yet to be described for MAIT cells.

human diseases and animal disease models. Undoubtedly, this will lead to a greatly improved understanding of the role of MAIT cells in both health and disease.
It will be important to determine the signals governing peripheral MAIT cell maturation and the maintenance of homeostasis. In addition to IL-18 and IL-23 (Constantinides et al., 2019; Koay et al., 2016; Wilson et al., 2015), it is worth investigating the role of IL-7 and IL-15, given MAIT cell responsiveness to these cytokines (Leeansyah et al., 2015; Sattler et al., 2015; Slichter et al., 2016; Tang et al., 2013; van Wilgenburg et al., 2016) and their function in homeostasis of the related iNKT cell population (Baev et al., 2004; Gordy et al., 2011; Matsuda et al., 2002; Ranson et al., 2003; Webster et al., 2014). MR1 does not appear to be required for the homeostasis of MAIT cells in the skin (Constantinides et al., 2019), but its role in other tissues has yet to be examined.

While MAIT cells can be activated through the TCR or with cytokines, or by a combination of both (Kurioka et al., 2016), which of these activation mechanisms predominate *in vivo* remains to be determined. It is plausible that cytokines may be key for MAIT cell activation, given the indiscriminate expression of their riboflavin metabolite ligands in both commensal and pathogenic bacteria (Kjer-Nielsen et al., 2012; Le Bourhis et al., 2010), and the need for both TCR and cytokine stimulation to induce robust secretion of proinflammatory cytokines (Slichter et al., 2016; Tang et al., 2013; Turtle et al., 2011). Despite expression of a variety of NK cell receptors, costimulatory molecules, and immune checkpoint molecules (Dias et al., 2017; Dusseaux et al., 2011), how stimulation of these receptors modulates MAIT cell activation, is essentially unknown.

Transcriptionally and functionally distinct MAIT cell subsets have been identified in mice (Koay et al., 2019a; 2016; Legoux et al., 2019b; Rahimpour et al., 2015). In

contrast, human MAIT cells exhibit a relatively uniform phenotype, and coexpress key TFs, suggesting a single MAIT cell subset (Dusseaux et al., 2011; Leeansyah et al., 2015). Nevertheless, recent studies indicate previously unappreciated phenotypic and functional heterogeneity in blood MAIT cells (Dias et al., 2017; Dusseaux et al., 2011; Lepore et al., 2014), as well as skewed or unique functions in certain tissues (Carolan et al., 2015; Gibbs et al., 2016; Leeansyah et al., 2014; Sobkowiak et al., 2018). Thus, the existence of specialised MAIT cell subsets, remains an open question. Additionally, it is unclear to what extent MAIT cells can show functional plasticity driven by environmental factors.

Recently, a number of studies have expanded our knowledge of MAIT cell interactions with other immune and non-immune cell populations (Bennett et al., 2017; Hegde et al., 2018; Li et al., 2018; Meierovics and Cowley, 2016; Salio et al., 2017). However, the role of MAIT cell interactions *in vivo* and their influence on the overall innate and adaptive immune response, is largely unknown.

Finally, we have limited understanding of the mechanistic role of MAIT cells in disease, despite their altered frequency and function in numerous infectious, inflammatory and malignant diseases (Toubal et al., 2019). Given that MR1 tetramers are now widely available, it is possible to address the role of MAIT cells *in vivo* in models of disease. MR1 ligands have been discovered amongst drugs and drug-like molecules (Keller et al., 2017). This indicates the potential for the future development of therapeutics to modulate MAIT cell function.

1.10.Conclusion

In conclusion, MAIT cells are an abundant population of innate-like T cells with diverse functionality. The MAIT cell field is growing rapidly, due to the recent development of the MR1 tetramer and hence the possibility to examine MAIT cells in WT mice. Given their abundance in humans, their capacity for rapid cytokine production in response to TCR and/or cytokine stimulation, and their interactions with other immune cell populations, MAIT cells are likely key players in the immune system, both in health and disease. In support of this, they show altered phenotype and function in numerous human diseases, and exhibit protective or deleterious roles in mouse models of infection or inflammation. Consequently, MAIT cells represent an attractive target for therapeutic manipulation, especially considering their high frequency in humans and their recognition of a monomorphic MHC-like molecule. To realise this goal, further research is necessary to develop a greater understanding of MAIT cell development, function, and regulation, and their specific roles in disease.

1.11.Aims and key questions of the thesis

Aim 1:

To characterise the transcriptome and chromatin accessibility landscape of human blood MAIT cells.

- What are the unique features of the MAIT cell transcriptome relative to T_{mem} cells?
- Are differences in gene expression between MAIT cells and T_{mem} cells associated with altered chromatin accessibility?

- What are the key TFs that regulate MAIT cell phenotype and function?

Aim 2:

To investigate heterogeneity in the transcriptional profile and TCR repertoire of human blood and liver MAIT cells at a single-cell level.

- Do MAIT cells comprise multiple subsets within the blood or liver?
- How does the transcriptional profile of MAIT cells vary between the blood and the liver?
- How does the TCR repertoire of MAIT cells vary between blood and liver, and between individuals?
- Is the transcriptional profile of MAIT cells associated with their TCR usage?

Aim 3:

To characterise human MAIT cell functional heterogeneity and its association with TCR usage at a single-cell level.

- Do MAIT cells comprise functionally distinct subsets?
- What are the similarities and differences in the MAIT cell response to TCR and cytokine stimulation?
- Does TCR usage influence the capacity of MAIT cells to respond to TCR or cytokine stimuli, or their functional response to stimulation?

Table 1. Characteristics of human and mouse MAIT cells.

	TCR usage	MHC-like molecule	Ligands	CD4/CD8	Naïve/ memory	Chemokine receptors	Cytokine receptors	Cytokines/ chemokines produced	Transcription factors
Human MAIT	Vα7.2-Jα33/12/20 (*TRAV1-2/TRAJ33/12/20*) TCRα. Biased repertoire of Vβ chains - Vβ2, Vβ13 (*TRBV20, TRBV6*). Oligoclonal CDR3β.	MR1	**Activating** - riboflavin metabolites (5-OP-RU, 5-OE-RU), drug metabolites (e.g. diclofenac metabolites). **Inhibitory** - folic acid metabolites (6-FP, Ac-6-FP), small organic molecules (e.g. salicylates).	70-90% $CD8^+$ (~50% CD8αα, ~50% CD8αβ), 10-20% DN^a, <5% $CD4^+$	$CD45RO^+$ $CCR7^-CD62L^{lo}$ effector memory	CCR2, CCR5, CCR6, CXCR6	IL-12R, IL-18R, IL-7R, IL-15R, IL-23R	IFNγ, TNFα, IL-17A, IL-26, CCL3, CCL4, IL-13 (after prolonged stimulation), IL-10 (in adipose tissue)	$PLZF^+$, $RORγt^+$, $T\text{-}bet^{int}$, $Eomes^{hi}$, $Helios^{int}$, $C/EBPδ^+$
Mouse MAIT	Vα19-Jα33 (*TRAV1/TRAJ33*) TCRα. Biased repertoire of Vβ chains - Vβ6, Vβ8 (*TRBV19, TRBV13*).	MR1	**Activating** - riboflavin metabolites (5-OP-RU, 5-OE-RU). **Inhibitory** - folic acid metabolites (6-FP, Ac-6-FP), small organic molecules (e.g. salicylates).	B6 mice - 55-90% DN, <25% $CD8^+$ (~50% CD8αα, ~50% CD8αβ), <20% $CD4^+$. Variation across tissues.	$CD44^{hi}CD62L^{lo}$ effector memory	CXCR6, CCR6	IL-12R, IL-18R, IL-7R	IFNγ, TNFα, IL-17A, CCL3, CCL4	$PLZF^+$, $RORγt^+$ (subset), $T\text{-}bet^+$ (subset)
[REFS]	(Lepore et al., 2014; Reantragoon et al., 2013; Tilloy et al., 1999; Treiner et al., 2003)	(Gold et al., 2010; Le Bourhis et al., 2010; Treiner et al., 2003)	(Corbett et al., 2014; Eckle et al., 2014; Keller et al., 2017; Kjer-Nielsen et al., 2012)	(Gherardin et al., 2018; Rahimpour et al., 2015; Reantragoon et al., 2013; Walker et al., 2012)	(Dusseaux et al., 2011; Martin et al., 2009; Rahimpour et al., 2015)	(Billerbeck et al., 2010; Cui et al., 2015; Dusseaux et al., 2011; Le Bourhis et al., 2010; Rahimpour et al., 2015)	(Billerbeck et al., 2010; Cui et al., 2015; Dusseaux et al., 2011; Le Bourhis et al., 2010; Rahimpour et al., 2015; Wilson et al., 2015)	(Carolan et al., 2015; Cui et al., 2015; Dusseaux et al., 2011; Hinks et al., 2019; Kelly et al., 2019; Leng et al., 2019; Rahimpour et al., 2015; Slichter et al., 2016)	(Dusseaux et al., 2011; Lee et al., 2018; Leeansyah et al., 2015; Rahimpour et al., 2015; Savage et al., 2008)

[a]DN = double negative/$CD4^-CD8^-$

2. Materials and methods

2.1. Data generation

2.1.1. PBMC isolation and cell enrichment

PBMCs were isolated from leukocyte cones (Oxford Blood Donor Centre) or whole blood by density gradient centrifugation (Lymphoprep, Axis-Shield) at 2000 rpm for 30 min with no brake. Cryopreserved (90% foetal calf serum [FCS], 10% dimethyl sulfoxide [DMSO]) PBMCs were used for most experiments, except for the staining of blood matched with colon biopsies, which was performed on freshly isolated cells. $CD8^+$ T cells were enriched from PBMCs using CD8 MicroBeads (Miltenyi Biotec) following manufacturer's instructions.

2.1.2. Media

Cells were thawed and washed in R10 (RPMI 1640 [Sigma-Aldrich], 10% FCS [Sigma-Aldrich], 1% Penicillin-Streptomycin [Thermo Fisher Scientific]). For scRNA-seq and single-cell TCR-sequencing (scTCR-seq) on activated MAIT cells, cells were incubated in complete medium (RPMI, 10% FCS, 1X non-essential amino acids [Thermo Fisher Scientific], 1 mM Na Pyruvate [Thermo Fisher Scientific], 10 mM HEPES [Thermo Fisher Scientific], 1% Penicillin-Streptomycin, and 50 µM 2-mercaptoethanol [Thermo Fisher Scientific]).

2.1.3. Human tissue collection and processing

Colon biopsies

Colon biopsies and matched blood were obtained from patients undergoing routine endoscopy at the John Radcliffe Hospital, Oxford. Biopsies were suspended in a solution of 1 mg/ml collagenase D and 0.1 mg/ml DNase in a gentleMACS C Tube (Miltenyi Biotec), homogenised on a gentleMACS Dissociator (Miltenyi Biotec), and incubated for 1 h at 37°C in a shaking incubator (200 rpm). Following incubation, samples were further homogenised on a gentleMACS Dissociator, and cells passed through a 70 µm cell strainer. Cells were washed three times in R10 before resuspension in FACS buffer (PBS, 0.05% BSA, 1 mM EDTA) for staining.

Liver biopsies and resections

The liver biopsy for optimising tetramer staining was collected during gastric bypass surgery at the Churchill Hospital, Oxford. Liver tissue and matched blood for combined scRNA-seq and scTCR-seq were obtained from patients (50-66 years old) undergoing liver resection for metastatic colorectal cancer (CRC; n = 1), haemangioma (n = 1), or hepatocellular carcinoma (HCC; n = 2) at the Churchill Hospital, Oxford. Patients had no chronic liver disease, active excess alcohol consumption (> 14g/day), infection, immunosuppression or family history of liver disease.

Disease-free liver tissue was collected from the resection margin. Liver tissue was cut into small pieces with a scalpel and ground through a 70 µm cell strainer. Cells were washed once with R10 and lymphocytes isolated by density gradient centrifugation on a discontinuous 35%/70% Percoll (GE Healthcare) gradient at

2000 rpm for 20 min with no brake. Isolated lymphocytes were washed once with R10 and red blood cells lysed with ACK for 3-5 min if required. Lymphocytes were washed with R10 and either resuspended in FACS buffer for staining, or resuspended in freezing medium (90% FCS, 10% DMSO) and cryopreserved at -80°C.

2.1.4. Ethics

All samples were obtained with written informed consent through the Oxford Gastrointestinal Illnesses Biobank (REC Ref: 16/YH/0247).

2.1.5. MR1 tetramer staining

Biotinylated human MR1/5-OP-RU and MR1/6-FP monomers were provided by the NIH Tetramer Core Facility (Emory University). Tetramers were generated using Streptavidin-Phycoerythrin or Streptavidin-Brilliant Violet 421 (BioLegend) following the standard NIH Tetramer Core Facility protocol. Except for tetramer optimisation experiments (see Chapter 3), single or double (PE and BV421) tetramer staining was performed at room temperature (RT) for 40 min prior to further surface/intracellular staining.

2.1.6. Surface and intracellular antibody staining for flow cytometry

Cells were washed once in FACS buffer (PBS, 0.05% BSA, 1 mM EDTA), and surface staining (including viability staining) performed in FACS buffer for 30 min at 4°C. For intracellular staining, cells were fixed and permeabilised for 20 min at 4°C

with Cytofix/Cytoperm (BD Biosciences), washed twice in Perm/Wash buffer (BD Biosciences), and stained in Perm/Wash buffer for 30 min at 4°C. Following intracellular staining, cells were washed twice in Perm/Wash buffer, resuspended in FACS buffer, and stored at 4°C until analysed on the flow cytometer. Samples were acquired on a BD LSRFortessa or a BD LSR II.

2.1.7. Staining for fluorescence-activated cell sorting (FACS)

When sorting for RNA-sequencing (RNA-seq) and the assay for transposase-accessible chromatin using sequencing (ATAC-seq), MR1 tetramer staining was performed as described above, together with staining for CCR7. Cells were washed twice in FACS buffer, and surface staining performed in Brilliant Stain Buffer (BD Biosciences) for 30 min at 4°C. Cells were washed once in FACS buffer and once in PBS with 0.5% BSA, before resuspension in pre-sort buffer (PBS, 1% BSA, 10 mM HEPES).

When sorting for scRNA-seq and scTCR-seq, MR1 tetramer staining was performed as described above (MR1 tetramer only used in Experiment 1). Cells were washed twice in FACS buffer, and surface staining performed in Brilliant Stain Buffer for 30 min at 4°C. Cells were washed twice in PBS with 0.5% BSA, resuspended in pre-sort buffer containing 5 nM SYTOX Green Nucleic Acid Stain (Thermo Fisher Scientific), and incubated for 20 min at 4°C prior to sorting.

2.1.8. FACS

Cells were sorted on a BD FACSAria III with a 70 µm nozzle. Sorted cells were collected in PBS, 2% BSA for RNA-seq and ATAC-seq (Chapter 4), or RPMI, 10%

FBS, 25 mM HEPES for scRNA-seq and scTCR-seq (Chapter 5 and 6). Sort purity was > 99%.

2.1.9. Staining panels

RNA-seq and ATAC-seq (Chapter 4)

Marker	Clone	Source/company
MR1/5-OP-RU (PE and BV421)		NIH Tetramer Core Facility
MR1/6-FP (PE and BV421) – in control panel		NIH Tetramer Core Facility
CCR7	150503	BD Biosciences
CD161	191B8	Miltenyi Biotec
CD3ε	UCHT1	BioLegend
CD4	OKT4	BioLegend
CD8α	SK1	BioLegend
CD14	M5E2	BioLegend
CD19	HIB19	BioLegend
CD45RO	UCHL1	BD Biosciences
TCR γδ	B1	BD Biosciences
LIVE/DEAD Fixable Near-IR Stain		Thermo Fisher Scientific

scRNA-seq and scTCR-seq Experiment 1 (Chapter 5)

Marker	Clone	Source/company
MR1/5-OP-RU (PE and BV421)		NIH Tetramer Core Facility
MR1/6-FP (PE and BV421) – in control panel		NIH Tetramer Core Facility
CCR7	150503	BD Biosciences
CD161	191B8	Miltenyi Biotec
CD3ε	UCHT1	BioLegend
CD4	OKT4	BioLegend
CD8α	SK1	BioLegend

CD14	M5E2	BioLegend
CD19	HIB19	BioLegend
CD45	HI30	BioLegend
CD45RO	UCHL1	BD Biosciences
CD69	FN50	BioLegend
TCR γδ	B1	BD Biosciences
TCR Vα24-Jα18	6B11	BioLegend
TCR Vδ2	B6	BioLegend
Vα7.2 (for donors GI3008 and GI2916)	3C10	BioLegend
SYTOX Green Nucleic Acid Stain		Thermo Fisher Scientific

scRNA-seq and scTCR-seq Experiment 2 (Chapter 6)

Marker	**Clone**	**Source/company**
CD161	191B8	Miltenyi Biotec
CD3ε	HIT3a	BioLegend
CD4	OKT4	BioLegend
CD8α	SK1	BioLegend
CD14	M5E2	BioLegend
CD19	HIB19	BioLegend
CD26	BA5b	BioLegend
CD69	FN50	BioLegend
TCR γδ	B1	BD Biosciences
TCR Vα24-Jα18	6B11	BioLegend
TCR Vδ2	B6	BioLegend
Vα7.2	3C10	BioLegend
SYTOX Green Nucleic Acid Stain		Thermo Fisher Scientific

Additional antibodies used in method development and assay optimisation

Marker	Clone	Source/company
MR1/5-OP-RU (PE and BV421)		NIH Tetramer Core Facility
MR1/6-FP (PE and BV421) – in control panel		NIH Tetramer Core Facility
CD161	191B8	Miltenyi Biotec
CD3ε	UCHT1	BioLegend
CD4	OKT4	BioLegend
CD8α	SK1	BioLegend
CD45	HI30	BioLegend
Vα7.2	3C10	BioLegend
LIVE/DEAD Fixable Near-IR Stain		Thermo Fisher Scientific

2.1.10. *In vitro* stimulation of enriched $CD8^+$ T cells for scRNA-seq and scTCR-seq

A Pierce streptavidin coated high capacity flat-bottom 96-well plate (Thermo Fisher Scientific) was coated with 50 µl biotinylated MR1/5-OP-RU monomer (obtained from the NIH Tetramer Core Facility) at 10 µg/ml in PBS overnight at 4°C. Cryopreserved PBMCs from three healthy donors were thawed and $CD8^+$ T cells enriched. Enriched $CD8^+$ T cells were diluted to 10 x 10^6 cells/ml and 100 µl added to the appropriate 96-well plate (MR1/5-OP-RU-coated plate for TCR-stimulated condition, round-bottom plate for unstimulated and cytokine-stimulated conditions). IL-12 (50 µg/ml; R&D Systems) and IL-18 (50 µg/ml; R&D Systems) were added to the cytokine-stimulated condition, αCD28 (1 µg/ml; clone CD28.2; BioLegend) to the TCR-stimulated condition, and complete medium to the unstimulated condition (final volume 200 µl in all wells). Cells were incubated for 20 h at 37°C, 5% CO_2.

2.1.11. RNA-seq library generation

RNA was extracted as previously described (Barnitz et al., 2013; Quigley et al., 2010), however standard 1.5 ml eppendorf tubes were used in place of phase-lock heavy gel tubes. RNA was quantified using the Qubit 3.0 Fluorometer (Thermo Fisher Scientific) with the Qubit RNA HS Assay Kit, and RNA quality was checked on a 2100 Bioanalyzer instrument (Agilent Technologies) using the Agilent RNA 6000 Nano Kit. RIN scores were ≥ 8.5. Poly(A)-enriched RNA-seq libraries were generated using the NEBNext Ultra II RNA Library Prep with Sample Purification Beads (New England Biolabs) kit following manufacturer's instructions. Final library size distribution was analysed on a 2100 Bioanalyzer instrument using the Agilent DNA 1000 Kit. RNA-seq libraries were quantified using the KAPA Library Quantification Kit (Roche). Libraries were diluted to 2 nM in nuclease-free H_2O and pooled for sequencing. Paired-end sequencing (40 base pairs [bp]) was performed on a NextSeq 500 using the High Output v2 kit (75 cycles; Illumina). Libraries were sequenced to a depth of 43-65 million reads/sample.

2.1.12. ATAC-seq library generation

Two different ATAC-seq protocols were used. The original Buenrostro protocol (Buenrostro et al., 2013; 2015a) was used for the optimisation steps and initial experiment, and the Corces Fast-ATAC protocol (Corces et al., 2016) was used for the final experiment.

Buenrostro et al., 2013 protocol

25,000-100,000 cells were sorted per replicate from total PBMCs or CD8-enriched PBMCs. ATAC-seq was performed as previously described (Buenrostro et al.,

2013; 2015a) with minor modifications. Briefly, cells were pelleted at 500 x g for 10 min at 4°C, washed once in 50 µl cold PBS (500 x g, 5 min, 4°C), and resuspended in 50 µl cold lysis buffer (10 mM Tris-HCl, pH 7.4, 10 mM NaCl, 3 mM $MgCl_2$, 0.1% IGEPAL CA-630). Nuclei were immediately pelleted at 500 x g for 15 min at 4°C, resuspended in 50 µl transposition mix (25 µl TD buffer, 2.5 µl TDE1 [FC-121-1030, Illumina, product discontinued], 22.5 µl nuclease-free H_2O), and incubated at 37°C for 30 min. In certain test experiments, an extra wash step (50 µl PBS, 500 x g, 5 min, 4°C) was performed prior to addition of the transposition mix. Transposed DNA was purified using a Qiagen MinElute PCR Purification Kit and eluted in 13 µl elution buffer (10 mM Tris-HCl, pH 8).

Corces et al., 2016 protocol

50,000 cells were sorted per replicate from CD8-enriched PBMCs. ATAC-seq was performed as previously described (Corces et al., 2016). Briefly, cells were pelleted at 500 x g for 10 min at 4°C and supernatant carefully removed. Cells were resuspended in 50 µl cold lysis and transposition mix (25 µl TD buffer, 2.5 µl TDE1 [FC-121-1030, Illumina, product discontinued], 22 µl nuclease-free H_2O, 0.5 µl 1% digitonin [Promega]), and incubated for 30 min at 37°C with agitation at 300 rpm (Thermo-Shaker TS-100, Biosan). Transposed DNA was purified using a Qiagen MinElute Reaction Cleanup Kit and eluted in 13 µl elution buffer (10 mM Tris-HCl, pH 8).

Both protocols

Purified library fragments were PCR amplified for five cycles with barcoded primers using NEBNext High-Fidelity 2X PCR Master Mix (New England Biolabs), followed by an additional six cycles (11 total) or the number of cycles determined by

quantitative PCR (qPCR; as discussed in Chapter 3). Amplified DNA was purified using a Qiagen MinElute PCR Purification Kit (23 µl elution volume) and PCR primer contamination removed using solid phase reverse immobilisation (SPRI) beads (5 min dry time, 15 µl elution volume; Agencourt AMPure XP PCR Purification, Beckman Coulter). Fragment size distribution was analysed on a 2100 Bioanalyzer instrument using the Agilent High Sensitivity DNA Kit. Libraries were quantified using the KAPA Library Quantification Kit (Roche). Paired-end sequencing (40 bp) was performed on a NextSeq 500 using the High Output v2 kit (75 cycles; Illumina). Libraries were sequenced to a depth of 41-58 million paired-end reads/sample (pilot experiment using original ATAC-seq protocol) and 215-271 million paired-end reads/sample (Fast-ATAC in Chapter 4).

2.1.13. 10x Genomics scRNA-seq and scTCR-seq library generation and sequencing

For Experiment 1 and Experiment 2, sequencing libraries were generated using the 10x Genomics Chromium Single Cell V(D)J Reagent Kits (v1.0 Chemistry) following manufacturer's instructions. Cells were loaded onto the Chromium Controller (10x Genomics) at a concentration of ~1 x 10^6 cells/ml, with 6,000-8,000 cells loaded per channel. In the majority of cases, one channel was loaded per donor. However, for each of donors GI3192 and GI3524 in Experiment 1, sorted T_{mem} cells and MAIT cells were mixed at an equal concentration and the cells loaded across two lanes of the Chromium Controller. Library quality and concentration was assessed using a TapeStation (Agilent) and Qubit 2.0 Fluorometer (Thermo Fisher Scientific). For Experiment 1, libraries were sequenced on an Illumina HiSeq 4000 to a mean depth of 39,013-46,998 reads/cell for scRNA-seq (sequencing saturation 72.7-78.1%) and

10,182-30,495 reads/cell for scTCR-seq. For Experiment 2, libraries were sequenced on an Illumina NovaSeq 6000 to a mean depth of 76,000-89,000 reads/cell for scRNA-seq (sequencing saturation 70.0-90.1%) and 11,196-23,869 reads/cell for scTCR-seq. Library generation and sequencing were performed at the Oxford Genomics Centre (Wellcome Centre for Human Genetics, University of Oxford).

2.2. Data analysis

2.2.1. Flow cytometry

Data was analysed in FlowJo 10 (FlowJo, LLC) and graphs were generated using GraphPad Prism 8 (GraphPad Software).

2.2.2. RNA-seq

Processing and differential expression analysis

Read quality was checked using FastQC (v0.11.5; Babraham Bioinformatics) and adapter sequences trimmed using Trimmomatic (v0.36) (Bolger et al., 2014). FastQ files from multiple sequencing lanes were merged, then reads mapped to hg38 using STAR (v2.5.3a) (Dobin et al., 2013). BAM files were filtered using samtools (v1.6) (Li et al., 2009) to retain only properly paired, uniquely mapped, reads. Duplicate reads were marked using Picard MarkDuplicates (v2.15.0; https://broadinstitute.github.io/picard/). Read summarisation was performed using featureCounts (v1.6.0) (Liao et al., 2014). Differentially expressed genes (log_2 fold

change [FC] > 1, false discovery rate [FDR] < 0.05) were identified using edgeR (v3.24.3) (Robinson et al., 2010).

Visualisation

Principal component analysis (PCA) was performed using log-transformed trimmed mean of M-values (TMM)-normalised counts per million (CPM). Heatmaps were generated using centred and scaled log-transformed TMM-normalised CPM.

Gene set enrichment analysis

Gene set enrichment analysis (GSEA) (Subramanian et al., 2005) was performed for all immunologic (C7) gene sets from the Molecular Signatures Database (MSigDB).

2.2.3. ATAC-seq

Preprocessing and differential accessibility analysis

Read quality was checked using FastQC (v0.11.5; Babraham Bioinformatics) and adapter sequences trimmed using Trimmomatic (v0.36) (Bolger et al., 2014). Reads were mapped to hg38 using Bowtie 2 (--very-sensitive -X 2000 --no-mixed --no-discordant --no-unal; v2.3.4.1) (Langmead and Salzberg, 2012). For each sample, BAM files from multiple sequencing lanes were merged using samtools (v1.6) (Li et al., 2009) and duplicate reads removed using Picard MarkDuplicates (v2.15.0; https://broadinstitute.github.io/picard/). BAM files were further filtered to remove reads mapping to the mitochondrial genome (samtools v1.6) and to remove blacklisted regions (bedtools v2.26.0) (Quinlan and Hall, 2010). Blacklisted regions comprised the ENCODE blacklist (Consortium, 2012) and a custom ATAC-seq

blacklist generated by Dr Jason Buenrostro (Buenrostro et al., 2015b). Read start sites were adjusted to correspond to the centre of the transposase binding site. To achieve this, reads on the forward strand were offset by +5 bp and reads on the reverse strand were offset by -4 bp using deepTools alignmentSieve (v3.1.0) (Ramírez et al., 2016). Cut sites were identified as the 5' ends of the forward and reverse reads. Peak calling was performed with MACS2 (Zhang et al., 2008) following the irreproducibility discovery rate (IDR) framework (Li et al., 2011). Briefly, subsampled pseudoreplicate BAM files were generated from each replicate BAM file (self-pseudoreplicates). Biological replicates were merged to produce a pooled BAM file, from which pseudoreplicates (pooled-pseudoreplicates) were also generated. Peaks were called on the replicate, pooled, and pseudoreplicate (self- and pooled-) BAM files using MACS2 with a relaxed p value threshold (--nomodel -f BAMPE --call-summits -p 0.1). IDR was run (using the command-line tool; https://github.com/nboley/idr) on all pairs of replicates, self-pseudoreplicates and pooled-pseudoreplicates (--soft-idr-threshold 0.05 --use-best-multisummit-IDR). The optimal peak list for each T cell subset was determined as the reproducible peak list obtained from IDR analysis of pooled-pseudoreplicates. Optimal peak lists were merged to generate the final peak list, and cut sites within peaks quantified using featureCounts (v1.6.0) (Liao et al., 2014). Peaks were annotated to nearby genes using HOMER annotatePeaks.pl (v4.8) (Heinz et al., 2010) with a custom protein-coding GTF file (see "FASTA and GTF files") and a custom transcription start site (TSS) file (see "Custom TSS file"). Differentially accessible peaks ($log_2FC > 1$, FDR < 0.05) were identified using edgeR (v3.24.3) (Robinson et al., 2010).

Visualisation

BAM files were converted into BigWig files using deepTools bamCoverage (--normalizeUsingRPKM --ignoreForNormalization chrM chrX chrY --binSize 1; v3.1.0)

(Ramírez et al., 2016). PCA was performed using log-transformed TMM-normalised CPM. Heatmaps were generated using centred and scaled log-transformed TMM-normalised CPM. Peak counts normalised using edgeR normalisation factors were used for calculating the Pearson correlation coefficients between pairs of technical and biological replicates.

Pathway enrichment analysis

Pathway enrichment analysis was performed on genes with increased accessibility in MAIT cells compared with T_{mem} cells using the ReactomePA R package (v1.26.0) (Yu and He, 2016). All accessible genes (have ≥ 1 peak with > 1 CPM in ≥ 3 samples) from naïve T cells, MAIT cells, and T_{mem} cells were used as background.

Motif enrichment analysis

Motif enrichment analysis was performed using HOMER (hypergeometric optimisation of motif enrichment) findMotifsGenome.pl (v4.8) (Heinz et al., 2010) and the AME (analysis of motif enrichment) tool from the MEME Suite (motif-based sequence analysis tools) (Bailey et al., 2009; McLeay and Bailey, 2010). Motif enrichment analysis was performed on the peaks with increased accessibility in MAIT cells relative to T_{mem} cells, with shared peaks (peaks with similar accessibility in MAIT cells and T_{mem} cells) used as background/control sequences. HOMER analysis was performed once with whole peaks and once with peak centres (± 100 bp around peak summits) using a repeat-masked hg38 genome sequence. Motifs with $p < 1 \times 10^{-12}$ were considered significantly enriched. AME motif enrichment analysis was performed on peak centres (sequence scoring method: average odds score; motif enrichment test: Fisher's exact test) using either the "Vertebrates (*In*

vivo and *in silico*)" or the "HOCOMOCO Human (v11 CORE)" motif database, as specified in the text.

CENTIPEDE TF footprinting

CENTIPEDE TF footprinting analysis was performed as described (Pique-Regi et al., 2011). Briefly, genome-wide instances of TF binding motifs were identified using HOMER scanMotifsGenome.pl (v4.8) (Heinz et al., 2010). The nearest TSS to each motif instance was determined using the bedtools closest command (-d -io -t first; v2.26.0) with hg38 genome-wide TSSs downloaded from the UCSC Table Browser. The average conservation score for nucleotides within each motif instance was determined using the UCSC bigWigAverageOverBed function and UCSC phastCons conservation scores (hg38.100way.phastCons.bw). Genome-wide motif instances were extended 100 bp at either end and the number of cut sites at each position of these extended regions was determined for a given sample (e.g. MAIT cells, replicate 1) using bedtools coverage (-d -s/-S; v2.26.0) (Quinlan and Hall, 2010). This was performed separately for cut sites on the same strand as the motif instance and cut sites on the opposite strand. Finally, the number of cut sites at each position was averaged for all motif regions on the strand. The position weight matrix (PWM) score from HOMER scanMotifsGenome.pl, the conservation score, and the nearest TSS for each motif instance, along with the average number of cut sites around the motif site of a given TF on each strand, were used as input for CENTIPEDE analysis.

2.2.4. ATAC-seq and RNA-seq

FASTA and GTF files

A soft-masked hg38 FASTA file was downloaded from UCSC and chr_alt chromosomes were removed. The corresponding gene annotation file (GTF format) was downloaded from Ensembl (genome-build GRCh38.p12). The Ensembl chromosome names in the GTF file were replaced with UCSC chromosome names using a chromosome mappings file from Dr Devon Ryan (https://github.com/dpryan79/ChromosomeMappings). For RNA-seq read summarisation using featureCounts (v1.6.0) (Liao et al., 2014) and for ATAC-seq peak annotation using HOMER (v4.8) (Heinz et al., 2010), the GTF file was filtered to contain only protein-coding genes (hereafter referred to as the protein-coding GTF file). The protein-coding GTF file was further modified for use with HOMER annotatePeaks.pl. We wanted to annotate each ATAC-seq peak to a gene, not a transcript, but annotatePeaks.pl uses transcript IDs for annotation. To overcome this issue, transcript IDs in the attribute column of the GTF file were replaced with the corresponding Ensembl gene IDs.

Custom TSS file

To generate a custom TSS file for HOMER annotatePeaks.pl (v4.8) (Heinz et al., 2010), BioMart (Durinck et al., 2005) was used to obtain the chromosome name, transcript IDs, TSS, and GENCODE basic annotation for each Ensembl gene ID (chromosome_name, ensembl_transcript_id, transcription_start_site, transcript_gencode_basic) from the protein-coding GTF file. Each Ensembl gene ID can be associated with multiple transcripts and thus TSSs. The cDNA coding start and stop sites (cdna_coding_start, cdna_coding_end) were obtained for each

transcript ID, and the length of the coding region calculated. For each Ensembl gene ID, only GENCODE basic transcripts were retained. From the remaining transcripts, the TSS from the transcript with the longest coding region was used to generate the custom TSS file for HOMER. The custom TSS file is essentially a BED file with 5 columns: the Ensembl gene ID, the UCSC chromosome name, the feature start site, the feature end site, and the strand. The TSS is interpreted as the mid-point between the start and end coordinates.

Spearman correlation plot

ATAC-seq peak counts (number of transposase cut sites per peak) were normalised by library size and the mean counts per peak calculated for naïve, MAIT and T_{mem} cells from the three biological replicates. Spearman's rank correlation coefficients (ρ) between mean peak counts were calculated for pairwise combinations of naïve, MAIT and T_{mem} cells. Similarly, RNA-seq counts were normalised by library size, mean counts per gene calculated for naïve, MAIT and T_{mem} cells, and Spearman's rank correlation coefficients determined for pairwise combinations of naïve, MAIT and T_{mem} cells. Each circle in the diagram (Fig. 4.3C) indicates a cell type, and the lines connecting pairs of cell types are scaled to one minus the Spearman's rank correlation coefficient. Thus, the longer the line, the weaker the correlation between cell types.

IMAGE

IMAGE (integrated analysis of motif activity and gene expression changes of transcription factors) analysis was performed according to guidelines (https://github.com/JesperGrud/IMAGE) (Madsen et al., 2018) using RNA-seq and ATAC-seq data from MAIT cells and T_{mem} cells. Gene expression data in the form of

a raw count matrix from featureCounts (v1.6.0) (Liao et al., 2014) was transformed into the required table format for IMAGE. Ensembl gene IDs were converted to RefSeq IDs and HGNC gene symbols using BioMart (Durinck et al., 2005). ATAC-seq raw counts from featureCounts were converted into TMM-normalised reads per kilobase million (RPKM) using edgeR (v3.24.3) (Robinson et al., 2010), and a table with the appropriate columns generated for IMAGE. The "Result" dataframe from IMAGE analysis was filtered to keep only high confidence causal TFs (CausalTF = 1) with motifs established using direct TF binding assays, such as ChIP-seq and SELEX (Evidence = Direct).

2.2.5. 10x Genomics scRNA-seq and scTCR-seq

Experiment 1

Pre-processing

10x Genomics Cell Ranger (https://github.com/10XGenomics/cellranger) analysis pipelines were used to generate single cell gene counts. Briefly, FASTQ files were generated from Illumina BCL (binary base call) files using cellranger mkfastq (v2.1.1 for donors GI2916 and GI3008, v3.0.1 for donors GI3524 and GI3192). Reads were aligned to the 10x Genomics human GRCh38 reference genome (v3.0.0) and quantified using cellranger count (v3.0.1). Output files were aggregated using cellranger aggr (v3.0.1) with read depth normalised (by number of confidently mapped reads per cell) between samples.

Quality control

Unfiltered feature-barcode matrices (raw_feature_bc_matrices) from cellranger aggr were imported into R and converted into a Seurat object using the Seurat R

package (v3.1.1) (Butler et al., 2018). Only cells expressing at least 200 genes and genes expressed in at least 10 cells were retained. The aggregated Seurat object was subset so that samples could be analysed individually. TCR (and BCR) genes were removed from the data to ensure that downstream clustering analysis was not influenced by TCR (or BCR) chain usage. Cells were filtered to remove those with very low or very high UMI counts. The lower UMI threshold, intended to remove dead cells and cell-free RNA, was set at the local minimum of the UMI distribution to the left of the mode UMI count. The upper UMI threshold was determined individually for each sample using a novel approach that made use of the matched TCR data (see Chapter 3 for further information). Appropriate setting of the upper UMI threshold is important to retain as many viable cells as possible, whilst removing cell multiplets (predominantly doublets). Further cell filtering was performed to remove outliers that were potentially dead cells or multiplets (cells with ≤ 500 genes, ≥ 2,500 genes and/or ≥ 10% mitochondrial reads were removed). Finally, cells with more than two TCRα and/or TCRβ chains were assumed to be multiplets and were removed from the analysis. The samples contained a median of 2,940 cells (range: 1,930-4,035) post-filtering. For combined analysis of all samples, TCR (and BCR) genes were removed from the aggregated Seurat object, then cells filtered to retain only those that passed quality filtering in the individual sample analyses.

Normalisation, dimensionality reduction and clustering

Analysis was performed using the Seurat R package (v3.1.1) (Butler et al., 2018). For individual sample analyses, count data was normalised and variable genes identified (top 3000 most variable) using SCTransform (Hafemeister and Satija, 2019) with mitochondrial read fraction regressed out. Dimensionality reduction was performed using PCA with the variable genes as input, and cell clusters were

identified using Seurat's graph-based clustering approach (0.5 resolution). Briefly, a shared nearest neighbour (SNN) graph was built on dimensionally reduced data (top 30 PCs), then clusters determined by optimising the standard modularity function. Data was visualised by uniform manifold approximation and projection (UMAP) (McInnes et al., 2018) on the top 30 PCs.

For samples containing MAIT cells and T_{mem} cells (i.e. samples from donors GI3192 and GI3524), the Seurat object was subset post-clustering to retain only MAIT cell clusters. MAIT cell clusters were defined as clusters where the majority of cells expressed a *TRAV1-2*/*TRAJ33*, *TRAV1-2*/*TRAJ12* or *TRAV1-2*/*TRAJ20* TCRα chain. Although GI2916 did not contain sorted T_{mem} cells, a non-*TRAV1-2* cluster of $CD8^+$ T cells was identified that lacked expression of classical MAIT cell markers, including *ZBTB16*, *IL23R*, *RORC*, and *DPP4*. Therefore, this cluster was assumed to be contaminating conventional T cells and was removed from the analysis. To identify clusters within the MAIT cell population, all analysis steps from SCTransform onwards were repeated on the subset Seurat object.

For the aggregated analysis of all samples, count data was normalised using SCTransform (Hafemeister and Satija, 2019). The variable gene list generated by SCTransform was replaced with a pooled list of the variable genes identified in the individual samples (8,378 genes total). Dimensionality reduction was performed by PCA, then the dimensionally reduced space modified using Harmony (Korsunsky et al., 2019) to correct for donor-specific effects. The top 30 Harmony dimensions were used for graph-based clustering and UMAP.

Differential expression analysis

In individual sample analyses, cluster markers (genes differentially expressed between each cluster and all other cells) were identified using the FindMarkers command with the Wilcoxon Rank Sum test ($log_2FC \geq 0.25$, adjusted [Bonferroni correction] p value < 0.01). In the combined analysis of all samples, differentially expressed genes between blood and liver MAIT cells, and between $CD8^+$ and DN MAIT cells, were identified using the FindMarkers command with the logistic regression test and donor as a latent variable ($log_2FC \geq 0.25$, adjusted [Bonferroni correction] p value < 0.01).

GSEA and pathway enrichment analysis

GSEA (GSEA Desktop v4.0.3) (Subramanian et al., 2005) was performed for published $CD8^+$ T_{RM} (Milner et al., 2017) and $CD4^+$ T_{RM} (Beura et al., 2019) gene signatures. Pathway enrichment analysis was performed using the ReactomePA R package (v1.30.0) (Yu and He, 2016). All MAIT cell-expressed genes were used as background.

Experiment 2

Pre-processing

10x Genomics Cell Ranger (https://github.com/10XGenomics/cellranger) analysis pipelines were used to generate single cell gene counts. Briefly, FASTQ files were generated from Illumina BCL files using cellranger mkfastq (v3.0.2). Reads were aligned to the 10x Genomics human GRCh38 reference genome (v3.0.0) and quantified using cellranger count (v3.0.1).

Quality control

Quality control was performed separately for each sample. Unfiltered feature-barcode matrices (raw_feature_bc_matrices) from cellranger count were imported into R and converted into a Seurat object using the Seurat R package (v3.1.1) (Butler et al., 2018). Only cells expressing at least 200 genes and genes expressed in at least 10 cells were retained. TCR (and BCR) genes were removed from the data to ensure that downstream clustering analysis was not influenced by TCR (or BCR) chain usage. Cells with UMI counts ≤ 2 median absolute deviations (MADs) below the median, gene counts ≤ 2 MADs below the median, and/or mitochondrial read fraction ≥ 2 MADs above the median were removed. Finally, cells with more than two TCRα and/or TCRβ chains were assumed to be multiplets and were removed from the analysis. The samples contained a median of 2,629 cells (range: 2,234-5,562) post-filtering. For combined analysis of multiple samples (e.g. the three stimulation conditions for each donor or all samples), TCR (and BCR) genes were removed from the combined Seurat object, then cells filtered to retain only those that passed quality filtering in the individual sample analyses.

Normalisation, dimensionality reduction and clustering

Analysis was performed using the Seurat R package (v3.1.1) (Butler et al., 2018). Count data was normalised and variable genes identified (top 3000 most variable) using SCTransform (Hafemeister and Satija, 2019). Dimensionality reduction was performed using PCA with the variable genes as input. For the analysis of individual samples, or multiple samples from the same donor, cell clusters were identified using Seurat's graph-based clustering approach (0.5 resolution). Briefly, an SNN graph was built on dimensionally reduced data (top 30 PCs), then clusters

determined by optimising the standard modularity function. Data was visualised by UMAP (McInnes et al., 2018) on the top 30 PCs.

For the combined analysis of multiple samples from different donors, PCA was performed and then a modified dimensionally reduced space was generated using Harmony (Korsunsky et al., 2019) to correct for donor-specific effects. The top 30 Harmony dimensions were used for graph-based clustering and UMAP.

Differential expression analysis

In individual sample analyses, cluster markers (genes differentially expressed between each cluster and all other cells) were identified using the FindMarkers command with the Wilcoxon Rank Sum test ($log_2FC \geq 0.25$, adjusted [Bonferroni correction] p value < 0.01). In the combined analysis of all samples, differentially expressed genes between unstimulated and TCR-stimulated, or unstimulated and cytokine-stimulated, MAIT cells were identified using the FindMarkers command with the logistic regression test and donor as a latent variable ($log_2FC \geq 0.25$, adjusted [Bonferroni correction] p value < 0.01).

Pathway enrichment analysis

Pathway enrichment analysis was performed using the ReactomePA R package (v1.30.0) (Yu and He, 2016). All genes expressed in unstimulated MAIT cells were used as background.

Experiment 1 and 2 scTCR-seq analysis

Pre-processing

scTCR-seq data was processed using Cell Ranger pipelines (v2.1.1-3.0.1 for Experiment 1, v3.0.2 for Experiment 2). Specifically, Illumina BCL files were converted to FastQ files using cellranger mkfastq. Single-cell V(D)J sequences were assembled using cellranger vdj. The filtered contigs output file (filtered_contig_annotations.csv) was filtered for downstream analysis. Only high-confidence, productive contigs associated with *TRA* or *TRB* chains were retained. The data was then collapsed by cell barcode to generate a table (one per sample) with cells as rows and TCR gene segments as columns.

TCR clonotype analysis

The table generated as described above was used to define nucleotide TCR clonotypes for each sample, or tables from multiple samples were combined for nucleotide and/or functional TCR clonotype analysis. This enabled nucleotide clonotype comparisons between matched blood and liver samples, and functional clonotype comparisons between donors in Experiment 1. In Experiment 2, nucleotide clonotypes were compared across stimulation conditions, while functional clonotypes were compared between donors.

Clonotypes were defined for cells that passed quality filtering and had a *TRAV1-2* TCRα chain and at least one TCRβ chain. Nucleotide clonotypes were defined as cells with identical CDR3α and CDR3β nucleotide sequences. If a cell had two CDR3α and/or two CDR3β sequences, both were required to match. TCRα and TCRβ only clonotypes were defined as cells with identical CDR3α or CDR3β nucleotide sequences, respectively. Functional TCR clonotypes were defined as

cells with at least one identical CDR3α and CDR3β amino acid sequence. Functional TCRα and TCRβ clonotypes were defined as cells with at least one matching CDR3α or CDR3β amino acid sequence, respectively.

TCR segment usage and CDR3 length/sequence analysis

Analysis of *TRAJ* segment usage, and CDR3α length and sequence, was performed for *TRAV1-2* TCRα chains from cells with a defined TCR clonotype. Analysis of TCRβ chain segment usage, and CDR3β length and sequence, was performed for all TCRβ chains from cells with a defined clonotype (all have a *TRAV1-2* TCRα chain).

Association of TCR usage with gene expression

TCR chain usage and clonotype assignments were imported as metadata into the Seurat object containing gene expression data for the corresponding cells. The location of cells expressing particular TCR chains or within a particular TCR clonotype could then be mapped onto the UMAP plot. Differentially expressed genes between clonotypes were identified using the FindMarkers command with the Wilcoxon Rank Sum test ($\log_2$FC ≥ 0.25, adjusted [Bonferroni correction] p value < 0.01).

3. Method development: optimisation of MR1 tetramer staining and the ATAC-seq protocol

3.1. Introduction

During my DPhil, I developed or optimised several methods to ensure that my data were of the highest possible quality. I optimised MR1 tetramer staining for the specific identification of MAIT cells and the ATAC-seq protocol for the generation of high quality chromatin accessibility data. In addition, I developed a method to more accurately exclude doublets from paired scRNA-seq and scTCR-seq data generated using the droplet-based approach from 10x Genomics.

3.1.1. MAIT cell identification

MAIT cells were originally identified as a population of T cells with an invariant Vα7.2-Jα33 TCRα chain (Porcelli et al., 1993; Tilloy et al., 1999). In 2009, Martin et al. developed a monoclonal antibody against the human Vα7.2 TCRα segment and showed that the canonical MAIT cell Vα7.2-Jα33 sequence was found exclusively within $V\alpha7.2^{+}CD161^{hi}$ T cells (Martin et al., 2009). Coexpression of Vα7.2 and high levels of CD161 became the principal method for the identification of MAIT cells. However, MAIT cells can be identified using other surrogate markers, for example as $V\alpha7.2^{+}$IL-18Rα^{+} (Le Bourhis et al., 2010) or $V\alpha7.2^{+}CD26^{+}$ (Sharma et al., 2015) T cells.

In 2012, MAIT cell ligands were identified as metabolites from the riboflavin biosynthetic pathway found in many bacteria and yeast (Kjer-Nielsen et al., 2012). Generation of these ligands was shown to require 5-A-RU, an early intermediate in this pathway (Corbett et al., 2014). 5-A-RU undergoes non-enzymatic condensation with the small metabolites glyoxal and methylglyoxal (derived from a number of metabolic pathways e.g. glycolysis) to generate the potent MAIT cell agonists 5-OE-RU and 5-OP-RU, respectively. Identification of MAIT cell ligands enabled the generation of tetramers for specific MAIT cell identification (Corbett et al., 2014; Reantragoon et al., 2013). MR1/5-OP-RU tetramers are now available from the NIH Tetramer Core Facility and MR1 tetramer staining is considered the gold standard approach for MAIT cell identification. Nevertheless, surrogate identification approaches remain widely used.

MAIT cells are predominantly $CD8^+$ or DN T cells. In peripheral blood, identification of MAIT cells as either $V\alpha7.2^+CD161^{hi}$ or MR1/5-OP-RU^+ defines essentially the same population amongst $CD8^+$ or DN T cells (Gherardin et al., 2018b). However, use of the MR1 tetramer is important in situations where MAIT cells are infrequent, such as during early development or in the $CD4^+$ T cell population. In cord blood, only 2-15% of $V\alpha7.2^+CD161^{hi}CD4^-$ T cells are MR1/5-OP-RU^+ (Ben Youssef et al., 2018). Within the $CD4^+$ T cell population in adult peripheral blood, a median of 33.7% of $V\alpha7.2^+CD161^{hi}$ T cells are MR1/5-OP-RU^+, with the percentage varying widely between individuals (Gherardin et al., 2018b; Kurioka et al., 2017).

The relative accuracy of MAIT cell identification using either surrogate MAIT cell markers or the MR1/5-OP-RU tetramer has yet to be systematically examined in human tissues or during disease. However, indirect evidence suggests that identification with the MR1 tetramer may be more accurate. $CD161^+$ T cells are

enriched in a number of tissues relative to blood, including the liver and gut (Billerbeck et al., 2010; Fergusson et al., 2015; Ishihara et al., 1999; O'Keeffe et al., 2004), and non-specific antibody staining is often increased in tissues. This can make it difficult to distinguish Vα7.2^{+}CD161^{hi} MAIT cells from conventional T cells. CD161 is downregulated on MAIT cells in response to *E. coli* stimulation *in vitro*, in correlation with cell division (Kurioka et al., 2015). Moreover, a clear population of Vα7.2^{+}CD161^{-} T cells can be observed after one-day *E. coli* stimulation of sorted Vα7.2^{+}CD161^{hi} MAIT cells (Leeansyah et al., 2013), implying CD161 downregulation. In HIV, the percentage of Vα7.2^{+}CD161^{-} cells amongst total T cells increases, while that of Vα7.2^{+}CD161^{hi} cells decreases (Fernandez et al., 2015; Leeansyah et al., 2013). This was suggested to result from CD161 downregulation on MAIT cells (Leeansyah et al., 2013), but a more recent study disputes this hypothesis, showing that Vα7.2^{+}CD161^{-} T cells in HIV patients are MR1/5-OP-RU^{-} (Fernandez et al., 2015). This finding should be verified in future studies. In early rheumatoid arthritis patients, the mean fluorescence intensity (MFI) of CD161 on MR1/5-OP-RU^{+} MAIT cells is decreased relative to healthy controls (Koppejan et al., 2019). Therefore, there is some evidence for a decrease in CD161 expression on MAIT cells following activation and/or in disease. If CD161 is downregulated, then gating on Vα7.2^{+}CD161^{hi} T cells would fail to capture the entire MAIT cell population.

In summary, the MR1 tetramer is now widely available, but many groups continue to use surrogate markers, such as Vα7.2 and CD161, for MAIT cell identification. While surrogate approaches detect the majority of MAIT cells in human peripheral blood (Gherardin et al., 2018b), their accuracy may be limited in some settings, for example in tissue. Thus, there is a need to optimise the protocol for MR1 tetramer

staining in human tissues and to compare the accuracy of MR1 tetramer staining with the use of surrogate markers for tissue MAIT cell identification.

3.1.2. ATAC-seq method

ATAC-seq is a method for assaying chromatin accessibility genome-wide, developed by Buenrostro and colleagues in 2013 (Buenrostro et al., 2013). The method uses a hyperactive Tn5 transposase that cuts and tags the genome with sequencing adapters in regions of open chromatin. Tagged DNA fragments are then PCR amplified, sequenced and mapped to the reference genome. Regions with a high density of mapping reads denote accessible sites, enriched for regulatory elements such as promoters and enhancers.

ATAC-seq has become a popular method for analysing DNA accessibility due to its advantages over previous methods (Tsompana and Buck, 2014), for example DNase-seq (Boyle et al., 2008; Song and Crawford, 2010) and FAIRE-seq (Giresi et al., 2007; Nagy et al., 2003). ATAC-seq is simple and quick to perform. The standard protocol uses 50,000 cells, but can be performed with as few as 500 cells, and single-cell protocols have now been developed (Buenrostro et al., 2015b; Cusanovich et al., 2015). This contrasts with the millions of cells typically required for DNase-seq and FAIRE-seq. Therefore, ATAC-seq can be performed on primary cells, in particular on rare populations or clinical samples, where cell numbers are often limiting. Moreover, ATAC-seq has a high signal-to-noise ratio, comparable to DNase-seq (and superior to FAIRE-seq), despite requiring three to five orders of magnitude fewer cells.

Nevertheless, a number of aspects of the ATAC-seq protocol can be optimised to improve data quality. ATAC-seq libraries must be PCR amplified to provide sufficient material for sequencing, but overamplification can introduce GC content and template length bias (Meyer and Liu, 2014). Therefore, it is important to determine an appropriate number of PCR cycles for each individual ATAC-seq library depending on the input cell number and concentration of transposed DNA. Studies have shown that mitochondrial DNA (mtDNA) is highly susceptible to transposase integration, resulting in a high fraction of ATAC-seq reads mapping to mtDNA (Buenrostro et al., 2015a). This may be explained by the lack of nucleosomes associated with mtDNA (Alexeyev et al., 2013). The mitochondrial read fraction in ATAC-seq libraries should be minimised in order to maximise the number of reads mapping to nuclear DNA. Finally, post-amplification ATAC-seq libraries may be contaminated with a high proportion of free primer. Free primers take up valuable space on the flow cell when sequencing and can increase the incidence of index hopping (Illumina technical note). Therefore, their removal is desirable.

In summary, ATAC-seq provides advantages over previous methods for the generation of chromatin accessibility data. However, there are still a number of steps within the protocol that can be optimised to improve data quality.

3.1.3. Removal of multiplets from paired scRNA-seq and scTCR-seq data

10x Genomics uses a droplet-based approach for single-cell genomic analysis, including paired scRNA-seq and scTCR-seq. Within the microfluidic channels of the

Chromium Controller, individual cells are combined with oligonucleotide-coated Gel Beads, reverse transcription reagents and oil to form isolated reaction chambers called Gel Beads-in-emulsion (GEM). Polyadenylated mRNAs are reverse transcribed within GEMs to produce full-length cDNA, then GEMs are broken and cDNA from all cells is PCR amplified together. Amplified cDNA is divided for the generation of gene expression and TCR-sequencing libraries.

During the formation of GEMs, more than one cell can inadvertently be captured within a single GEM, thus generating multiplets i.e. GEMs containing more than one cell. During reverse transcription within GEMs, a cell-specific barcode (10x Barcode) is incorporated into the cDNA. In the case of multiplets, DNA derived from all cells will be labelled with an identical 10x Barcode and therefore will appear to originate from the same cell. To minimise the formation of multiplets, cells are loaded onto the Chromium Controller at a limiting dilution such that most droplets are empty, with the remaining GEMs predominantly containing a single cell. Nevertheless, some droplets will still contain two (or more) cells.

Multiplets, which are predominantly doublets, can have a major impact on the downstream analysis (e.g. clustering, differential gene expression analysis) and interpretation of single-cell data (Ilicic et al., 2016). For example, doublets containing cells from distinct developmental stages would exhibit a mixed transcriptional profile and could be misinterpreted to be cells of an intermediate developmental state. Thus, the identification and removal of multiplets is an essential analysis step.

Potential multiplets are typically removed from scRNA-seq data by excluding cells with a UMI count or gene count that exceeds an arbitrary threshold (Luecken and

Theis, 2019), as described in the analysis tutorials for Seurat (https://satijalab.org/seurat/), a popular R package for single-cell analysis. UMI and gene counts can vary between cells for biological (e.g. cell type, cell cycle stage) and technical (e.g. mRNA capture efficiency) reasons (Stegle et al., 2015). Although true multiplets do exhibit an increased average UMI count relative to singlets, their UMI counts vary widely, with the distribution of UMI counts in multiplets overlapping with that of singlets (Stoeckius et al., 2018). Therefore, this approach fails to unambiguously identify multiplets.

Over the last couple of years, a number of methods have been developed to improve the accuracy of doublet detection in scRNA-seq (Gehring et al., 2018; Guo et al., 2019; Kang et al., 2018; McGinnis et al., 2019a; 2019b; Shin et al., 2019; Stoeckius et al., 2018; Wainer-Katsir and Linial, 2019; Wolock et al., 2019). Many of these approaches require multiple samples to be combined in a single assay, as doublets are detected as “cells” (labelled with a single 10x Barcode in the 10x Genomics system) displaying markers of more than one sample. For example, samples can be labelled with distinct oligonucleotide barcodes prior to pooling and sequencing (Gehring et al., 2018; Guo et al., 2019; McGinnis et al., 2019b; Stoeckius et al., 2018). Alternatively, cells from different donors can be distinguished based on natural genetic variation (Kang et al., 2018). The number of unidentifiable doublets, namely doublets containing cells from a single sample, is inversely correlated with the number of multiplexed samples. Contrasting with the methods dependent on genomic variation, oligonucleotide labelling approaches could theoretically be applied to a single sample. Cells could be divided, labelled with distinct barcodes, and repooled for sequencing. However this is somewhat impractical and is unfeasible for samples with limited cell number.

Additionally, computational approaches have been developed to identify doublets based on their mixed transcriptional profile (DePasquale et al., 2019; McGinnis et al., 2019a; Wolock et al., 2019) or biallelic expression pattern (Wainer-Katsir and Linial, 2019). These approaches also have limitations. For example, the former method can only detect heterotypic doublets containing transcriptionally distinct cells, while the latter approach requires high read coverage and the availability of genomic data for each donor.

Overall, there is no optimal method for the identification of multiplets in scRNA-seq data and the applicability of each approach is sample-specific. Thus, new approaches are required, particularly those compatible with previously generated data. Ultimately, combining several multiplet detection methods is likely to be the most effective approach.

3.2. Results

3.2.1. MAIT cell identification

Tetramer titration

Biotinylated MR1/5-OP-RU (MR1 bound to activating ligand) and MR1/6-FP (MR1 bound to non-activating ligand) monomers were each separately tetramerised with streptavidin-PE and streptavidin-BV421, according to protocols from the NIH Tetramer Core Facility (https://tetramer.yerkes.emory.edu). MR1/5-OP-RU tetramers should bind to all MAIT cells, whereas MR1/6-FP tetramers should not bind to MAIT cells and can therefore be used to examine the degree of non-specific tetramer staining (Corbett et al., 2014).

MR1/5-OP-RU and MR1/6-FP conjugated to PE and BV421 were titrated on healthy human PBMCs from one donor. The frequency of MAIT cells detected using MR1/5-OP-RU on PE and BV421 (hereafter MR1/5-OP-RU^{PE} and MR1/5-OP-RU^{BV421}, respectively) was comparable (Fig. 3.1A). As tetramer dilution increased from 1/100 to 1/1600, the MFI of PE and BV421 decreased (Fig. 3.1B-E). MR1/5-OP-RU^{+} T cells were clearly distinguishable from non-MAIT CD3^{+} T cells even at the highest tetramer dilution and background staining with MR1/6-FP was low (< 0.3% of CD3^{+} T cells) (Fig. 3.1B,C). The stain index (stain index = [MFI positive peak - MFI negative peak]/2 x standard deviation negative peak) for MR1/5-OP-RU^{PE} and MR1/5-OP-RU^{BV421} was highest at a 1/100 dilution. However, as MR1/5-OP-RU^{+} and MR1/5-OP-RU^{-} T cells remained distinct at higher dilutions, I opted to use MR1/5-OP-RU^{PE} at 1/400 and MR1/5-OP-RU^{BV421} at 1/800 for future experiments to save reagent. MR1/6-FP tetramers were used at equivalent concentrations. As an aside, the donor used for MR1 titration had an unusual population of MR1/5-OP-RU^{+} T cells that were CD161$^{low\text{-}neg}$ (Fig. 3.1A-C). This should not have interfered with tetramer titration.

Optimisation of tetramer staining conditions

Using the titrated concentrations of MR1/5-OP-RU$^{PE/BV421}$ and MR1/6-FP$^{PE/BV421}$, I next optimised the MR1 tetramer staining protocol on healthy control PBMCs. I optimised the time and temperature of tetramer staining, and tested whether tetramer staining should be performed prior to surface antibody staining, or whether tetramer and surface antibody staining can be performed simultaneously. In addition, in an attempt to increase the specificity of MAIT cell detection, I performed combined staining with MR1/5-OP-RU^{PE} and MR1/5-OP-RU^{BV421}.

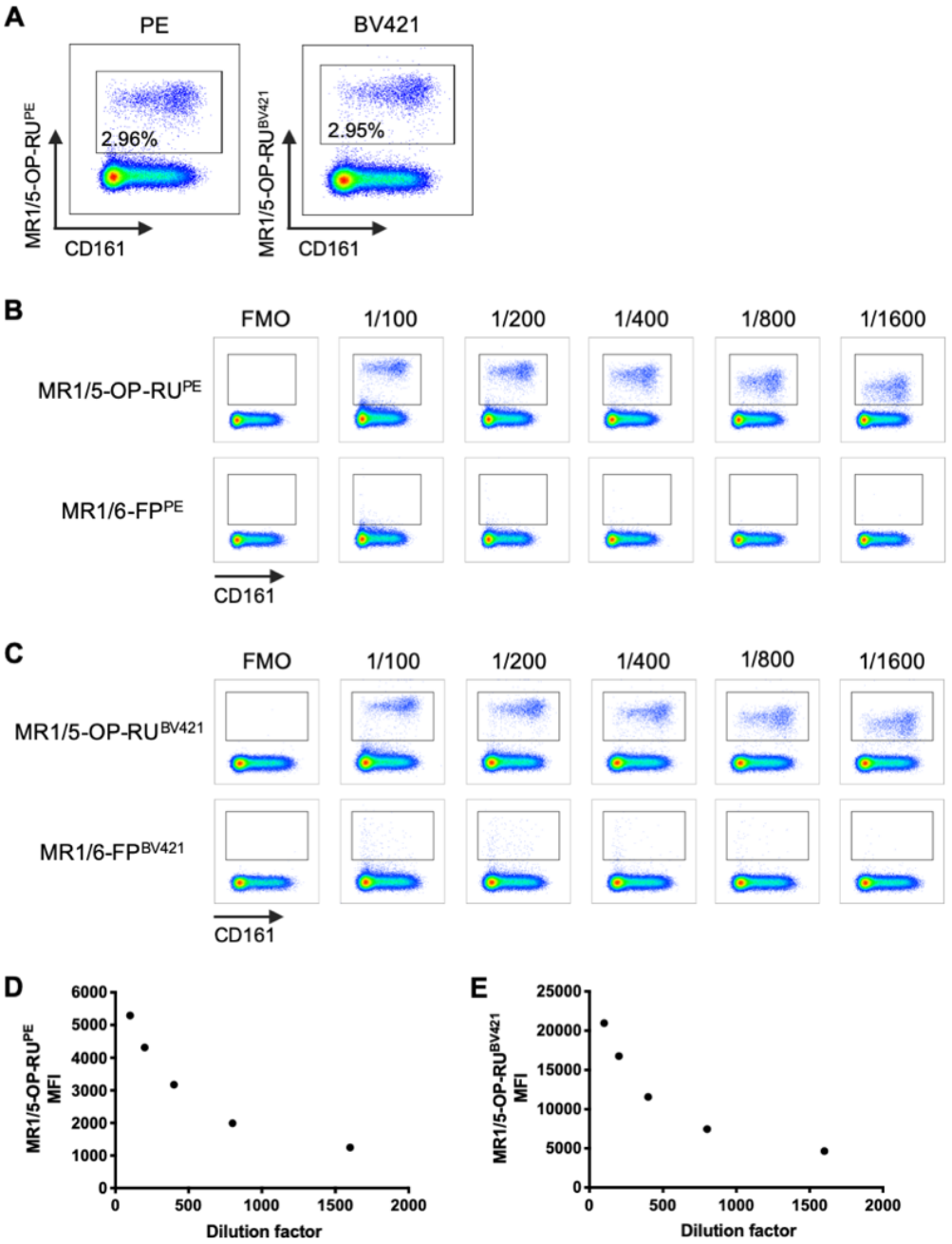


Figure 3.1. Titration of MR1/5-OP-RU and MR1/6-FP conjugated to PE and BV421. **(A)** Representative FACS plots demonstrating an equivalent frequency of MR1/5-OP-RU^{PE+} and MR1/5-OP-RU^{BV421+} T cells. **(B,C)** FACS plots showing MR1/5-OP-RU^{PE} and MR1/6-FP^{PE} (B) or MR1/5-OP-RU^{BV421} and MR1/6-FP^{BV421} (C) staining at two-fold dilutions from 1/100 to 1/1600, along with an FMO control for each tetramer. **(D,E)** MFI of MR1/5-OP-RU^{PE} (D) and MR1/5-OP-RU^{BV421} (E) at two-fold dilutions from 1/100 to 1/1600.

MR1 tetramer staining was examined under the following conditions:

1. MR1 tetramer (PE or BV421) combined with surface panel
2. MR1 tetramer (40 min at 4°C) followed by surface panel
3. MR1 tetramer (40 min at RT) followed by surface panel
4. MR1 tetramer (40 min at 37°C) followed by surface panel
5. MR1 tetramer (40 min at RT) followed by surface panel containing Vα7.2
6. PE and BV421 MR1 tetramers combined (40 min at RT), followed by surface panel
7. Surface panel containing Vα7.2 and CD161 (no MR1 tetramer)

Surface panel staining was performed for 30 minutes at 4°C. Barring Condition 1, staining with the MR1 tetramer was performed for 40 minutes, as recommended by the NIH Tetramer Core Facility. Under each staining condition, I measured the MFI of MR1/5-OP-RU$^{PE/BV421}$ and the degree of background staining with MR1/6-FP$^{PE/BV421}$. In addition, the frequency of MR1/5-OP-RU^{+} T cells was compared with the frequency of Vα7.2^{+}CD161^{hi} T cells detected in Condition 7 (no MR1 tetramer control). The frequency of these two populations should be equivalent in healthy adult peripheral blood (Gherardin et al., 2018b).

The frequency of MAIT cells detected using the MR1 tetramer was comparable in Condition 1 and 2 (PE: 3.26% and 3.14%, BV421: 3.14% and 3.26% of CD3^{+} T cells for Condition 1 and 2, respectively), however the MFI of MR1/5-OP-RU$^{PE/BV421}$ was higher when tetramer staining was performed prior to (Condition 2), rather than together with (Condition 1), surface antibody staining (Fig. 3.2A). While the duration of tetramer staining was slightly shorter in Condition 1 compared with Condition 2 (30 min vs. 40 min), this difference is unlikely to account for the observed difference in tetramer MFI.

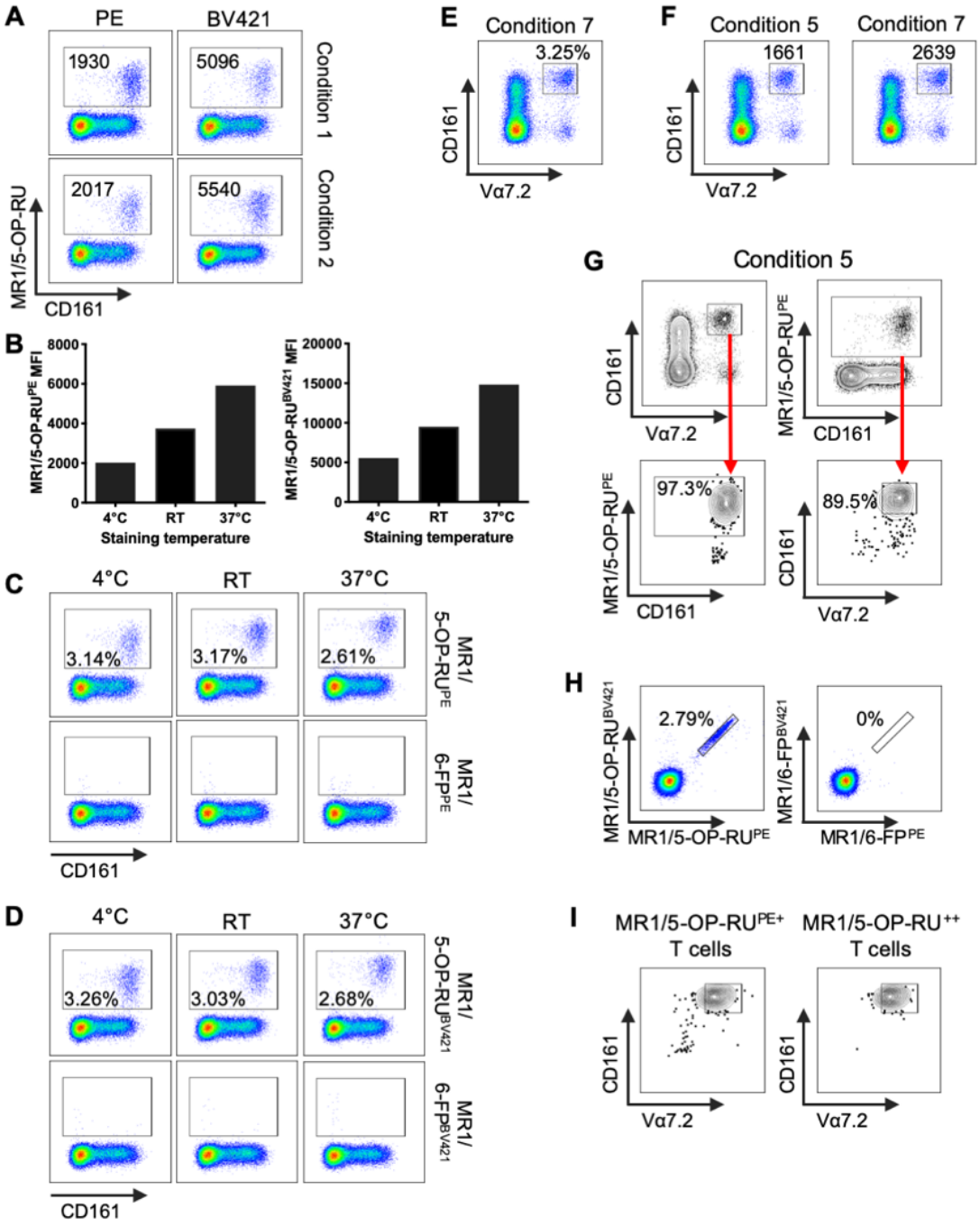


Figure 3.2. Optimisation of the MR1 tetramer staining protocol. (A) MFI of PE/BV421 on MR1/5-OP-RU$^{PE/BV421+}$ T cells when tetramer staining was performed prior to (bottom) or together with (top) surface antibody staining. **(B)** Bar plot showing the MFI of PE/BV421 on MR1/5-OP-RU$^{PE/BV421+}$ T cells when staining was performed at 4°C, RT or 37°C. **(C,D)** FACS plots indicating the frequency of MR1/5-OP-RU^{PE+} (C) or MR1/5-OP-RU^{BV421+} (D) T cells detected at different staining temperatures. **(E)** FACS plot showing the frequency of Vα7.2^{+}CD161^{hi} T cells (staining panel without the MR1 tetramer). **(F)** Vα7.2 MFI when staining was performed subsequent to (left), or in the absence of (right), MR1/5-OP-RU tetramer staining. **(G)** FACS plots showing the fraction of MR1/5-OP-RU^{PE+} T cells amongst Vα7.2^{+}CD161^{hi} T cells (left) and vice versa (right). **(H)** MAIT cell identification using two MR1/5-OP-RU tetramers (left) and background staining with MR1/6-FP (right). **(I)** Vα7.2 and CD161 staining on MR1/5-OP-RU^{PE+} (left) and MR1/5-OP-RU^{++} (right) T cells. FACS plots in A, C-F and H are gated on live CD3^{+} T cells.

The MFI of MR1/5-OP-RU$^{PE/BV421}$ increased with staining temperature (Fig. 3.2B-D), while MR1/6-FP background staining was low at all temperatures (Fig. 3.2C,D). At 4°C (Condition 2) and RT (Condition 3), the frequency of MR1/5-OP-RU^{+} T cells was similar to the frequency of Vα7.2^{+}CD161^{hi} T cells (Condition 7; Fig. 3.2C-E). In contrast, MR1/5-OP-RU^{+} T cell frequency was considerably reduced when staining was performed at 37°C (Condition 4), most likely due to internalisation of the TCR.

In Condition 5, cells were stained with the MR1 tetramer and Vα7.2. While the MFI of Vα7.2 was decreased in Condition 5 relative to Condition 7, Vα7.2^{+}CD161^{hi} T cells remained clearly identifiable (Fig. 3.2F). As expected, the majority of Vα7.2^{+}CD161^{hi} T cells were MR1/5-OP-RU^{+} and vice versa (Fig. 3.2G). A minor fraction of MR1/5-OP-RU^{+} T cells were negative for CD161 and/or Vα7.2 (Fig. 3.2G). The cells lacking Vα7.2 may represent rare Vα7.2^{-}MR1/5-OP-RU^{+} MR1-restricted T cells as identified in several studies (Gherardin et al., 2018b; 2016; Koay et al., 2019b; Meermeier et al., 2016).

Finally, combined staining with PE and BV421 MR1 tetramers was tested (Condition 6). Two-colour tetramer staining allowed highly specific MAIT cell detection (Fig. 3.2H). The frequency of MR1/5-OP-RU^{PE+}MR1/5-OP-RU^{BV421+} (hereafter MR1/5-OP-RU^{++}) T cells (Fig. 3.2H) was reduced compared with the frequency of MAIT cells detected using a single tetramer (Condition 3; Fig. 3.2C-D) or Vα7.2 and CD161 (Fig. 3.2E). This may reflect reduced non-specific staining when using two MR1 tetramers. In accordance with this hypothesis, in a separate experiment where cells were stained with MR1/5-OP-RU^{PE}, MR1/5-OP-RU^{BV421} and Vα7.2, I observed a lower frequency of Vα7.2^{-} MAIT cells within the MR1/5-OP-RU^{++} population compared with the single MR1/5-OP-RU^{+} population (Fig. 3.2I).

Based on these results, MR1 tetramer staining for all subsequent experiments was performed for 40 minutes at RT prior to surface and/or intracellular antibody staining. When highly specific MAIT cell identification was required, combined staining was performed with MR1/5-OP-RU^{PE} and MR1/5-OP-RU^{BV421}.

Tissue tetramer staining

As my project involved staining of MAIT cells in human tissue, as well as blood, it was important to test my optimised staining protocol in different tissues. Tetramer staining was performed on matched blood and colon biopsy samples from three donors. The staining panel included both MR1/5-OP-RU tetramers (PE and BV421) and Vα7.2, so that identification of MAIT cells using Vα7.2 and CD161, or MR1/5-OP-RU, could be compared in tissue. In the colon samples, MAIT cells did not form a distinct Vα7.2^{+}CD161^{hi} population amongst CD3^{+} T cells (Fig. 3.3A). Moreover, a large fraction of Vα7.2^{+}CD161^{hi} T cells did not stain with either of the two MR1 tetramers in the three donors (38.6-76.3%; Fig. 3.3B). In contrast, MAIT cells were easily and accurately identified by two-colour MR1/5-OP-RU tetramer staining (Fig. 3.3C). The majority of MR1/5-OP-RU^{++} T cells were Vα7.2^{+}CD161^{hi} (data not shown) and background staining with MR1/6-FP was low (Fig. 3.3C).

I next tested the tetramer staining protocol in a liver biopsy sample with matched blood. Similar to gut, MR1/5-OP-RU^{++} T cells could be clearly identified in the liver (Fig. 3.3D) and were CD161^{hi} as expected (Fig. 3.3E). Consistent with previous studies (Dusseaux et al., 2011; Tang et al., 2013), MAIT cell frequency amongst CD3^{+} T cells was increased in the liver compared with the blood (Fig. 3.3D). Due to the low number of liver lymphocytes, I was unable to perform a tetramer staining control with MR1/6-FP. However, I have demonstrated low background staining with MR1/6-FP in the liver in subsequent experiments (see Chapter 5, Fig. 5.1D).

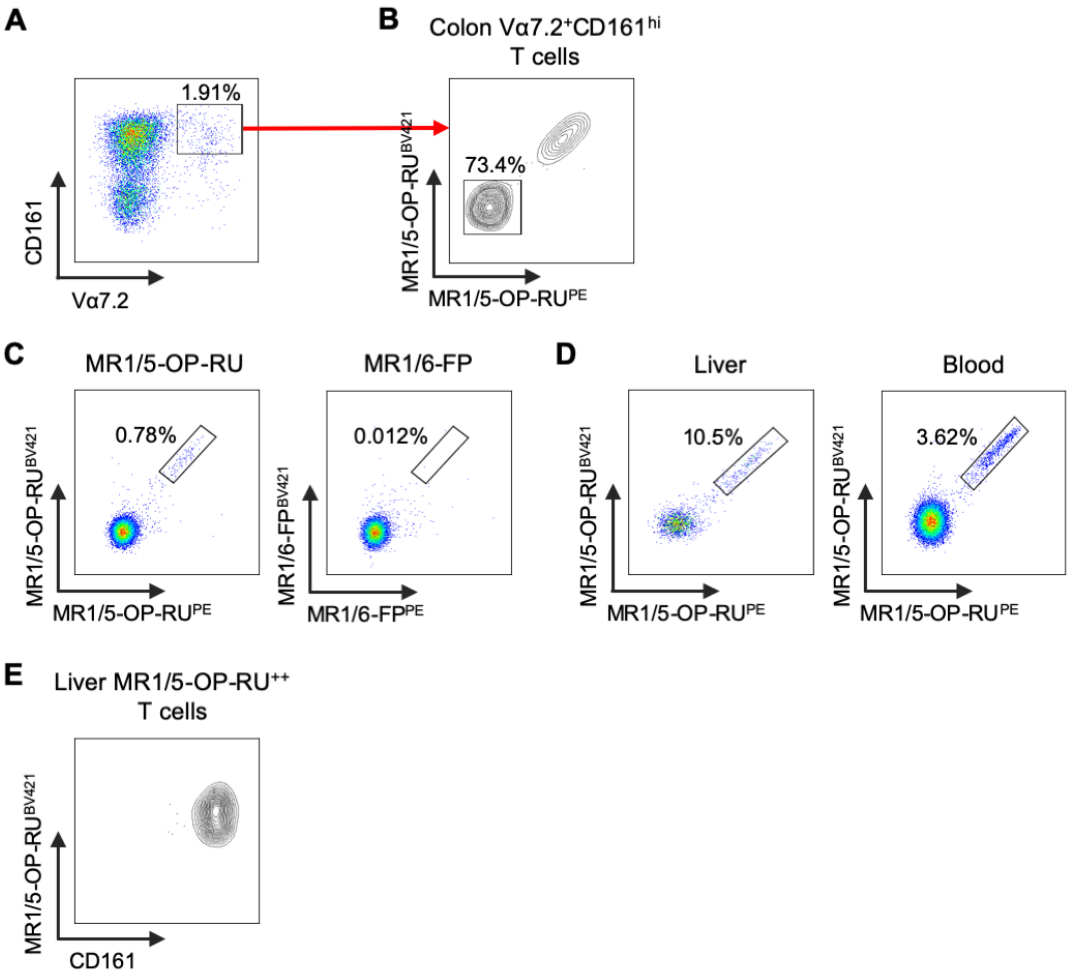


Figure 3.3. MAIT cell detection in tissues using either Vα7.2 and CD161, or the MR1/5-OP-RU tetramer. (A) Vα7.2 and CD161 staining of colonic T cells. **(B)** Representative FACS plot indicating the fraction of MR1/5-OP-RU^{+} MAIT cells amongst colonic Vα7.2^{+}CD161^{hi} T cells. **(C)** FACS plots showing two-colour MR1/5-OP-RU (left) and MR1/6-FP (right) tetramer staining of colonic T cells. **(D)** FACS plots showing two-colour MR1/5-OP-RU tetramer staining of liver (left) and matched blood (right) T cells. **(E)** FACS plot showing CD161 expression on liver MR1/5-OP-RU^{++} T cells. FACS plots in A and C were gated on live CD45^{+}CD3^{+} T cells, and FACS plots in D were gated on CD45^{+}CD14^{-}CD19^{-}TCRγδ^{-}CD3^{+} T cells.

In summary, two-colour MR1/5-OP-RU tetramer staining can clearly identify MAIT cells in different tissues. As a distinct Vα7.2^{+}CD161^{hi} T cell population cannot be detected in some tissues, such as the gut, use of the MR1/5-OP-RU tetramer is essential for accurate MAIT cell identification.

3.2.2. ATAC-seq method

Cycles of PCR amplification

Naïve T cells (CD8^{+}CD4^{-}CCR7^{+}CD45RO^{-}), MAIT cells (CD8^{+}CD4^{-}CCR7^{-}MR1/5-OP-RU^{++}), and T_{mem} (CD8^{+}CD4^{-}CCR7^{-}MR1/5-OP-RU^{-}) cells were sorted from human PBMCs (Fig. 3.4A). Sort purity was > 99.5% for all cell populations (Fig. 3.4B). ATAC-seq was performed on 50,000 cells per sample using the original Buenrostro et al. protocol (Buenrostro et al., 2013; 2015a) with the addition of an extra wash step (500 x g, 5 min, 4°C) after transposition. Addition of this step was recommended by colleagues from Professor Lars Fugger's group to reduce the fraction of mitochondrial reads in ATAC-seq libraries.

ATAC-seq libraries were amplified for an initial five PCR cycles, then 1/10th of each library used for qPCR. According to the Buenrostro protocol, the number of further PCR cycles required is equal to the PCR cycle number at 1/3rd of the maximal qPCR fluorescence intensity. The required number of extra cycles was 12, 8 and 8 for naïve T cells, MAIT cells and T_{mem} cells, respectively (Fig. 3.5A). After amplification for the prescribed number of additional cycles, MAIT and T_{mem} cell libraries showed the expected fragment length distribution on an Agilent Bioanalyzer, but the naïve T cell library that was amplified by 12 extra cycles (shown in red) was skewed towards shorter DNA fragments, indicating overamplification (Fig. 3.5B). Library

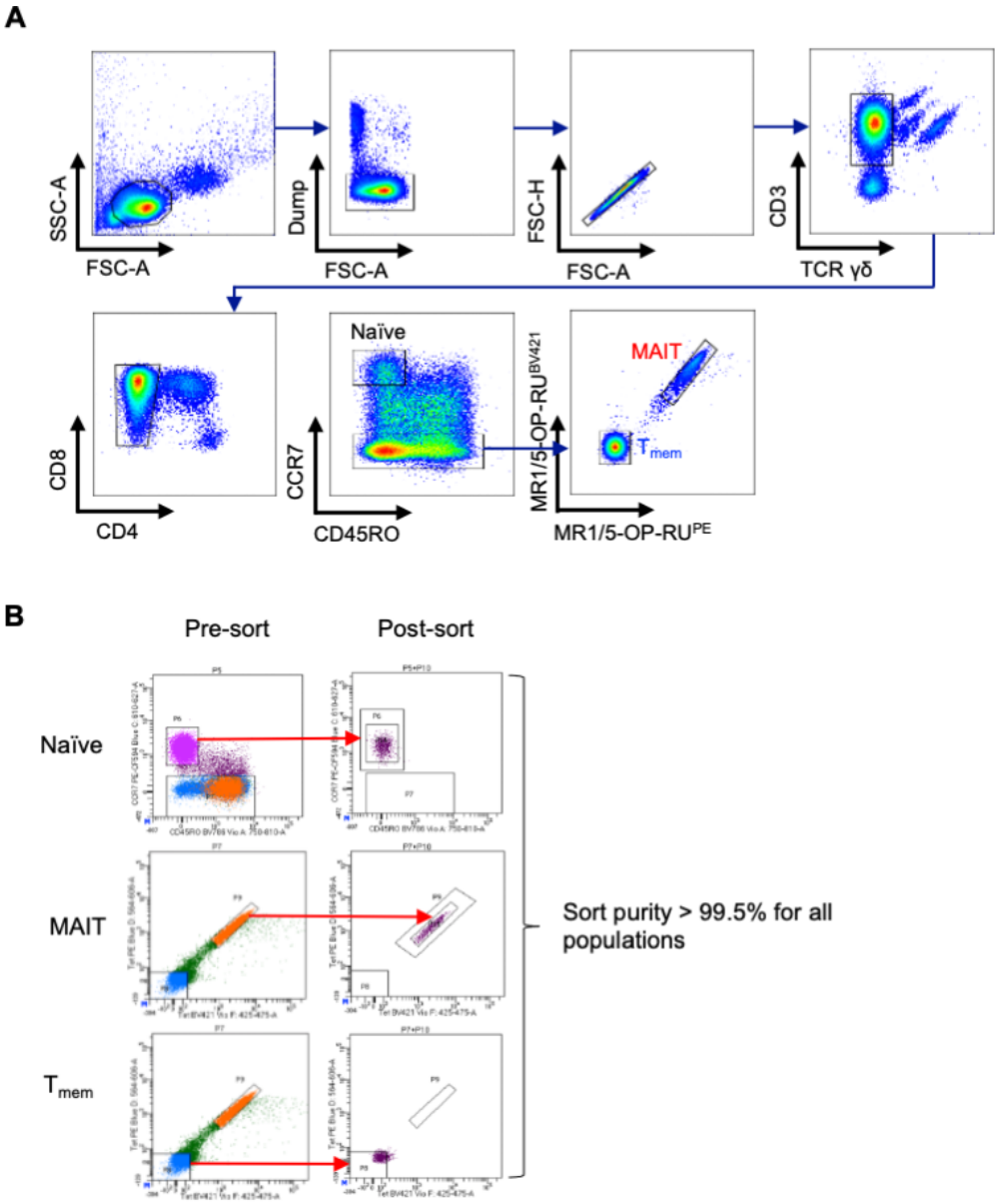
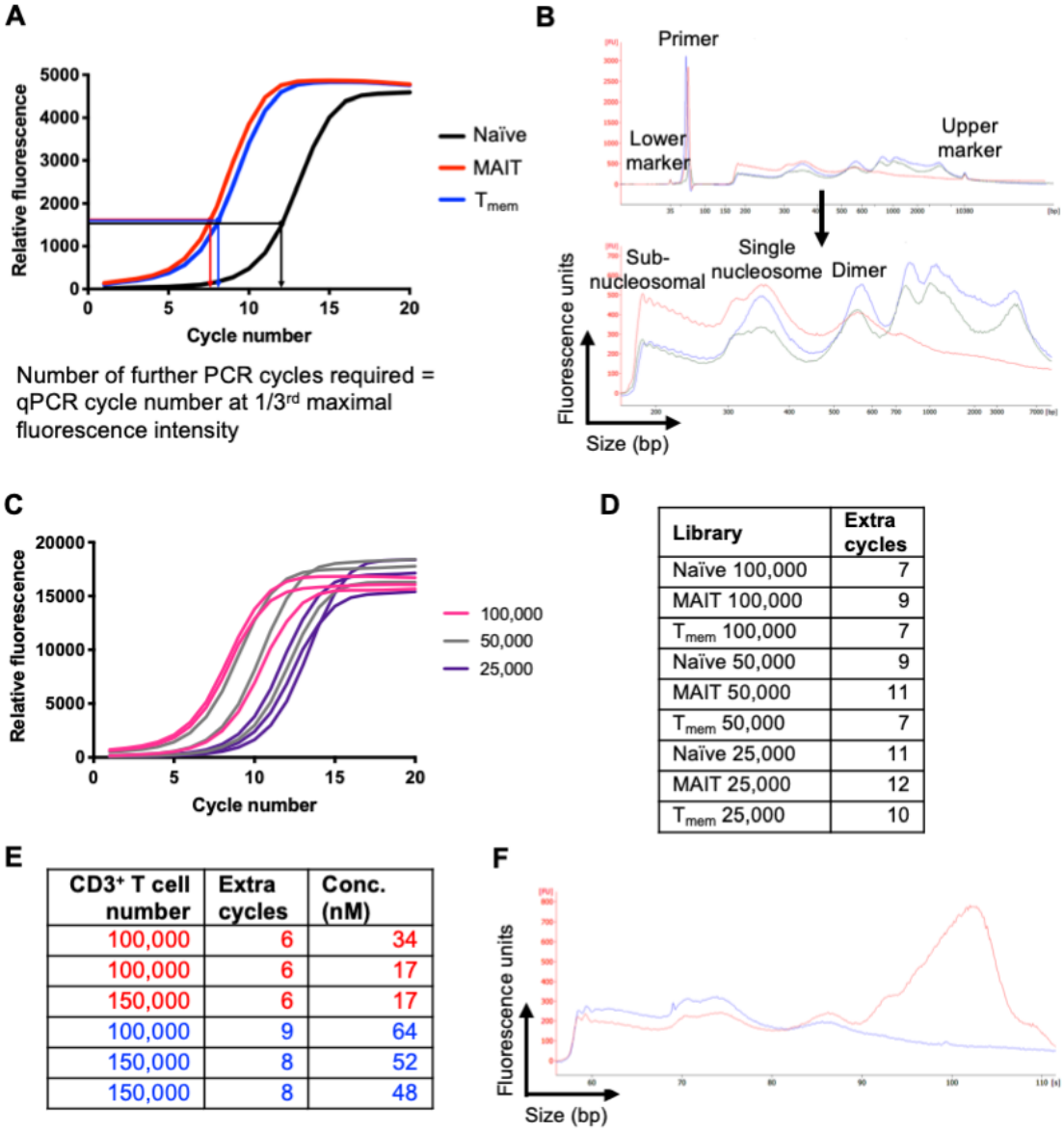


Figure 3.4. Sorting of naïve T cells, MAIT cells and T_{mem} cells for ATAC-seq. (A) Gating strategy for sorting naïve T cells, MAIT cells and T_{mem} cells from human CD8-enriched PBMCs. **(B)** FACS plots showing sort purity for the three T cell populations.

concentrations ranged from 26-92 nM. The expected concentration is around 30 nM (with five-fold variation possible) (Buenrostro et al., 2015a).

Buenrostro and colleagues found that library complexity became an issue when greater than six extra PCR cycles (11 PCR cycles total) were required (Buenrostro et al., 2015a). Therefore, to try to reduce the number of extra PCR cycles required to six or less, I performed ATAC-seq on 100,000 naïve T cells, MAIT cells and T_{mem} cells. In addition, I performed ATAC-seq on equivalent samples of 50,000 and 25,000 cells, so that I could look at the association between input cell number and the number of extra PCR cycles required. Although the number of extra PCR cycles required was on average lower with a higher input cell number, not every sample conformed with this trend (Fig. 3.5C,D). Moreover, the number of required extra cycles was always more than the recommended maximum of six.

I next decided to increase the maximum input cell number to 150,000 in a further attempt to reduce the number of required extra PCR cycles to six or less. For ease, ATAC-seq was performed on total $CD3^+$ T cells, with 150,000 or 100,000 cells per sample (in triplicate). According to qPCR, 8-9 extra PCR cycles were required for all samples. However, I decided to amplify half of the samples by six extra PCR cycles and half by the recommended number of cycles, so that I could compare the fragment distribution and concentration of the amplified libraries. The concentration of the libraries amplified by six extra PCR cycles was 17-34 nM in 20 µl (Fig. 3.5E). This is plenty for sequencing on an Illumina NextSeq, which requires 2 nM of each library in 10 µl. Whilst the libraries amplified by six extra cycles showed the desired fragment length distribution, the $1/3^{rd}$-maximally amplified libraries were biased towards shorter DNA fragments (Fig. 3.5F).



Library	Extra cycles
Naïve 100,000	7
MAIT 100,000	9
T_{mem} 100,000	7
Naïve 50,000	9
MAIT 50,000	11
T_{mem} 50,000	7
Naïve 25,000	11
MAIT 25,000	12
T_{mem} 25,000	10

$CD3^+$ T cell number	Extra cycles	Conc. (nM)
100,000	6	34
100,000	6	17
150,000	6	17
100,000	9	64
150,000	8	52
150,000	8	48

Figure 3.5. Optimising the degree of ATAC-seq library amplification. (A) Buenrostro et al. qPCR-based method to determine the number of further PCR amplification cycles (above the initial 5) required for ATAC-seq libraries (Buenrostro et al., 2015a). Naïve, MAIT and T_{mem} cell libraries needed 12, 8 and 8 extra PCR cycles, respectively. **(B)** Bioanalyzer traces showing the fragment length distribution of naïve (red), MAIT (blue) and T_{mem} (green) cell libraries. **(C,D)** qPCR plot (C) and table (D) showing the number of recommended extra PCR cycles for libraries of 100,000, 50,000 and 25,000 naïve, MAIT and T_{mem} cells. **(E)** Table showing the number of $CD3^+$ T cells in six ATAC-seq libraries, the number of extra PCR cycles by which the libraries were amplified, and the final concentration of the libraries. **(F)** Bioanalyzer traces demonstrating two representative ATAC-seq libraries from (E), amplified by six (red) or 1/3rd maximal (blue) extra PCR cycles.

In conclusion, libraries from up to 150,000 T cells required more than the recommended six extra PCR cycles. However, amplification of libraries from 100,000 or 150,000 cells by six extra PCR cycles, regardless of the 1/3rd maximal number of PCR cycles (up to a maximum of 9), produced libraries of the desired concentration and fragment distribution. Therefore, for future experiments, I decided to forego the 1/3rd maximal amplification calculation and generate ATAC-seq libraries from 100,000 cells per sample, all amplified by six extra PCR cycles.

Generating ATAC-seq libraries

ATAC-seq libraries were generated from 100,000 naïve T cells, MAIT cells and T_{mem} cells from three donors in triplicate (27 libraries total), with all libraries amplified by 11 PCR cycles (six extra cycles). To limit loss of DNA, I removed the additional wash step after transposition (previously added as a recommendation by the Fugger group).

Generated libraries showed the expected fragment length distribution (Fig. 3.6A). However, as previously observed, all libraries contained a large primer peak at around 50 bp that comprised 20-40% of the library (Fig. 3.6A). To remove the primer peak, size selection was performed using SPRI beads. The primer peak was successfully removed from all samples (Fig. 3.6B), with final libraries containing ≤ 1% primer DNA. The concentration of the libraries was suitably high for sequencing, in the range 70-200 nM.

From the 27 libraries, I chose 12 for sequencing. These included naïve, MAIT and T_{mem} cell libraries from three donors (biological replicates), in addition to technical replicate libraries for each cell type from one donor. Inclusion of technical replicates allowed me to examine the impact of technical variability on ATAC-seq data.

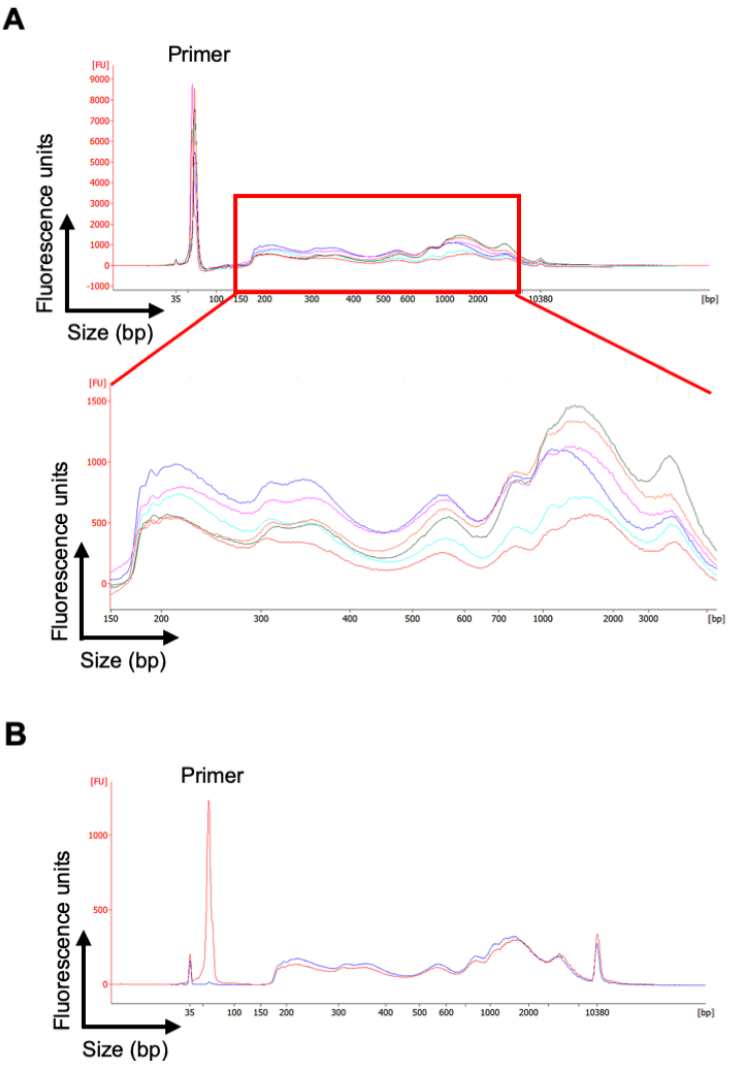


Figure 3.6. Removal of the primer peak from ATAC-seq libraries. (A) Bioanalyzer traces showing six representative ATAC-seq libraries from 100,000 naïve, MAIT and T_{mem} cells. **(B)** Bioanalyzer traces showing a representative ATAC-seq library before (red) and after (blue) primer removal.

Sequenced ATAC-seq libraries were of high quality but contained a large fraction of reads mapping to mtDNA

The number of concordantly mapped paired-end reads per library (prior to filtering) was in the range 39-57 million. The overall alignment rate (percentage of reads mapped concordantly ≥ one time) to hg38 was very high at > 98%, however 40-70% of reads mapped to mtDNA. Following mapping, duplicate reads were discarded along with reads mapping to mtDNA and to blacklisted regions (Amemiya et al., 2019; Consortium, 2012). Reads were filtered to retain only those with high mapping quality scores (MAPQ ≥ 30). This left 10-20 million paired-end reads per library, considerably below the ENCODE standard of 25 million (https://www.encodeproject.org/atac-seq/).

The libraries could have been resequenced until they reached the required 25 million paired-end reads. However, I aimed to obtain a larger number of paired-end reads per library, namely around 75 million, in order to perform TF footprinting analysis. This would have required a large number of sequencing lanes and thus would have been very costly. Each time the libraries are sequenced, 40-70% of the reads would have to be discarded due to their mitochondrial origin. Furthermore, the number of reads gained per lane of sequencing would decrease, as an increasing proportion of reads would be duplicates of previously sequenced reads and would be discarded during data analysis.

Despite the high fraction of reads mapping to mtDNA, my ATAC-seq libraries were of high quality. The fragment length distributions showed the expected nucleosomal periodicity (Fig. 3.7A). ATAC-seq peak intensity was highly correlated between biological replicates (Fig. 3.7B), with an even higher correlation between technical

replicates (Fig. 3.7C). A combined total of 61,064 regions of chromatin accessibility were identified from the three cell types.

In summary, ATAC-seq libraries generated using the original ATAC-seq protocol were of high quality but contained a sizeable fraction of reads mapping to mtDNA, limiting the number of useable reads. Technical replicates were very highly correlated and therefore it was not necessary to include these in further experiments.

Generating Fast-ATAC libraries

With the aim of reducing the proportion of reads mapping to mtDNA, I switched to an optimised ATAC-seq approach known as Fast-ATAC (Corces et al., 2016). Fast-ATAC is very similar to the original ATAC-seq method, however cell lysis and transposition are performed in a single step rather than in separate steps, and a different detergent (Digitonin) is used for cell lysis. Libraries generated using Fast-ATAC have approximately five-fold fewer reads mapping to mtDNA compared with libraries generated using the original ATAC-seq protocol, in addition to reduced noise and a five-fold increased fragment yield per cell (Corces et al., 2016).

Fast-ATAC libraries were generated in duplicate from 50,000 (standard Fast-ATAC input cell number) sorted $CD8^+$ naïve T cells, MAIT cells and T_{mem} cells from the CD8-enriched PBMCs of three donors. At the same time, cells were sorted for matched RNA-seq, the results of which are discussed in Chapter 4. qPCR was performed as before to calculate the recommended number of extra PCR cycles, which was eight or nine for all libraries. Given my previous success with libraries that showed similar qPCR results, I decided to amplify the libraries by just six extra PCR cycles. All libraries showed the expected fragment length distribution by

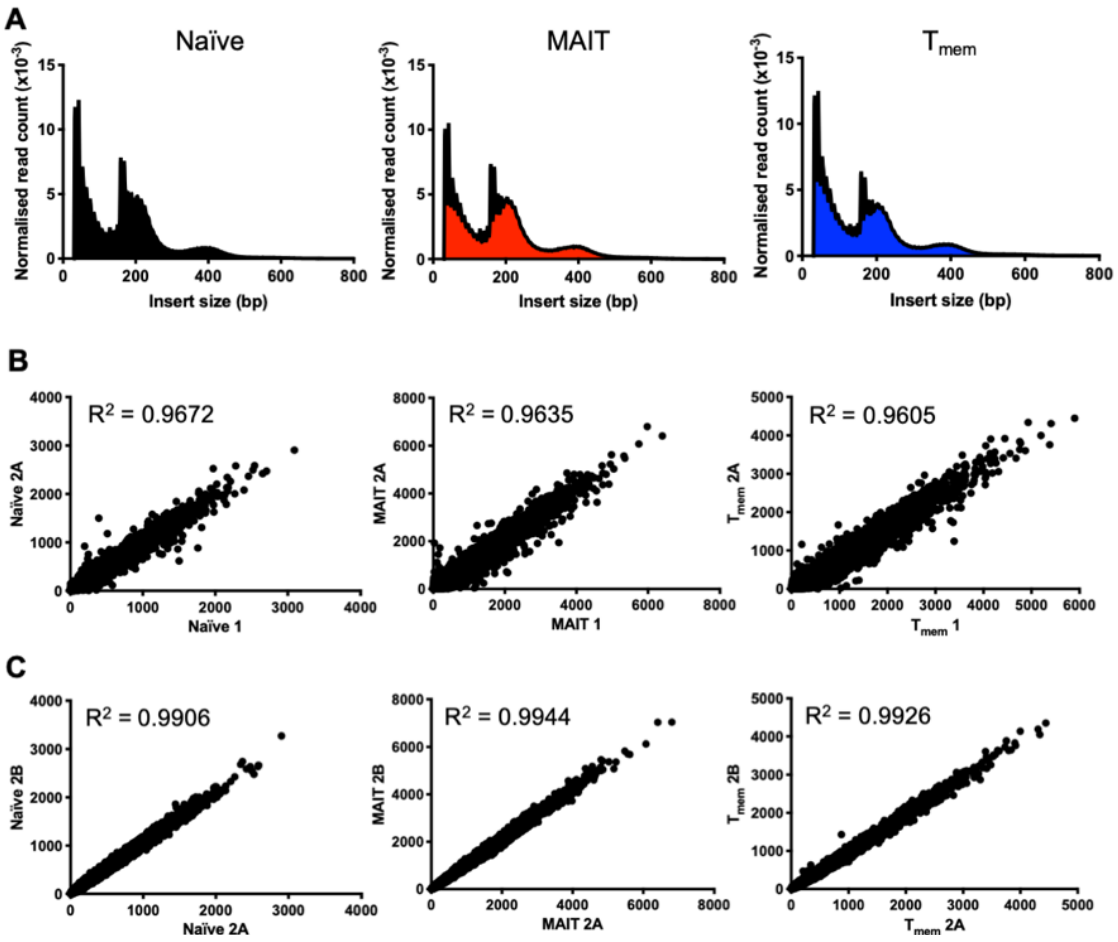


Figure 3.7. ATAC-seq libraries are of high quality. (A) Fragment length distributions of ATAC-seq libraries from naïve, MAIT and T_{mem} cells from a representative donor. Read counts normalised to library size. **(B,C)** Correlation in normalised peak counts (number of transposase cut sites per peak) between biological replicate (B) and technical replicate (C) ATAC-seq libraries. Pearson R squared value is shown.

Bioanalyzer (Fig. 3.8A). Nine libraries (one per cell type per donor) were size selected and quantified. Concentrations were in the range 18-42 nM, ample for sequencing.

Sequenced Fast-ATAC libraries were of high quality and contained a significantly reduced fraction of reads mapping to mtDNA

After one lane of sequencing, I assessed Fast-ATAC library quality using several measures, including the fraction of reads mapping to mtDNA and the appearance of the fragment length distributions. As the quality appeared high, I ran a further three lanes of sequencing in order to increase read depth.

The quality of the final Fast-ATAC libraries (pooled from four lanes) was assessed according to ENCODE data quality standards. High quality ATAC-seq libraries should: (1) have a minimum of 25 million non-duplicate non-mitochondrial paired-end reads, (2) contain a mononucleosomal peak in the fragment length distribution, (3) have an alignment rate > 95%, (4) have IDR (Li et al., 2011) self-consistency and rescue ratios < 2 and IDR peak count > 50,000, and (5) have a fraction of reads in peak regions (FRiP) score > 0.3.

I obtained 203-261 million concordantly mapped paired-end reads per library (prior to filtering) with an alignment rate to hg38 of > 95%. Compared with the original ATAC-seq protocol, Fast-ATAC libraries showed a significantly reduced fraction of reads mapping to mtDNA, in the range 6-25% (Fig. 3.8B). After removing duplicate reads, mtDNA and blacklisted regions (Amemiya et al., 2019; Consortium, 2012), and retaining only high quality read pairs (MAPQ ≥ 30), I was left with 59-83 million paired-end reads per sample. Libraries showed the expected fragment length

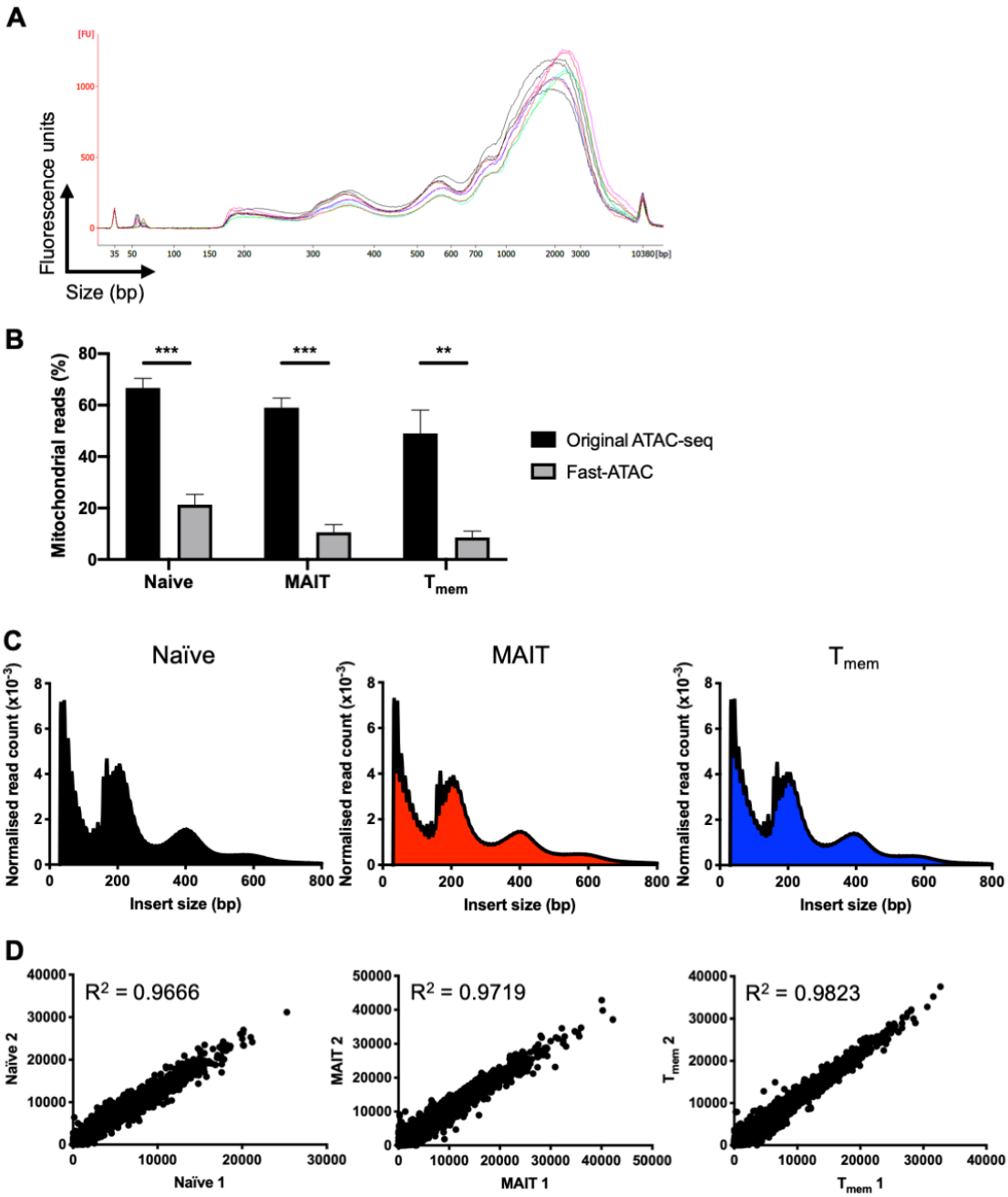


Figure 3.8. Fast-ATAC libraries are of high quality and have a reduced mitochondrial read fraction compared with standard ATAC-seq libraries. (A) Bioanalyzer traces of representative Fast-ATAC libraries. **(B)** Comparison of the percentage of mitochondrial reads in ATAC-seq libraries generated using the original or Fast-ATAC approach (n = 3 per group). Mean ± SD is shown. Unpaired t-test with Holm-Šídák correction for multiple comparisons. ** $p < 0.01$, *** $p < 0.001$. **(C)** Fragment length distribution of naïve, MAIT and T_{mem} cell libraries from a representative donor. Read count normalised by library size. **(D)** Correlation in normalised peak counts (number of transposase cut sites per peak) between biological replicate Fast-ATAC libraries. Pearson R squared value is shown.

distribution (Fig. 3.8C) and peak intensity was highly correlated between biological replicates (Fig. 3.8D).

IDR analysis was performed to check consistency between the replicates in my experiment and to produce an optimal peak list for naïve T cells, MAIT cells and T_{mem} cells. The IDR self-consistency ratio and rescue ratio (IDR QC scores) were < 2 for each of the three cell types, indicating high consistency between replicates. The number of optimal IDR peaks was 54,645 for naïve T cells, 52,138 for MAIT cells and 54,645 for T_{mem} cells, resulting in a total of 74,927 peaks in the global peak list (merged from the three separate peak lists). All samples had a high FRiP score (0.57-0.66).

Overall, Fast-ATAC libraries were of high quality and depth, and contained a relatively small proportion of reads mapping to mtDNA. The biological findings from this Fast-ATAC experiment (along with the paired RNA-seq data) will be discussed in Chapter 4.

3.2.3. Removal of multiplets from paired scRNA-seq and scTCR-seq data

Developing a new method for multiplet identification and removal

To identify potential doublets (or other multiplets) in my 10x data, I took advantage of the fact that I had gene expression and TCR data from the same individual cells. Most T cells express a TCR comprising an α chain paired with a β chain (named αβ T cells). Due to the simultaneous recombination of TCRα gene segments (V and J) on both human chromosomes, single T cells can express mRNA for two TCRα

chains. Single-cell sequencing studies estimate that 10-19% of T cells express two productive TCRα chains (Dash et al., 2011; Eltahla et al., 2016; Han et al., 2014; Stubbington et al., 2016). A smaller fraction of T cells, around 6-7% (Eltahla et al., 2016; Stubbington et al., 2016), express two TCRβ chains due to "leakiness" in allelic exclusion at the TCRβ locus. While T cells can express two TCRα and/or two TCRβ chains, it is impossible for a single diploid human cell to express three different TCRα or TCRβ chains. Therefore, cells with three TCRα and/or three TCRβ chains must be multiplets and would be expected to show a higher average UMI count compared with cells with fewer TCR chains.

Comparing the UMI distributions of cells with zero, one, two or three TCRα chains within an individual sample (prior to any cell filtering), I observed a positive association between the number of TCRα chains detected per cell and the average UMI count, irrespective of the tissue of origin of the T cells (Fig. 3.9A,B). Compared with cells expressing a single TCRα chain (which were presumed to be predominantly live single cells), the UMI count was significantly decreased and increased, respectively, in cells with zero or three TCRα chains (Fig. 3.9A,B). Cells with zero TCRα chains were enriched for dead cells, as demonstrated by their increased mitochondrial read fraction (Fig. 3.9C,D). This explains their low average UMI count. Comparable results were obtained for cells grouped by TCRβ chain frequency (data not shown).

Using these data, the upper UMI threshold was derived empirically for each sample. The optimal upper UMI threshold was defined as the threshold that removed the majority of cells with greater than two TCRα and/or two TCRβ chains, while retaining most cells with a single TCRα and/or TCRβ chain. These criteria were met when

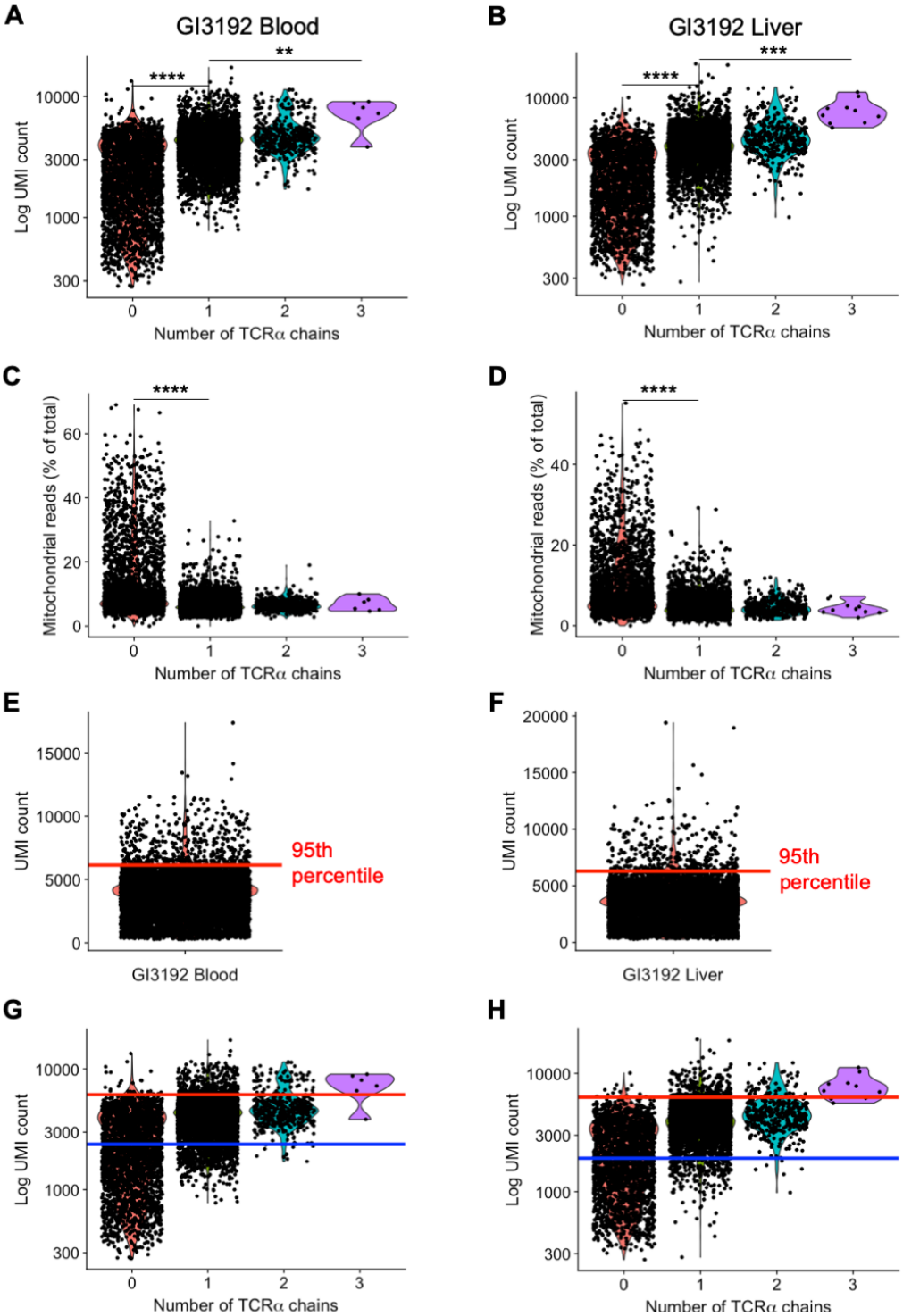


Figure 3.9. Novel method for the exclusion of multiplets from paired scRNA-seq and scTCR-seq data. Violin plots showing blood (left) and liver (right) samples from a representative donor (GI3192). Dots indicate individual cells. **(A-D)** Log UMI counts (A,B) and mitochondrial read percentage (C,D) for cells with zero, one, two or three TCRα chains. **(E,F)** UMI counts for all cells with one TCRα and/or one TCRβ chain. The upper UMI threshold is the 95th percentile of this distribution, as indicated by the red line. **(G,H)** Violin plots from (A,B) showing the upper (red) and lower (blue) UMI threshold. Wilcoxon Rank Sum test (A,B). p values in (A) adjusted for multiple comparisons (Bonferroni correction). **** $p < 0.0001$, *** $p < 0.001$, ** $p < 0.01$.

the UMI threshold was set at the 95th percentile of the UMI distribution for all cells with one TCRα chain and/or one TCRβ chain (Fig. 3.9E-H).

Validation of novel multiplet identification method using a separate dataset from inflamed colonic T cells

The above approach was validated on a separate scRNA-seq and scTCR-seq dataset (generated using 10x Genomics technology), which comprised T cells from the inflamed mucosa of four ulcerative colitis patients. Again, UMI count was positively associated with the number of expressed TCRα (Fig. 3.10A) and TCRβ (Fig. 3.10B) chains, and mitochondrial read fraction was increased in the population of cells lacking expression of either TCR chain (Fig. 3.10C,D). Based on my definition of the optimal upper UMI threshold, the upper UMI threshold for this dataset was set at the 85th percentile of the UMI distribution for all cells with one TCRα chain and/or one TCRβ chain (Fig. 3.10E-G).

In summary, the upper UMI thresholds determined using my novel approach removed most unequivocal multiplets (i.e. cells with more than two TCRα and/or TCRβ chains), while retaining the large majority of presumed live single cells (i.e. cells with a single TCRα and/or TCRβ chain) from the two datasets. As demonstrated using a second dataset, the optimal UMI threshold must be derived empirically in each case.

3.3. Discussion

In conclusion, I have optimised an MR1 tetramer staining protocol for accurately identifying MAIT cells in blood and tissue, and optimised the ATAC-seq protocol for

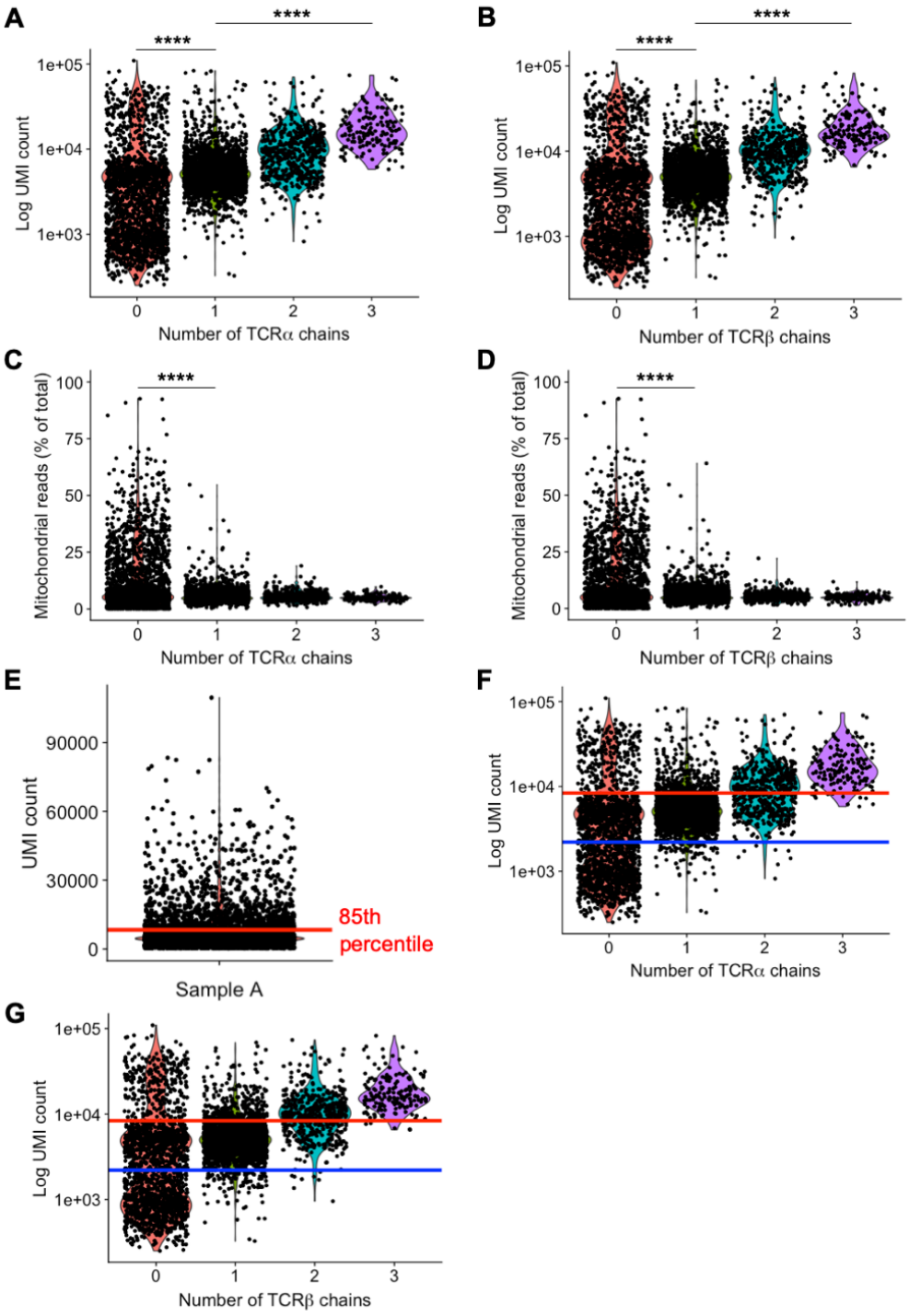


Figure 3.10. Validation of the novel approach to multiplet identification using a combined scRNA-seq and scTCR-seq dataset of colonic T cells from four patients with active ulcerative colitis. (A-D) Violin plots for a representative sample ("Sample A") showing the log UMI counts (A,B) and percent mitochondrial reads (C,D) for cells with zero, one, two or three TCRα (A,C) or TCRβ (B,D) chains. **(E)** Violin plot showing the UMI counts for cells with one TCRα and/or one TCRβ chain. The upper UMI threshold is set at the 85th percentile, as indicated by the red line. **(F,G)** Violin plots from (A) and (B) showing the upper (red) and lower (blue) UMI threshold. The upper UMI threshold removes the majority of cells with more than two TCRα (F) or TCRβ (G) chains, while retaining most cells with a single TCRα (F) or TCRβ (G) chain. Wilcoxon Rank Sum test. p values in (A,B) adjusted for multiple comparisons (Bonferroni correction). **** $p < 0.0001$.

an input of 50,000 T cells. In addition, I have devised a novel method for the exclusion of multiplets from droplet-based paired scRNA-seq and scTCR-seq data that makes use of the number of TCRα and TCRβ chains detected per cell.

The MR1/5-OP-RU staining protocol that I have developed is suitable for accurately identifying MAIT cells in human blood, gut and liver. While it may be necessary to further optimise the protocol for other tissues, the fact that it worked well for gut biopsies, which contain a wide variety of different cell types (e.g. lymphocytes, epithelial cells, stromal cells) and a relatively high frequency of dead cells, suggests that the protocol is likely to be appropriate for other tissues.

My data supports the use of MR1/5-OP-RU for the specific identification of MAIT cells in tissues. In the colon, a large fraction (38.6-76.3%) of Vα7.2^{+}CD161^{hi} T cells did not stain with MR1/5-OP-RU (Fig. 3.3B), indicating that not all Vα7.2^{+}CD161^{hi} T cells in the gut are bona fide MAIT cells. One limitation of this experiment was that both MR1/5-OP-RU and Vα7.2 were included in the colon staining panel. In my optimisation experiments on PBMCs, Vα7.2 MFI was reduced in the presence of MR1/5-OP-RU (Fig. 3.2F). Therefore, inclusion of both MR1/5-OP-RU and Vα7.2 in the colon staining panel might have made it more difficult to identify Vα7.2^{+}CD161^{hi} T cells. However, it was necessary to include MR1/5-OP-RU in order to determine the fraction of Vα7.2^{+}CD161^{hi} T cells that were not MR1/5-OP-RU^{+} MAIT cells. Numerous studies of MAIT cells in human intestine in both health and disease have identified MAIT cells using surrogate markers, most commonly Vα7.2 combined with CD161 (Cosgrove et al., 2013; Haga et al., 2016; Hama et al., 2019; Hiejima et al., 2015; Ling et al., 2016; Serriari et al., 2014; Slichter et al., 2016; Sundström et al., 2015; Tominaga et al., 2017; Won et al., 2016; Zabijak et al., 2015). Despite its limitations, my data suggests that the findings from these studies need to be

validated using the MR1 tetramer. Moreover, the accuracy of surrogate methods of MAIT cell identification should be examined in other adult tissues.

Since generating my Fast-ATAC libraries, an updated ATAC-seq protocol, named Omni-ATAC, has been published by the labs of William Greenleaf and Howard Chang (Corces et al., 2017). Omni-ATAC is suitable for use with a wide variety of sample types, including diverse cell lines and tissues, and frozen tissue sections. The quality of Omni-ATAC data is improved relative to data generated using Fast-ATAC or the original ATAC-seq protocol. However, for isolated $CD4^+$ T cells, Omni-ATAC shows only minor improvements over Fast-ATAC in several measures of library quality, including library size and the percentage of reads mapping to mtDNA. As the quality of my Fast-ATAC data is high and I sequenced to a high read depth, the information content of my data is unlikely to be meaningfully reduced relative to what could be achieved with Omni-ATAC on the same samples. The biggest advantages of using the Omni-ATAC protocol appear to be in the context of challenging samples, such as frozen tissues.

My novel approach to the removal of multiplets from paired scRNA-seq and scTCR-seq data successfully removed the majority of multiplets with three TCRα and/or three TCRβ chains from the two datasets examined (Fig. 3.9G,H and 3.10F,G). However, multiplets were likely also present amongst cells with zero, one or two TCRα or TCRβ chains. Given the lack of data on the true identity of multiplets in these samples, I was unable to assess the overall accuracy of multiplet removal.

The alternative approaches for doublet identification discussed in the introduction (Gehring et al., 2018; Guo et al., 2019; Kang et al., 2018; McGinnis et al., 2019a; 2019b; Shin et al., 2019; Stoeckius et al., 2018; Wainer-Katsir and Linial, 2019;

Wolock et al., 2019) are very new and were largely unavailable at the time of data collection. Besides, some of these methods would be unsuitable for this study. For example, I lacked genomic information from the donors and the cell population under study was too homogenous for computational techniques that detect heterotypic doublets. "Cell Hashing", a multiplexing approach that uses barcoded antibodies against ubiquitous surface antigens (Stoeckius et al., 2018), is now commercially available and would be relatively simple to incorporate into future assays if cell number is not limiting. Overall, while not optimal, the approach that I have developed for the identification of doublets appears to be the most appropriate for my dataset, and is superior to the filtering of data at an arbitrary UMI or gene count threshold, as is commonly used (Luecken and Theis, 2019). Furthermore, as my approach is purely computational, it can easily be applied to previously generated scRNA-seq and scTCR-seq datasets.

3.4. Summary

- Optimised the MR1 tetramer staining protocol in blood and tissue.
- Optimised the ATAC-seq protocol for 50,000 T cells.
- Developed a novel approach for the removal of multiplets from scRNA-seq data that makes use of matched scTCR-seq data.

4. Comparison of the transcriptional profile and chromatin accessibility landscape of human blood CD8$^+$ MAIT cells and T_{mem} cells

MAIT cells have a characteristic phenotype and function that is reflected in a unique gene expression profile amongst CD8$^+$ T cells in humans. Compared with CD161$^+$ and CD161$^-$ CD8$^+$ T cells, CD161^{++} (or CD161^{hi}) CD8$^+$ T cells, comprising largely MAIT cells, show altered expression of chemokine and cytokine receptors, TFs, and cytokines (Billerbeck et al., 2010). In particular, MAIT cells show high expression of a number of type 17-associated genes, including *IL23R*, *IL18R1*, *RORC*, *CCR2* and *CCR6*. Along with the type 17 TF RORγt, MAIT cells also express the type 1 TF T-bet, Helios, and PLZF, which is thought to regulate their innate-like functionality, analogous with its function in iNKT cells (Koay et al., 2016; Kovalovsky et al., 2008; Leeansyah et al., 2015; Savage et al., 2008). To gain a deeper insight into MAIT cell biology, recent studies have employed genome-wide approaches to characterise the resting transcriptional profile of blood and tissue MAIT cells (Salou et al., 2018) and to examine the genes induced following MAIT cell activation with different stimuli (Hinks et al., 2019; Lamichhane et al., 2019b; Leng et al., 2019).

While MAIT cells exhibit defined transcriptional differences to T_{mem} cells, whether these differences in gene expression are associated with distinct epigenetic landscapes has not been addressed. One epigenetic property that can be measured is chromatin accessibility. This is easily achieved using modern techniques, such as ATAC-seq, as described in Chapter 3. Open chromatin is

enriched in regulatory regions, such as promoters and enhancers (Klemm et al., 2019), and genome-wide chromatin accessibility data can provide important insights into cell biology, particularly relating to the regulation of gene expression. Additionally, compared with transcriptional data, chromatin accessibility data is better able to distinguish different cell types or cell states (Sen et al., 2016), due to the increased frequency of open regions compared with expressed genes.

Accessibility of chromatin at a given promoter or enhancer is necessary but not sufficient for expression of the associated gene. Gene expression requires other factors, including the binding of appropriate TFs (Lambert et al., 2018a). Therefore, some genes will exhibit altered expression in the absence of changes in chromatin accessibility, while in other cases changes in expression will be accompanied by altered accessibility. Genes that are accessible, but not expressed, under given conditions may provide an insight into the functional potential of the cell type under study. Chromatin accessibility and gene expression analysis therefore provide complementary information and are most useful when performed in parallel.

Whole-genome chromatin accessibility profiles can potentially be used to predict which TFs are important for regulation of a given cell state. The binding preferences of a large number of human TFs are defined as position frequency matrices (PFM) or PWMs that are typically generated from ChIP-seq experiments. PFMs/PWMs are stored in a number of publicly available databases, such as HOCOMOCO (Kulakovskiy et al., 2018) and JASPAR (Fornes et al., 2019). Consequently, it is possible to look for the enrichment of particular TF motifs either within the open chromatin regions of one cell type relative to another, or in open chromatin regions relative to the background genome.

Using a more complex technique, known as TF footprinting (Sung et al., 2016), which was originally developed for DNase-seq data, it may be possible to deduce sites at which a given TF is bound to DNA in your cell type or condition of interest. This method assumes that TF binding will block accessibility of the Tn5 transposase to the underlying DNA and therefore cause a local decrease in accessibility at the TF binding motif. TF footprinting has been used in a number of ATAC-seq publications (Buenrostro et al., 2013; Corces et al., 2016; 2018; Hosoya et al., 2018; Moskowitz et al., 2017; Pauken et al., 2016; Ramirez et al., 2017; Sen et al., 2016; Varshney et al., 2017). However, recent studies have questioned the overall utility of TF footprinting (for both DNase-seq and ATAC-seq) due to a number of limitations, the most significant of which is the lack of detectable footprints for a large fraction of TFs (Goldstein and Hager, 2018; Sung et al., 2016). Furthermore, some studies suggest that ATAC-seq data is inferior to DNase-seq data for the identification of TF footprints (Karabacak Calviello et al., 2019; Schwessinger et al., 2017), but this appears to depend on the computational approach used for footprint identification (Li et al., 2019; Quach and Furey, 2016). This area of research is relatively new and is rapidly progressing. Thus, further studies are required to test the suitability of ATAC-seq data for TF footprinting, as well as to determine the most effective computational approaches for identifying footprints.

Using information from both RNA-seq and ATAC-seq can enable more accurate predictions of the key TF regulators in a given cell type or experimental condition. As TFs from the same family often have a similar PWM (Todeschini et al., 2014), enrichment of a particular motif in accessible chromatin generally suggests a role for one or several members of a TF family. Potential targets can then be narrowed down to the TFs that are actually expressed in the setting of interest. More

sophisticated methods that use both gene expression and chromatin accessibility data to identify significant TFs, have also been developed. For example, IMAGE (Madsen et al., 2018) uses a machine-learning approach that takes into account both the change in accessibility of TF motifs in the local region around a gene, as well as the expression of that gene, to predict the TFs regulating global gene expression changes along differentiation pathways, following cell activation, or between different cell types.

In summary, MAIT cells have been characterised at a transcriptional level, but remain to be explored at the epigenomic level. Combined analysis of transcriptional and chromatin accessibility data can provide information on both the genes expressed under a given condition (from RNA-seq data) and the genes with the potential to be expressed under alternative conditions (from ATAC-seq data). While the role of several MAIT cell TFs is known or presumed, much remains to be explored regarding the regulation of MAIT cell phenotype and function by TFs. This can be investigated by ATAC-seq through TF motif enrichment and perhaps footprinting.

4.1. Aim

To characterise the relationship between the transcriptional and chromatin accessibility profile of MAIT cells, and to define key TFs important for their regulation.

4.2. Key questions

1. Do $CD8^+$ MAIT cells have a unique chromatin accessibility landscape relative to conventional $CD8^+$ T_{mem} cells?
2. To what extent are changes in gene expression between MAIT cells and T_{mem} cells associated with changes in chromatin accessibility?
3. What is the role of known and novel TFs in regulating MAIT cell biology?

4.3. Experimental design

Matched RNA-seq and ATAC-seq on human peripheral blood $CD8^+$ naïve, MAIT and T_{mem} cells from three healthy donors.

Justifications for design:

- MAIT cells and naïve T cells are highly distinct in phenotype and function, and therefore naïve T cells provided a control T cell population that was expected to show large differences in expression and chromatin accessibility compared with MAIT cells.
- MAIT cells have an effector memory phenotype, and show phenotypic and functional similarities with T_{mem} cells (Dusseaux et al., 2011). Therefore comparison of these two cell types enables the unique features of MAIT cells to be defined.
- MAIT cells are predominantly $CD8^+$ (70-90%), but can be DN (10-20%), with a minor population of $CD4^+$ cells (Gherardin et al., 2018b; Martin et al., 2009; Reantragoon et al., 2013). DN and $CD4^+$ MAIT cells show some differences relative to $CD8^+$ MAIT cells (Dias et al., 2018; Gherardin et al., 2018b;

Kurioka et al., 2017). Therefore, to ensure that the population under study was as homogeneous as possible, I focused on the predominant $CD8^+$ subset of MAIT cells. $CD8^+$ MAIT cells also have a natural comparator population, namely $CD8^+$ T_{mem} cells, while the DN population of conventional T cells is highly heterogeneous and relatively poorly characterised (Brandt and Hedrich, 2018).

4.4. Results

$CD8^+$ naïve T cells, MAIT cells and T_{mem} cells (50,000 cells/population) were sorted for matched RNA-seq and ATAC-seq as described in Chapter 3 (sorting strategy in Fig. 3.4A). The ATAC-seq data generated using the optimised Fast-ATAC protocol was of high quality (Fig. 3.8) and therefore suitable for addressing the key questions of this study.

4.4.1. MAIT cells are transcriptionally distinct from T_{mem} and naïve T cells

PCA of the RNA-seq data showed that MAIT cells were transcriptionally distinct from T_{mem} cells and naïve T cells (Fig. 4.1A). Biological replicates were tightly clustered on the PCA plot, indicating that cells of the same type were transcriptionally similar across individuals and that the RNA-seq data was of high quality. MAIT cells and T_{mem} cells were more similar to one another than either was to naïve T cells, as PC1 explained ~61% of the variance in the data, whilst PC2 explained ~22%. This was also apparent in the heatmap of genes differentially expressed ($log_2FC > 1$, FDR < 0.05) between any pair of the three cell types (Fig.

4.1B). While there were clusters of genes that showed similar expression levels in MAIT and T_{mem} cells, naïve T cells were largely distinct, as supported by hierarchical clustering of the samples (Fig. 4.1B).

Compared with naïve T cells, MAIT cells and T_{mem} cells showed shared upregulated expression of 1613 protein-coding genes. These encoded effector cytokines (e.g. *IFNG*, *TNF*), cytotoxic molecules (e.g. *GZMB*, *GNLY*), tissue-homing chemokine receptors (e.g. *CXCR6*, *CCR5*), cytokine receptors (*IL12RB2*, *IL18R1*), as well as TFs that regulate the effector function and memory phenotype of $CD8^+$ T cells (e.g. *TBX21*, *EOMES*) (Intlekofer et al., 2005; Pearce et al., 2003; Sullivan et al., 2003). Furthermore, upregulated genes were enriched for gene sets with increased expression in activated vs. resting $CD4^+$ T cells (Fig. 4.1C) and in $CD4^+$ effector memory T (T_{EM}) cells vs. $CD4^+$ central memory T (T_{CM}) cells (Fig. 4.1D), as demonstrated by GSEA (Godec et al., 2016; Mootha et al., 2003; Subramanian et al., 2005). This is consistent with MAIT cells and T_{mem} cells sharing effector memory-related characteristics.

Naïve T cells showed upregulated expression of 1431 protein-coding genes compared with both T_{mem} and MAIT cells. These included genes encoding lymph node homing receptors (e.g. *CCR7*, *SELL*), cytokine receptors (e.g. *IL7R*), costimulatory molecules (e.g. *CD27*), and TFs associated with naïve T cell fate (e.g. *BACH2*, *TCF7*, *LEF1*) (Tsukumo et al., 2013; Willinger et al., 2006). Naïve T cells showed enriched expression of gene sets upregulated in naïve $CD4^+$ T cells compared with T_{CM} and T_{EM} $CD4^+$ T cells (Fig. 4.1E,F). These findings support the reliability of my transcriptional data, and validate the use of this data to identify differences in the transcriptional profiles of $CD8^+$ MAIT cells and T_{mem} cells.

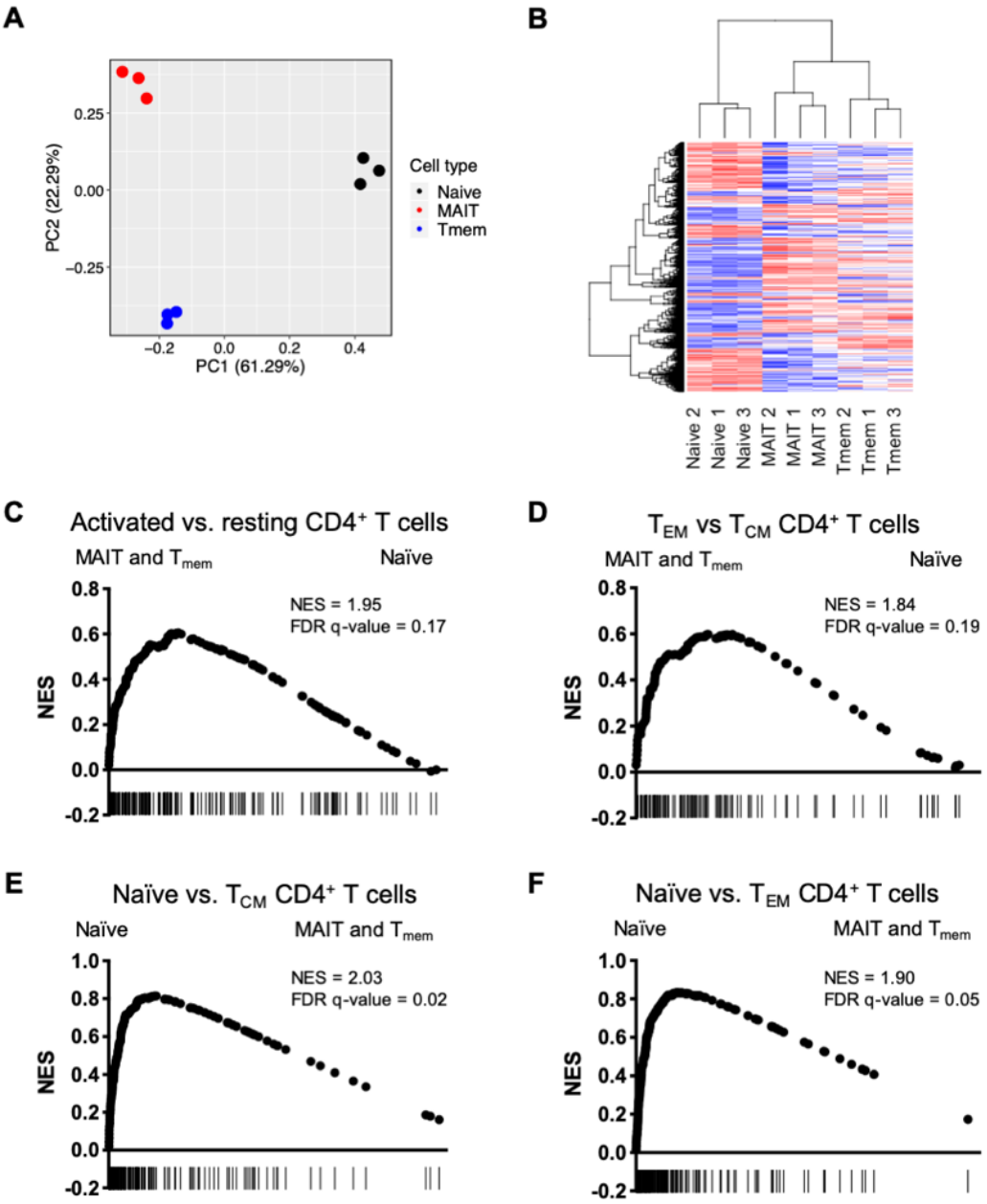
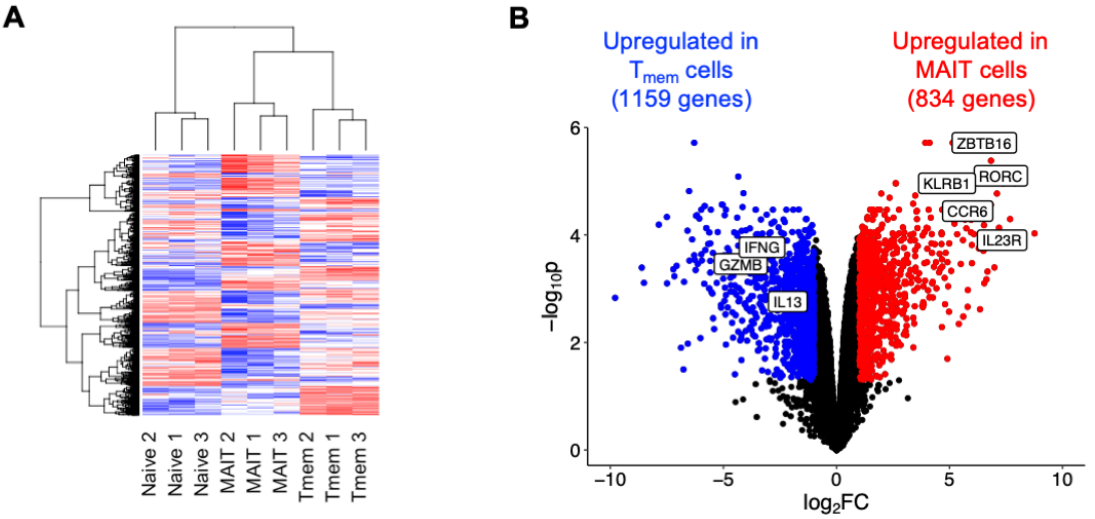


Figure 4.1. MAIT cells are transcriptionally distinct from T_{mem} and naïve T cells.
(A) PCA plot showing the relatedness of RNA-seq samples from naïve, MAIT and T_{mem} cells. **(B)** Heatmap of the genes differentially expressed between any pair of the three cell types ($log_2FC > 1$, FDR < 0.05). **(C,D)** GSEA enrichment plots showing two pathways upregulated in MAIT cells and T_{mem} cells compared with naïve T cells. **(E,F)** GSEA enrichment plots showing two pathways upregulated in naïve T cells compared with MAIT cells and T_{mem} cells.

In total, 1,993 genes were differentially expressed between $CD8^+$ MAIT cells and T_{mem} cells, with 834 genes upregulated in MAIT cells and 1,159 genes upregulated in T_{mem} cells (Fig. 4.2A). The most highly upregulated genes in MAIT cells included classic MAIT cell markers, such as *IL23R*, *RORC* (encodes RORγt), *ZBTB16* (encodes PLZF) and *KLRB1* (encodes CD161) (Fig. 4.2B). Amongst the genes downregulated in MAIT cells compared with T_{mem} cells were *GZMB*, *GNLY* and *GZMH*, consistent with the unique cytotoxic profile of MAIT cells (Kurioka et al., 2015; Leeansyah et al., 2015). Despite high expression of IFNγ by activated MAIT cells (Dusseaux et al., 2011; Ussher et al., 2014), *IFNG* gene expression was downregulated in MAIT cells relative to T_{mem} cells, suggesting distinct transcriptional regulation. Downregulation of TCR-signalling molecules (e.g. *CD8B*, *CD3G*, *IKBKE*), combined with upregulation of negative regulators (e.g. *SOCS3*, *DUSP5*, *SPRY1*), was consistent with previous publications (Tang et al., 2013; Turtle et al., 2011). Other downregulated genes encoded immune checkpoint molecules (e.g. *CTLA4*, *TIGIT*) and chemokines/chemokine receptors (e.g. *CCL3*, *CCL4L2*, *CCL3*, *CX3CR1*, *CXCR3*), suggesting differential trafficking of MAIT cells and T_{mem} cells. *CX3CR1* was one of the most highly downregulated genes in MAIT cells compared with T_{mem} cells (> 55-fold) and was expressed at similar levels to naïve T cells. Expression of CX3CR1 has been used to define three populations of memory T cells in peripheral blood, of which classical T_{EM} cells express the highest levels (Gerlach et al., 2016). Given the T_{EM} phenotype of MAIT cells, their low expression of *CX3CR1* is unexpected. T_{RM} cells are $CX3CR1^{-/low}$ and are thought to be derived from peripheral $CX3CR1^-$ effector T cells (Gerlach et al., 2016), perhaps suggesting that MAIT cells have an increased capacity for T_{RM} differentiation. Thus, further research is required to understand the functional implications of low CX3CR1 expression on MAIT cells.



C

Top 10 upregulated TFs in MAIT cells and T_{mem} cells

MAIT	T_{mem}
1. *RORC*	1. *MEIS1*
2. *ZBTB16*	2. *ZNF365*
3. *CEBPD*	3. *BNC2*
4. *HES4*	4. *ZNF683*
5. *NR4A3*	5. *PRDM5*
6. *NR1D1*	6. *SETBP1*
7. *MYC*	7. *KLF7*
8. *RUNX2*	8. *TSHZ2*
9. *EPAS1*	9. *ZNF322*
10. *GLIS3*	10. *IRF4*

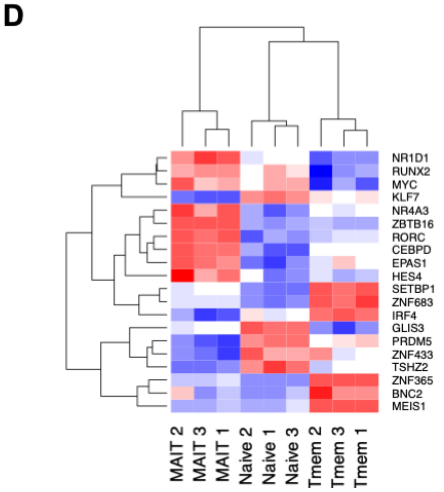


Figure 4.2. Comparison of the transcriptional profile of CD8⁺ MAIT cells and T_{mem} cells. **(A,B)** Heatmap of the 1993 genes differentially expressed between MAIT cells and T_{mem} cells ($log_2FC > 1$, FDR < 0.05). **(B)** Volcano plot showing the log_2FC and $-log_{10}p$ value obtained from differential expression analysis between MAIT cells and T_{mem} cells. Genes significantly upregulated in MAIT cells are shown in red and genes significantly downregulated in MAIT cells are shown in blue. **(C)** Table listing the top 10 upregulated TFs in MAIT cells compared with T_{mem} cells and vice versa. **(D)** Heatmap showing the expression of the top 10 TFs upregulated in MAIT cells compared with T_{mem} cells and vice versa.

Focussing on the TFs that were differentially expressed between MAIT cells and T_{mem} cells, 72 TFs show increased expression in MAIT cells and 135 in T_{mem} cells (Figure 4.2C,D; Table 4.1 and 4.2). In addition to recognised MAIT cell TFs, such as *RORC* (Dusseaux et al., 2011), *ZBTB16* (Savage et al., 2008), and *CEBPD* (Lee et al., 2018), the top 10 upregulated TFs in MAIT cells included a number of TFs whose role in MAIT cells has not been investigated, for example *NR1D1*, *MYC* and *RUNX2* (Fig. 4.2C,D; Table 4.1). Functions of these TFs described in other T cell populations may provide clues to their role in MAIT cells. For example, *NR1D1* encodes REV-ERBα, which has been shown to negatively regulate the proinflammatory function of Th17 cells (Amir et al., 2018). c-Myc (encoded by *MYC*) is predominantly known for its role in reprogramming T cell metabolism to enable rapid proliferation (Buck et al., 2015), but other functions have been described, including in the generation (Haque et al., 2016) and homeostasis (Bianchi et al., 2006) of memory $CD8^+$ T cells, and as a negative regulator of activation-induced cell death (Shi et al., 1992). A T cell-intrinsic role for Runx2 has recently been described in the maintenance of antiviral $CD8^+$ memory T cells in mice (Olesin et al., 2018).

Downregulated TFs in MAIT cells compared with T_{mem} cells (Figure 4.2C,D; Table 4.2) included the TCR-induced TFs *IRF4* and *BATF*, which bind cooperatively to DNA. IRF4 and BATF are required for prolonged antiviral $CD8^+$ T cell effector function, however high levels of these TFs can promote T cell exhaustion (Man et al., 2017). This could explain the upregulated expression of the immune checkpoint molecules *CTLA4* and *TIGIT* in T_{mem} cells compared with MAIT cells. In conventional $CD8^+$ T cells, IRF4 is required for the expression of CD122 and CD127 (Huang et al., 2017), both of which are expressed by MAIT cells (Dusseaux et al.,

2011). *IRF4* was expressed at low levels in MAIT cells, perhaps suggesting distinct mechanisms for the regulation of cytokine receptor expression in the two cell types.

4.4.2. The chromatin accessibility landscape of MAIT cells is distinct from that of naïve T cells and T_{mem} cells

$CD8^+$ naïve T cells, MAIT cells and T_{mem} cells showed distinct chromatin accessibility profiles (Fig. 4.3A). Consistent with the phenotype and function of the three cell types, the chromatin accessibility profiles of MAIT cells and T_{mem} cells were relatively similar, whilst that of naïve T cells was more distinct (Fig. 4.3A,B). Compared with the RNA-seq data, the ATAC-seq data was better able to distinguish the three cell populations, as previously reported (Sen et al., 2016). This is illustrated in Figure 4.3C, which shows the degree of correlation between each pair of the three cell types for both the ATAC-seq and RNA-seq data. The line lengths are proportional to one minus ρ, where ρ is the Spearman's rank correlation coefficient. Hence, the longer the line connecting any given pair of cell types, the weaker the correlation between them.

To characterise the differences between MAIT cells and T_{mem} cells further, I performed differential accessibility analysis. In total, 14,853 peaks were differentially accessible between MAIT cells and T_{mem} cells, comprising 7,019 and 7,834 peaks (3,222 and 3,879 genes) with increased and decreased accessibility in MAIT cells, respectively (Fig. 4.4A). Genes with increased accessibility included characteristic MAIT cell markers, such as *ZBTB16* (Fig. 4.4B), *RORC*, *KLRB1* and *IL23R*, while genes with decreased accessibility included *GZMB*, *IFNG*, *IL13* (Fig. 4.4C) and *IL4*. Decreased accessibility of *GZMB* suggests that the distinct cytotoxic

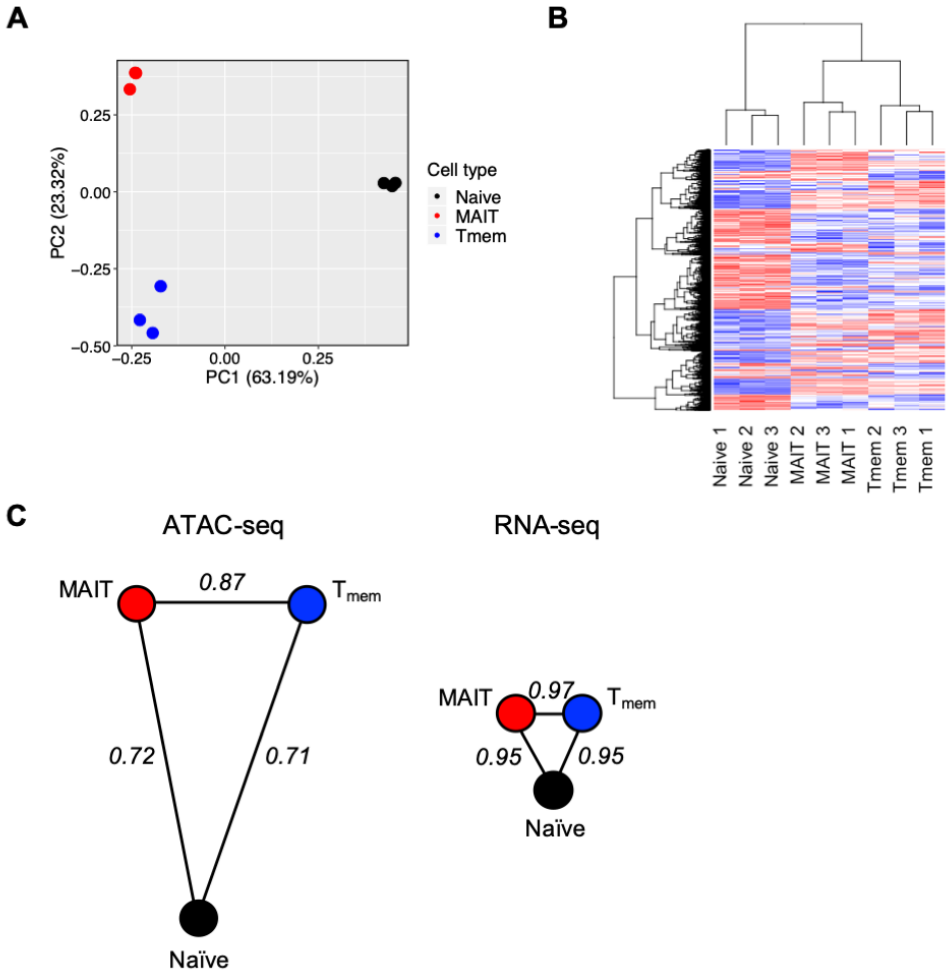


Figure 4.3. The chromatin accessibility profile of CD8⁺ MAIT cells is distinct from that of naïve T cells and T_{mem} cells. (A) PCA plot showing the relatedness of ATAC-seq samples from naïve, MAIT and T_{mem} cells. **(B)** Heatmap of the peaks differentially accessible ($log_2FC > 1$, FDR < 0.05) between any pair of the three cell types. **(C)** Correlation between the ATAC-seq (left) and RNA-seq (right) profiles of naïve T cells, MAIT cells and T_{mem} cells. Value of Spearman's rank correlation coefficient (ρ) is indicated for each comparison. The length of the line connecting each pair of circles (i.e. cell types) is proportional to 1 - ρ.

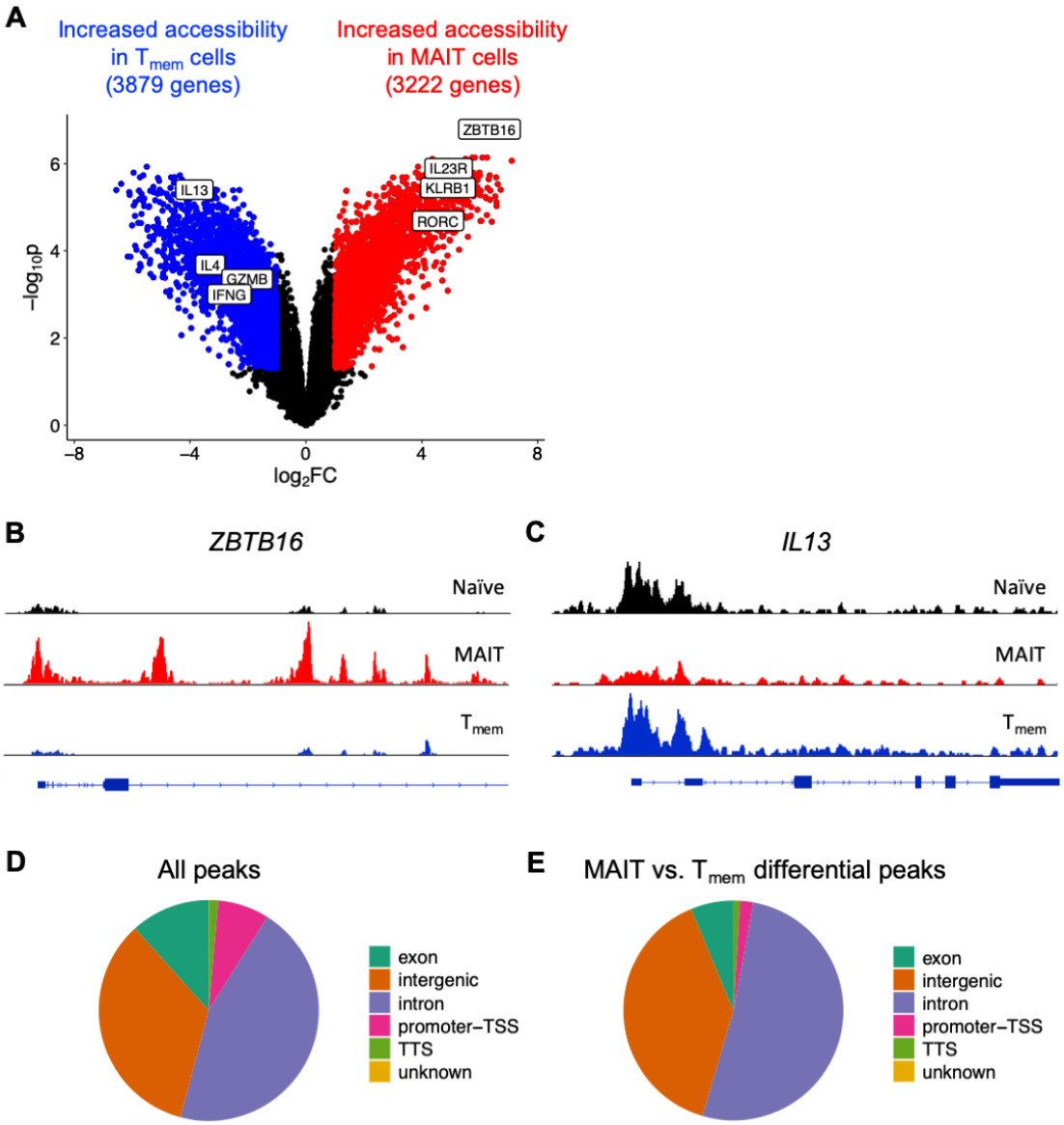


Figure 4.4. Differences in chromatin accessibility between MAIT cells and T_{mem} cells. **(A)** Volcano plot showing the log_2FC and $-log_{10}p$ value obtained from differential accessibility analysis between MAIT cells and T_{mem} cells. Genes with significantly increased accessibility in MAIT cells are shown in red and genes with significantly decreased accessibility in MAIT cells are shown in blue. **(B,C)** Representative ATAC-seq tracks showing the upstream region of the *ZBTB16* gene (B) and the *IL13* locus (C). **(D)** Genomic location of all ATAC-seq peaks. **(E)** Genomic location of the ATAC-seq peaks differentially accessible between MAIT cells and T_{mem} cells.

phenotype of MAIT cells (Kurioka et al., 2015; Leeansyah et al., 2015) may be regulated by changes in chromatin accessibility, rather than simply at the level of transcription. MAIT cells do not typically produce type 2 cytokines, although a fraction of MAIT cells can be induced to express IL-13 and IL-5 following sustained *in vitro* stimulation (Kelly et al., 2019). The reduced accessibility of type 2 cytokine genes in MAIT cells compared with T_{mem} cells, and the time delay in the expression of these cytokines, suggests that chromatin remodelling may be necessary for MAIT cells to acquire type 2 functionality.

The majority of ATAC-seq peaks were within introns and intergenic regions, with a smaller fraction located in exons and promoters (Fig. 4.4D). As a fraction of the differentially accessible peaks between MAIT cells and T_{mem} cells, intronic and intergenic peaks were further increased, whilst exonic and promoter peaks were decreased (Fig. 4.4E). Approximately 90% of all differentially accessible peaks were in introns and intergenic regions, indicating that phenotypic differences between MAIT cells and T_{mem} cells are governed largely by TF activity at distal regulatory elements, rather than at promoters.

Pathway enrichment analysis on the genes with increased accessibility in MAIT cells compared with T_{mem} cells identified interleukin signalling as the most upregulated pathway (Fig. 4.5). Although the IL-4 and IL-13 signalling pathway was enriched, the majority of the 43 enriched genes belonging to this pathway were not specific to the type 2 response and instead included genes from multiple cytokine signalling pathways, for example *RORC*, *IL12B*, *IL17A*, *IL17F*, *MYC*, *IRF4*, *IL23A* and *OSM*. Therefore the pathway enrichment analysis is not inconsistent with the downregulated accessibility of *IL13* and *IL4* in MAIT cells compared with T_{mem} cells (Fig. 4.4A,C). MAIT cells are highly cytokine responsive and can be activated by

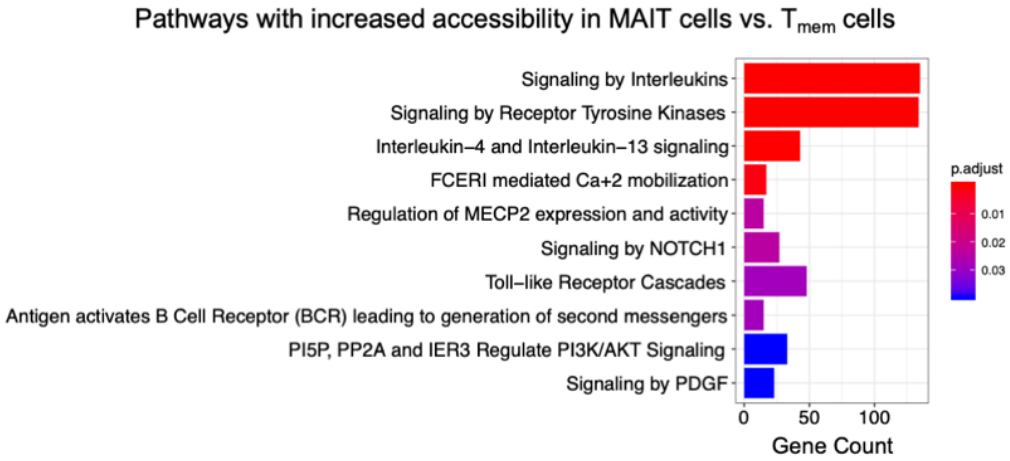


Figure 4.5. Interleukin signalling genes show increased accessibility in MAIT cells compared with T_{mem} cells. Reactome pathway enrichment analysis of the genes with increased accessibility in MAIT cells compared with T_{mem} cells (top 10 enriched pathways are shown).

various combinations of cytokines in the absence of TCR stimulation, including IL-12 + IL-18 and IL-15 + IL-18 (Sattler et al., 2015; Ussher et al., 2014; van Wilgenburg et al., 2016). MAIT cells also secrete a variety of cytokines upon activation, such as IFNγ and IL-17A (Dusseaux et al., 2011). The increased accessibility of genes involved in cytokine signalling in MAIT cells compared with T_{mem} cells suggests that the chromatin accessibility landscape of MAIT cells is primed for both responsiveness to, and production of, cytokines. In addition, MAIT cells express high levels of chemokine receptors such as CCR2 and CCR5, which are also found within the enriched interleukin signalling pathway, indicating that the tissue-homing potential of MAIT cells is also embedded within their chromatin landscape. Pathway analysis on the genes with reduced accessibility in MAIT cells compared with T_{mem} cells identified only a single significantly enriched pathway ("Diseases of signal transduction", adjusted p = 0.047).

Comparison of differentially expressed genes and differentially accessible peaks provides an insight into the regulation of MAIT cell gene expression. Having separately analysed changes in gene expression and chromatin accessibility between MAIT cells and T_{mem} cells, I wanted to compare the results from the two analyses. This would allow me to identify genes for which the difference in expression between MAIT cells and T_{mem} cells was associated with changes in chromatin accessibility, and those for which differences in expression occurred in the absence of altered chromatin accessibility.

Of the 834 genes upregulated in MAIT cells compared with T_{mem} cells, roughly half (441 genes) also showed increased chromatin accessibility (Fig. 4.6A,E; Table 4.3). Genes that showed both increased expression and chromatin accessibility encompassed the majority of classical MAIT cell markers, including *RORC*,

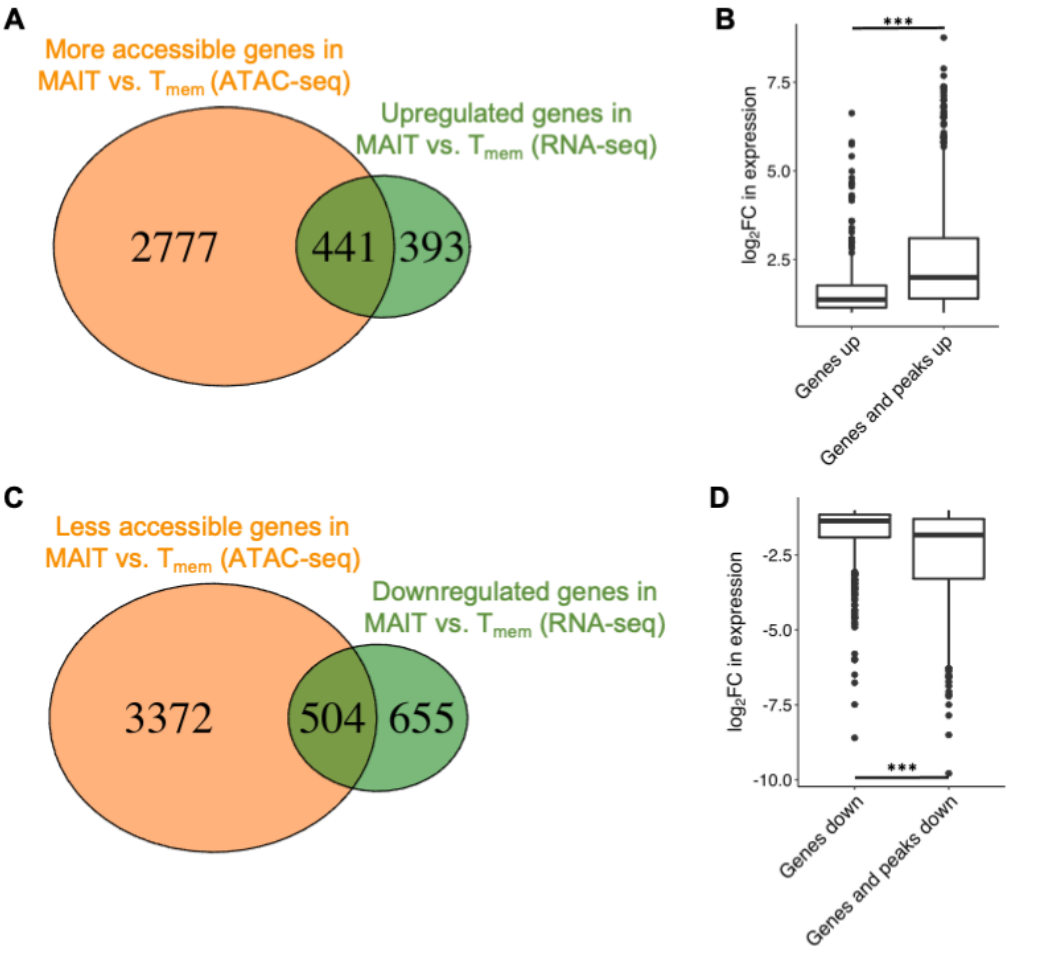
ZBTB16, *IL23R*, *KLRB1*, *DPP4*, *CCR2* and *CCR6*. Genes with increased expression but not accessibility included *CD69*, *LTA*, *NFKB2*, *ICAM1* and *IL2RB*, genes that are upregulated upon activation in both MAIT cells and T_{mem} cells. Genes whose increased expression was accompanied by increased accessibility showed a significantly greater average fold change in expression compared with genes that showed increased expression but not accessibility (Fig. 4.6B). Finally, there were 2,777 genes (5,442 peaks) that showed increased accessibility but no change in expression between MAIT cells and T_{mem} cells. These comprised a number of genes that are upregulated upon MAIT cell activation, for example the effector molecules *IFNG*, *IL17A* and *IL17F*, and the cytokine receptors *IL12RB1* and *IL15RA*. The increased accessibility but not expression of these genes suggests that MAIT cells are primed for their expression in response to the appropriate stimuli.

I performed a similar analysis to compare the genes and peaks with decreased expression and accessibility, respectively, in MAIT cells compared with T_{mem} cells. Similar to upregulated genes, approximately half (504/1,159 genes) of downregulated genes showed reduced accessibility in MAIT cells compared with T_{mem} cells (Fig. 4.6C,E; Table 4.4). These included cytotoxic proteins (e.g. *GZMB*, *GZMH*, *GNLY*), chemokines (e.g. *CCL3*, *CCL4*) and cytokines (e.g. *IFNG*, *IL13*), as well as positive regulators of TCR signalling (e.g. *ITK*, *MAL*, *CD8B*, *KRAS*, *IKBKE*). Consistent with these findings, MAIT cells have a unique cytotoxic profile compared with other $CD8^+$ T cells (Kurioka et al., 2015; Leeansyah et al., 2015) and exhibit downregulated expression of TCR signalling genes (Turtle et al., 2011). Given that MAIT cells require prolonged stimulation to induce the expression of type 2 cytokines (Kelly et al., 2019), the reduced accessibility and expression of *IL13* by resting MAIT cells suggests that chromatin remodelling is required for MAIT cells to

acquire type 2 functionality. MAIT cells rapidly produce large amounts of *IFNG* upon activation at levels similar to T_{mem} cells (Dusseaux et al., 2011), so it is surprising that both the expression and accessibility of *IFNG* was reduced in MAIT cells relative to T_{mem} cells. However, as expected, *IFNG* expression and accessibility was significantly increased in MAIT cells relative to naïve T cells. Genes with decreased expression but not accessibility included *CD3G*, and the SLAM family genes *SLAMF6* and *SLAMF7*, while genes with reduced accessibility but not expression included the type 2-related genes *IL4* and *GATA3*. As with upregulated genes, the fold change in expression between MAIT cells and T_{mem} cells for genes with downregulated expression and accessibility was significantly greater than for those with reduced expression but not accessibility (Fig. 4.6D).

4.4.3. Identification of TF motifs with increased accessibility in MAIT cells compared with T_{mem} cells using HOMER *de novo* motif discovery

To identify TFs with a potential role in regulating MAIT cell-specific phenotype and function, I performed motif enrichment analysis on the peaks with increased accessibility in MAIT cells relative to T_{mem} cells. Shared peaks (peaks present in both MAIT cells and T_{mem} cells) were used as the background sequences. Specifically, I used HOMER software (Heinz et al., 2010) to perform *de novo* motif discovery. In the *de novo* approach, HOMER searches for motifs with differential enrichment between target and background sequences, in this case between MAIT cell-enriched and shared peaks, respectively, and then looks for similarity between identified motifs and known TF binding motifs. It is important to bear in mind that not all TFs have a known binding motif. For example, *ZBTB16*, which is highly



E

	Expression	Expression and accessibility	Accessibility
Up	*CD69*, *LTA*, *NFKB2*, *ICAM1*, *IL2RB*	*RORC*, *ZBTB16*, *IL23R*, *KLRB1*, *DPP4*, *CCR2*, *CCR6*	*IFNG*, *IL17A*, *IL17F*, *IL12RB1*, *IL15RA*
Down	*CD3G*, *SLAMF6*, *SLAMF7*	*GZMB*, *GZMH*, *GNLY*, *CCL3*, *CCL4*, *IFNG*, *IL13*, *ITK*, *MAL*, *CD8B*, *KRAS*, *IKBKE*	*IL4*, *GATA3*

Figure 4.6. Overlap of differentially expressed genes and differentially accessible peaks between MAIT cells and T_{mem} cells. (A) Venn diagram showing the overlap of genes with increased expression and increased accessibility in MAIT cells compared with T_{mem} cells. **(B)** Bar plot showing the average log_2FC in expression of genes with increased expression but not accessibility ("Genes up") compared with genes with increased expression and accessibility ("Genes and peaks up"). **(C)** Venn diagram showing the overlap of genes with decreased expression and decreased accessibility in MAIT cells compared with T_{mem} cells. **(D)** Bar plot showing the average log_2FC in expression of genes with decreased expression but not accessibility ("Genes down") compared with genes with decreased expression and accessibility ("Genes and peaks down"). Error bars show mean ± SEM. Wilcoxon Rank Sum test in (B,D). *** $p < 0.001$. **(E)** Table showing example genes with increased (up) or decreased (down) expression, expression and accessibility, or accessibility only, in MAIT cells compared with T_{mem} cells.

expressed by MAIT cells and regulates their innate-like functionality (Koay et al., 2016; Kovalovsky et al., 2008; Savage et al., 2008), does not have a binding motif in any of the main public motif databases and thus will not be identified by motif enrichment analysis, despite its important role in MAIT cell biology.

I performed the motif analysis twice, first (Analysis 1) using just the central peak regions (± 100 bp around peak summits; Fig. 4.7A) where the majority of TF binding sites are expected to be located, and the second time (Analysis 2) using whole peaks (Fig. 4.7B). In both analyses, the most enriched motif matched a RORγt binding motif derived from ChIP-seq (Fig. 4.7A,B). RORγt is highly expressed in MAIT cells (Dusseaux et al., 2011), and is required for the differentiation and function of Th17 cells (Castro et al., 2017; Ivanov et al., 2006). Thus, enrichment of the RORγt motif provided confidence in the results of my HOMER analysis.

The second most upregulated motif in both analyses was associated with a different member of the activator protein-1 (AP-1) family, namely FOSL2 (also known as Fra-2) in the peak centres analysis (Fig. 4.7A) and c-Jun in the whole peaks analysis (Fig. 4.7B). However, the sequence of the enriched motif was highly similar in both instances (Fig. 4.7C). The AP-1 complex is a dimer predominantly comprising proteins from the Fos, Jun and ATF protein families, and is involved in the regulation of numerous cellular processes, including proliferation, differentiation and apoptosis (Hess et al., 2004). Different AP-1 dimers show relatively similar sequence specificity. Therefore, the detection of FOSL2 and c-Jun motifs by the two analyses might more generally reflect the involvement of the AP-1 TF family in MAIT cells, as opposed to these specific factors. Nevertheless, FOSL2 is a promising candidate, as it was upregulated in expression in MAIT cells compared with T_{mem} cells in my

RNA-seq data, and is essential for the normal development and function of iNKT cells (Lawson et al., 2009), a population of innate-like T cells closely related to MAIT cells (Garner et al., 2018). In sum, particular AP-1 complexes may have an important role in the regulation of MAIT cell development, phenotype and/or function. The exact composition of these complexes will need to be determined experimentally.

The third most enriched motif in MAIT cells was RUNX2 in Analysis 1 and RUNX1 in Analysis 2. As with the AP-1 motifs, the motifs identified in the two analyses were similar (Fig. 4.7D). *RUNX2* expression was highly upregulated in MAIT cells compared with T_{mem} cells in my RNA-seq data (Fig. 4.2C), while *RUNX1* was highly expressed in both T cell subsets. While RUNX1 has diverse roles in T cells (Wong et al., 2011), RUNX2 has mostly been studied for its role in osteoblast differentiation (Komori et al., 1997). However, more recent studies have identified specific roles for RUNX2 in T cells, including in the long term maintenance of LCMV-specific $CD8^+$ T cells in mice (Olesin et al., 2018) and a possible role in thymic development (Vaillant et al., 2002). Given the upregulated expression of and motif enrichment for RUNX2 in MAIT cells, these findings suggest a potentially important and specific role for RUNX2 in MAIT cells.

4.4.4. AME motif enrichment analysis gives similar results to HOMER *de novo* motif discovery

To provide further support for my findings with HOMER, I performed motif enrichment analysis using the AME tool from the MEME Suite (Bailey et al., 2009; McLeay and Bailey, 2010). AME identifies enrichment of known TF binding motifs

A MAIT enriched vs. MAIT and T_{mem} shared – peak centres (Analysis 1)

Best match motif	Motif source	p value	% of targets	% of background
1. RORγt	ChIP-seq	1×10^{-253}	10.69	2.81
2. FOSL2	ChIP-seq	1×10^{-35}	15.09	10.70
3. RUNX2	ChIP-seq	1×10^{-32}	20.27	15.43
4. ZEB1	JASPAR	1×10^{-30}	16.78	14.49
5. DMRT3	JASPAR	1×10^{-15}	0.57	0.13
6. CEBPE	JASPAR	1×10^{-15}	5.15	3.46
7. MafB	JASPAR	1×10^{-14}	0.16	0.01

B MAIT enriched vs. MAIT and T_{mem} shared – whole peaks (Analysis 2)

Best match motif	Motif source	p value	% of targets	% of background
1. RORγt	ChIP-seq	1×10^{-194}	17.96	6.41
2. c-Jun	ChIP-seq	1×10^{-37}	27.10	20.5
3. RUNX1 motif 1	ChIP-seq	1×10^{-25}	21.68	16.33
4. RUNX1 motif 2	ChIP-seq	1×10^{-17}	9.31	6.36
5. MEIS2	JASPAR	1×10^{-16}	0.94	0.23
6. ZEB1	JASPAR	1×10^{-13}	51.39	46.39

C

Motif 2 in Analysis 1



Motif 2 in Analysis 2



D

Motif 3 in Analysis 1



Motif 3 in Analysis 2



Figure 4.7. HOMER *de novo* motif discovery on MAIT cell-enriched ATAC-seq peaks. **(A,B)** Table showing HOMER significantly enriched TF motifs in peaks with increased accessibility in MAIT cells compared with T_{mem} cells when using shared peaks (i.e. peaks with similar accessibility in MAIT cells and T_{mem} cells) as background sequences. In Analysis 1 (A), motif enrichment was performed using the central peak regions (± 100 bp around peak summits), and in Analysis 2 (B), the whole peak regions were used. **(C,D)** Sequence logos of example enriched motifs identified by HOMER. Motif 2 from (A) and (B) is shown in (C). Motif 3 from (A) and (B) is shown in (D).

(obtained from user-specified databases) in test sequences relative to control sequences. As with HOMER analysis, I looked for enrichment of TF motifs in peaks (± 100 bp around peak centres) with increased accessibility in MAIT cells compared with T_{mem} cells, using shared peaks as background.

For Analysis 1 (Fig. 4.8A), I selected the "Vertebrates (*In vivo* and *in silico*)" motif database, which contains motifs from JASPAR (Fornes et al., 2019), UniPROBE (Hume et al., 2015) and Jolma 2013 (Jolma et al., 2013). For Analysis 2 (Fig. 4.8B), I selected the "HOCOMOCO Human (v11 CORE)" database, which as the name suggests contains the core human motif collection from HOCOMOCO (Kulakovskiy et al., 2018). Both analyses gave very similar results to HOMER (Fig. 4.8A,B). The top 20 enriched motifs from Analysis 1 and Analysis 2 included several members of the retinoic acid-related orphan receptor (ROR) family, including RORC; AP-1 TFs, such as FOSL2 and JUNB; and members of the runt-related TF (RUNX) family. Multiple members of each family were identified due to the similarity of motif binding sites for related TFs (Lambert et al., 2018b). Thus, as discussed above, it is not possible to determine which member(s) of each family has a central role in MAIT cells, although looking at whether the TF is expressed can help to reduce the list of potential candidates. The motif for NR1D1 was also identified as enriched in Analysis 2. *NR1D1* (encodes REV-ERBα) was the sixth most upregulated TF in MAIT cells compared with T_{mem} cells in my RNA-seq data, and as discussed, REV-ERBα negatively regulates Th17 cell functions, including production of IL-17A (Amir et al., 2018).

Overall, the results of my motif enrichment analysis suggest roles for TFs from the ROR, AP-1 and RUNX families as key regulators of blood MAIT cell phenotype.

A

MAIT cell-enriched – vertebrates database (Analysis 1)

Enriched motif	E-value
RORC	1.52×10^{-265}
RORB	1.41×10^{-196}
RORA(var.2)	5.75×10^{-135}
RORA	2.40×10^{-113}
FOSL1::JUND	6.08×10^{-69}
FOS::JUN	3.20×10^{-62}
FOSL1::JUNB	1.42×10^{-61}
FOSL2::JUNB	3.19×10^{-59}
FOSL2::JUND	9.18×10^{-59}
FOSB::JUNB	1.55×10^{-58}
FOSL2::JUN	4.67×10^{-58}
FOS	4.96×10^{-58}
FOS::JUNB	2.53×10^{-57}
FOSL1	8.17×10^{-57}
JUN::JUNB	1.15×10^{-56}
JUN(var.2)	1.68×10^{-56}
FOSL1::JUN	4.69×10^{-56}
RUNX1	4.23×10^{-54}
FOS::JUND	9.77×10^{-53}
JUNB	5.42×10^{-52}

B

MAIT cell-enriched – HOCOMOCO database (Analysis 2)

Enriched motif	E-value
RORG (RORC)	6.60×10^{-163}
FOSB	7.17×10^{-69}
RUNX3	3.25×10^{-67}
JUNB	3.79×10^{-67}
JUND	1.57×10^{-66}
FOS	6.63×10^{-66}
JUN	9.85×10^{-64}
RUNX1	4.34×10^{-62}
FOSL2	4.46×10^{-60}
FOSL1	2.35×10^{-59}
RORA	9.71×10^{-47}
RUNX2	3.43×10^{-45}
PEBB/CBFB	2.70×10^{-35}
NF2L1	1.80×10^{-34}
BACH2	1.36×10^{-23}
NR1D1	3.74×10^{-21}
FOXK1	5.41×10^{-15}
FOXM1	1.28×10^{-13}
TCF7	1.38×10^{-13}
NF2L2	2.47×10^{-13}

Figure 4.8. AME motif enrichment analysis produces similar results to HOMER *de novo* motif discovery. (A,B) Table showing the top 20 AME significantly enriched TF motifs in peaks with increased accessibility in MAIT cells compared with T_{mem} cells when using shared peaks (i.e. peaks with similar accessibility in MAIT cells and T_{mem} cells) as background sequences. In Analysis 1 (A), motif enrichment was performed using the "Vertebrates (*In vivo* and *in silico*)" TF motif database, and in Analysis 2 (B), the "HOCOMOCO Human (v11 CORE)" database was used. E-value = the expected number of motifs that would be as enriched in the test sequences as this one. The E-value is the adjusted p value multiplied by the number of motifs in the motif file(s).

4.4.5. CENTIPEDE TF footprinting analysis

RORγt did not produce a detectable footprint

While motif analysis enables you to identify TF motifs that are enriched within particular genomic sequences, for example in ATAC-seq peaks, it does not provide information on which of these motifs are actually bound by the corresponding TF in your cell type or condition of interest. Having identified some candidates for key regulators of MAIT cell phenotype, I wanted to identify at which genes these TFs were bound in resting MAIT cells. While this could be achieved using ChIP-seq, this would require further experimental work and cost, and thus would be feasible only for a few TFs. In contrast, using the computational TF footprinting approach, it is theoretically possible to profile the binding sites of many TFs using one set of experimental data, for example from DNase-seq or ATAC-seq.

To perform footprinting, I used the CENTIPEDE algorithm (Pique-Regi et al., 2011), which employs a motif-centric footprinting approach. The CENTIPEDE algorithm first identifies candidate binding sites for a given TF based on their similarity with the PWM and then generates an unsupervised Bayesian mixture model to categorise each motif instance as either “bound” or “unbound”. In reality, “bound” sites are those at which the TF is bound in a large fraction of cells at that particular time. As well as the PWM match score and chromatin accessibility data, in my case from ATAC-seq, the model considers additional genomic information, such as evolutionary sequence conservation and distance to a TSS.

To test the effectiveness of CENTIPEDE, I started with RORγt, as *RORC* was the most upregulated TF relative to T_{mem} cells in my RNA-seq data (Fig. 4.2C) and its

binding motif was highly enriched amongst MAIT cell-specific ATAC-seq peaks (Fig. 4.7A,B and 4.8). Using the first MAIT cell ATAC-seq replicate, CENTIPEDE predicted that 4,292 of the 87,086 RORγt motif instances across the genome were bound in MAIT cells. Visualising the average number of Tn5 cut sites around bound RORγt motifs (9 bp motif ± 100 bp) revealed a clear preference for cutting at specific sites within the motif (Fig. 4.9A). The number of cut sites at all positions within the motif is expected to be lower than the surrounding region, due to the bound TF protecting the underlying DNA from Tn5 cutting, thus generating a "footprint" (Buenrostro et al., 2013; Sung et al., 2016). However, some sites within the RORγt motif actually showed an increased frequency of cutting relative to the surrounding region.

I next visualised the average number of Tn5 cut sites around unbound RORγt motifs. Unexpectedly, the pattern of cutting was very similar to the bound sites, although the overall accessibility of the unbound motifs was reduced (Fig. 4.9B). This suggested that the pattern observed at RORγt motifs was not a consequence of TF binding, but instead reflected the sequence preference of the Tn5 transposase. I obtained a similar result with the other two MAIT cell ATAC-seq replicates. Thus, I was not confident in the results of the CENTIPEDE analysis for RORγt.

CTCF produces a clear footprint at bound sites but not at unbound sites

To assess whether CENTIPEDE might work for other TFs, I inferred the binding sites for CTCF, as this TF has previously been shown to produce a clear footprint with ATAC-seq data (Buenrostro et al., 2013). Of the 45,013 CTCF motif instances

across the genome, CENTIPEDE predicted 11,805-12,327 of these to be bound in the three replicates of MAIT cells. A clear footprint was detected at bound CTCF motif sites (Fig. 4.9C), but not at unbound motif sites (Fig. 4.9D), suggesting that CENTIPEDE was able to accurately infer CTCF binding. As observed for RORγt, the Tn5 transposase displayed a preference for cutting at particular sites within the CTCF motif, irrespective of whether they were bound by the TF.

Given my success in identifying CTCF binding sites, I decided to test CENTIPEDE on other TFs with known or potential roles in MAIT cells, including T-bet, FOSL2 and NR1D1. However, none of these TFs produced a clear footprint (Fig. 4.9E-G).

Identification of causal TFs using IMAGE

As CENTIPEDE did not work for the TFs in which I was interested, I looked for other tools that could provide an insight into MAIT cell gene regulation. IMAGE uses a machine learning approach to predict the TFs driving observed differences in expression between cell types or experimental conditions of interest, as well as the binding sites and target genes of these TFs (Madsen et al., 2018). To achieve this, IMAGE uses a custom PWM database, and requires as input information on gene expression and the location of regulatory sequences, in my case provided by RNA-seq and ATAC-seq data, respectively. I used IMAGE to look for TFs that regulate the differences between MAIT cells and T_{mem} cells. Factors identified as high confidence causal TFs (i.e. TFs responsible for expression differences between MAIT cells and T_{mem} cells) are shown (Fig. 4.10A). The motif activity is a score determined by IMAGE for each TF motif and indicates how much the activity of a certain motif contributes to the activity of all enhancers. Increased motif activity in MAIT cells relative to T_{mem} cells indicates a greater importance of that motif for

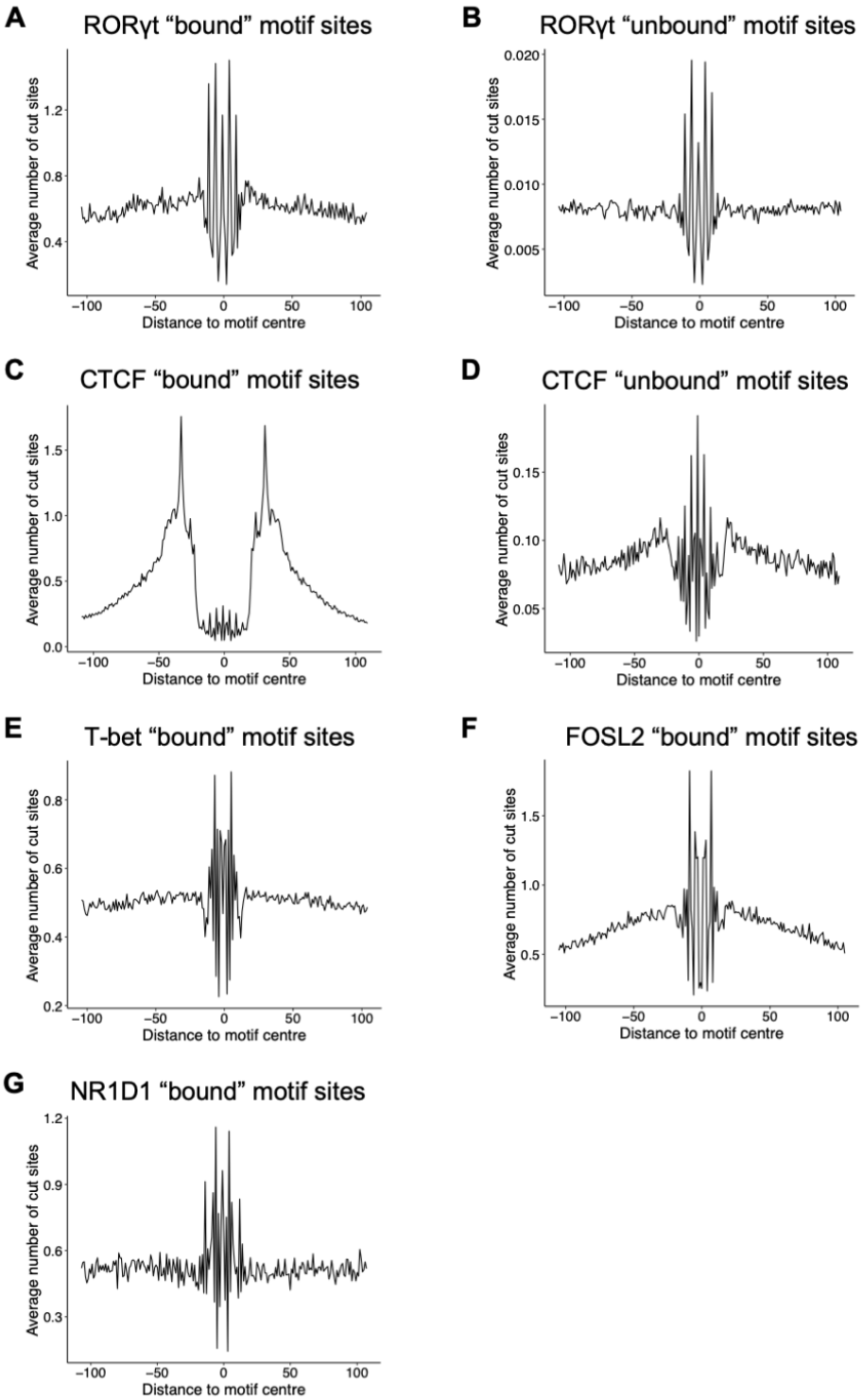


Figure 4.9. CENTIPEDE TF footprinting analysis was effective for CTCF but not for other tested TFs. (A-D) Average number of Tn5 cut sites in the region ± 100 bp around RORγt (A,B) or CTCF (C,D) motif sites that were deduced by CENTIPEDE to be "bound" (A,C) or "unbound" (B,D). **(E-G)** Average number of Tn5 cut sites in the region ± 100 bp around T-bet (E), FOSL2 (F) and NR1D1 (G) motif sites that were deduced by CENTIPEDE to be "bound". All plots show results from a representative MAIT cell sample.

driving gene expression in MAIT cells compared with T_{mem} cells, and vice versa. Along with its highly increased expression in MAIT cells compared with T_{mem} cells, the motif activity of RORγt (encoded by *RORC*) was highly increased in MAIT cells compared with T_{mem} cells, suggesting a key role for RORγt in regulating MAIT cell phenotype. The expression of RORγt was highly positively correlated with its motif activity, suggesting that it functions as an enhancer of gene expression. Given the assumed role of RORγt in regulating the type 17 functionality of MAIT cells (Dusseaux et al., 2011; Ivanov et al., 2006), this finding validates the use of IMAGE analysis for identifying important transcriptional regulators of MAIT cell phenotype.

Both FOS and FOSL2 were upregulated in expression and had a highly increased motif activity in MAIT cells compared with T_{mem} cells, and their expression was positively correlated with motif activity. This supports the hypothesis that specific AP-1 complexes have a key role in MAIT cell regulation, as suggested by my motif enrichment analysis. Furthermore, consistent with my motif enrichment analysis, RUNX2 showed increased expression and motif activity in MAIT cells compared with T_{mem} cells, and *RUNX2* expression was positively correlated with motif activity. NR1D1, which was identified in one of my motif enrichment analyses (Fig. 4.8B), was increased in expression in MAIT cells compared with T_{mem} cells. IMAGE identified two NR1D1 motifs as important causal transcriptional regulators of the differences in gene expression between MAIT cells and T_{mem} cells. Interestingly, one NR1D1 motif showed increased activity in MAIT cells and NR1D1 expression was positively associated with its activity, while the other motif showed reduced activity in MAIT cells and NR1D1 expression was negatively associated with its activity. This suggests that NR1D1 has increased enhancer functionality in MAIT cells and increased suppressor functionality in T_{mem} cells.

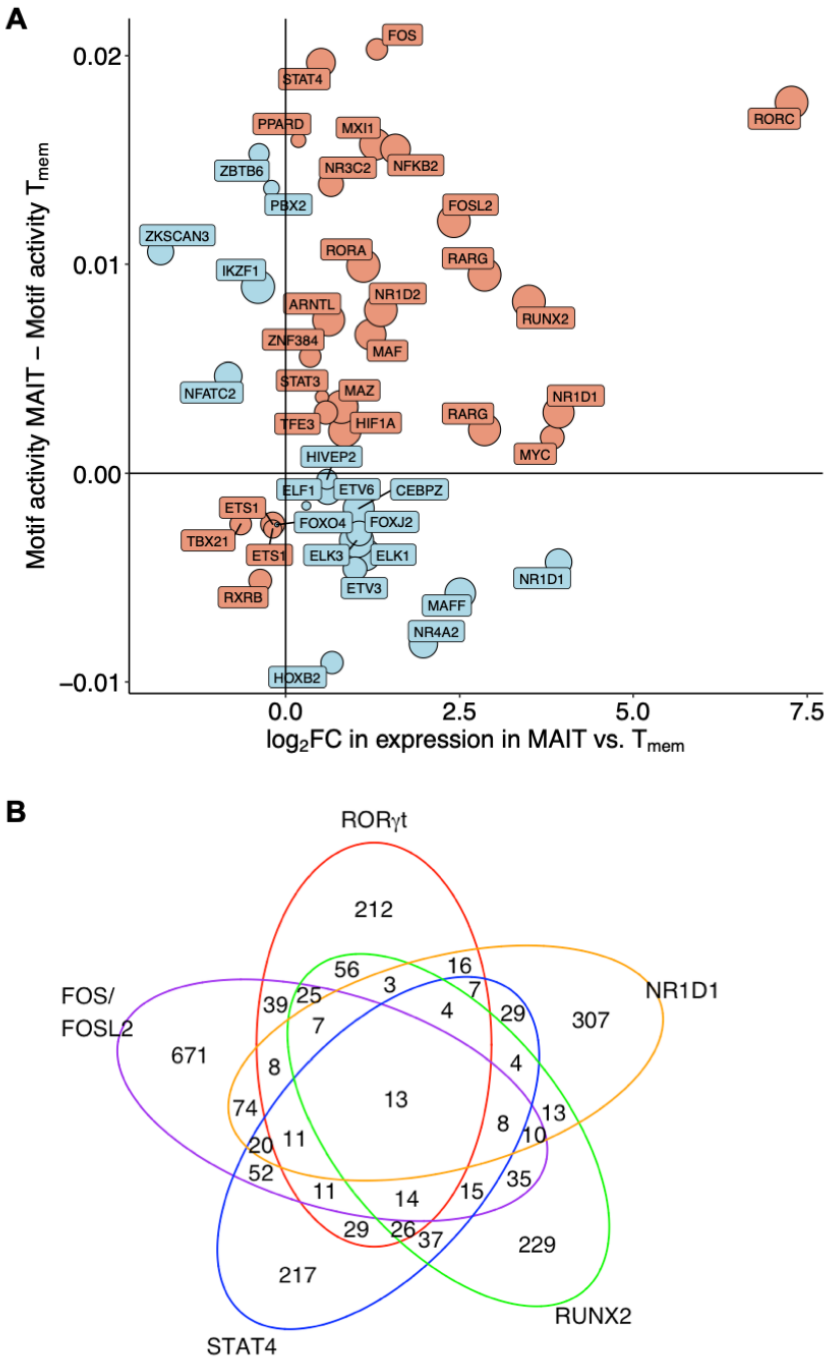


Figure 4.10. IMAGE analysis identifies known and novel TFs as key regulators of MAIT cell phenotype and/or function. (A) Results from IMAGE analysis. The gene names of the causal TFs designated as high confidence by IMAGE are shown. The motif activity value indicates how much each TF motif contributes to the activity of all enhancers. The x axis shows the log_2FC in expression of the TF in MAIT cells compared with T_{mem} cells from my RNA-seq data, and the y axis shows the difference in motif activity between MAIT cells and T_{mem} cells. Circles in red and blue indicate a positive and negative Pearson correlation coefficient, respectively, between the expression pattern of the TF and the motif activity of the associated motif. The size of the circles indicates the absolute magnitude of the Pearson correlation coefficient. **(B)** Venn diagram showing the overlap of target genes for TFs predicted by IMAGE analysis to be key regulators of MAIT cell phenotype.

STAT4 showed one of the largest increases in motif activity in MAIT cells compared with T_{mem} cells and its expression was positively correlated with motif activity (Fig. 4.10A). STAT4 is activated downstream of the IL-12R, a key cytokine involved in MAIT cell activation (Ussher et al., 2014). Interestingly, expression of *STAT4* was not significantly different between MAIT cells and T_{mem} cells, and the STAT4 motif was not enriched in my HOMER or AME motif enrichment analyses. This suggests that regulation of STAT4 motif activity may not be governed by *STAT4* expression or the number of STAT4 motifs in open chromatin regions.

As well as identifying causal transcriptional regulators, IMAGE outputs a list of predicted target genes for these TFs. IMAGE predicted target genes for RORγt included known RORγt-regulated genes (Castro et al., 2017), for example *IL23R*, *CCR6* and *CCL20*, supporting the validity of IMAGE target gene prediction. Therefore, I compared the IMAGE predicted target genes for RORγt, the AP-1 factors FOS and FOSL2, STAT4, RUNX2 and NR1D1 (only included targets of the NR1D1 motif with an increased activity score in MAIT cells compared with T_{mem} cells; Fig. 4.10B). Thirteen genes were regulated by all six TFs, including genes specifically associated with MAIT cell phenotype, such as *DPP4* (encodes CD26), *CCR6* and *CXCR6*. In total, 2202 genes were regulated by one or more of these TFs. The majority of gene targets were unique to a particular TF, suggesting that these TFs could collectively explain a large amount of MAIT cell gene expression. Importantly, as mentioned earlier, the TF PLZF that regulates the innate-like functionality of MAIT cells (Koay et al., 2016; Kovalovsky et al., 2008; Savage et al., 2008), lacks a known binding motif. Therefore, some of the genes identified as targets for the tested factors could also be regulated by PLZF.

Overall, the results of IMAGE support my motif enrichment analysis. Together, they suggest potentially important roles for AP-1 complexes, RUNX2, NR1D1 and STAT4 in regulating MAIT cell phenotype, along with the known role for RORγt. Further experimental work is required to test and validate these findings.

4.5. Discussion

As well as confirming their distinct transcriptional profile (Billerbeck et al., 2010; Salou et al., 2018), my results demonstrate that MAIT cells and conventional T_{mem} cells have a distinct chromatin accessibility landscape. Furthermore, I have identified a list of candidate TFs with potentially important roles in governing the unique phenotype and function of MAIT cells. These are worth investigating in future studies.

Consistent with their effector memory phenotype, the transcriptome and chromatin accessibility landscape of MAIT cells was more similar to that of T_{mem} cells than to that of naïve T cells. Classical MAIT cell markers, such as *ZBTB16*, *RORC* and *KLRB1* were highly upregulated in expression but also tended to show large increases in chromatin accessibility relative to T_{mem} cells. The 441 genes with significantly increased expression and accessibility in MAIT cells compared with T_{mem} cells provide a refined MAIT cell signature gene set. This will be useful for characterising precise changes in the MAIT cell transcriptome in different diseases and may enable the identification of gene targets for therapeutic modulation of MAIT cell function.

Compared with T_{mem} cells, MAIT cells showed enriched accessibility at genes associated with cytokine signalling, indicating that the sensitivity of MAIT cells to cytokine stimulation and/or their capacity to rapidly produce large amounts of cytokines (Garner et al., 2018), is encoded in their resting chromatin accessibility landscape. While the genes for type 1 and type 17 cytokines were generally accessible, type 2 cytokines such as IL-13 showed reduced accessibility in MAIT cells relative to T_{mem} cells (Figure 4.4A,C). The need for chromatin opening in order for MAIT cells to acquire type 2 functionality may explain why such cytokines are only produced following prolonged stimulation (Kelly et al., 2019). As well as being primed for cytokine responsiveness and/or production, the unique cytotoxic profile of MAIT cells (Kurioka et al., 2015; Leeansyah et al., 2015), including low resting *GZMB* expression, was also encoded within the chromatin landscape.

TF motif enrichment analysis with HOMER and AME produced relatively similar results. In addition to RORγt, a known MAIT cell TF (Dusseaux et al., 2011), and regulator of type 17 differentiation and function (Castro et al., 2017; Ivanov et al., 2006), my analysis identified several other TF families that may be important regulators of MAIT cell phenotype and function. These included members of the AP-1 family, RUNX factors and NR1D1, the roles of which have not been investigated in MAIT cells. In addition to supporting these findings, IMAGE analysis revealed a potential role for STAT4 in governing the characteristic MAIT cell phenotype. STAT4 is activated downstream of IL-12 signalling (Watford et al., 2004), so its activation in MAIT cells following IL-12 stimulation would be expected, however a role for STAT4 under homeostatic conditions is unknown. STAT4 is an interesting candidate as its expression was not significantly increased in MAIT cells

compared with T_{mem} cells, so its potential role would not have been identified by RNA-seq alone.

4.6. Limitations

While ATAC-seq data can provide useful insights into cellular regulation, there are a number of limitations that should be taken into account when considering using ATAC-seq in your research or when interpreting ATAC-seq data.

ATAC-seq peaks are largely found at sites of distal enhancers in intergenic and intronic regions (Klemm et al., 2019). To regulate gene expression, enhancers are typically brought into physical contact with promoters. The relative position of an enhancer relative to its target gene(s) is not fixed and can occur over distances of hundreds of kilobases (Schoenfelder and Fraser, 2019). Moreover, enhancers can interact with multiple targets and the function of enhancers is highly cell type-specific. Thus, assignment of enhancers to their target genes is challenging and methods to improve the accuracy of enhancer-gene pairing are under constant development. For the best pairing predictions, functional data, for example from conformation capture or imaging techniques, is often combined with complex computational analysis (Hariprakash and Ferrari, 2019; Kempfer and Pombo, 2019). In the majority of ATAC-seq studies, such functional data is not available. Instead, most peak annotation software, including HOMER which was used in this study, assigns ATAC-seq peaks to their nearest gene. Hence, the assignment of ATAC-seq peaks to genes is imperfect and this may affect interpretation of the results.

There are also several limitations of motif enrichment analysis. As discussed, members of a TF family usually have highly related DNA-binding motifs (Todeschini et al., 2014), so an enriched motif can only be accurately assigned to a family of TFs. Identifying the relevant TF from this family requires further experimental data, for example on the expression of the TFs. Moreover, motif enrichment analysis does not provide information on which of the TF motifs are actually bound by the TF, or at which sites TF binding is having a functional effect. Other TFs, including PLZF which is highly expressed in MAIT cells, lack a known binding motif (Lambert et al., 2018a).

ChIP-seq can provide detailed information on the binding sites of an individual TF. However, it is generally only feasible to perform ChIP-seq for a few TFs in a given study, due to the time and cost of generating and analysing the data. When first described, TF footprinting was thought to solve this limitation, enabling the genome-wide binding sites of the majority of TFs (those with a known DNA-binding motif) to be identified in a given cell type or condition with a single set of experimental data. While originally designed for DNase-seq data, ATAC-seq data was also shown to be suitable for TF footprinting (Buenrostro et al., 2013). However, recent studies indicate that the majority of TFs do not have a measurable footprint. This was also demonstrated through analysis of my ATAC-seq data using CENTIPEDE. Only CTCF and none of the TFs of interest produced a measurable footprint, while a signature reflecting the cutting preference of the Tn5 transposase was identified at the motif binding sites of all TFs (Fig. 4.9). The ability to identify a footprint for a particular TF has been shown to vary depending on cell type (Karabacak Calviello et al., 2019).

A number of studies have suggested that DNase-seq data is more amenable to TF footprinting than ATAC-seq data (Karabacak Calviello et al., 2019; Schwessinger et al., 2017). However, a similar accuracy of TF footprinting was achieved with ATAC-seq and DNase-seq data using a couple of methods. One of these takes into account the sequence preferences of the transposase (Li et al., 2019), while the other uses a supervised learning approach that accounts for heterogeneity in the footprint signal at individual TF binding sites (Quach and Furey, 2016). The ATAC-seq experimental protocol also affects TF footprinting, with Omni-ATAC data being superior to that generated using Fast-ATAC (as performed in my study) or the original ATAC-seq protocol (Li et al., 2019). Therefore, the success of TF footprinting appears to be considerably dependent on the chosen experimental and computational analysis approach.

My study compared the relatively homogeneous MAIT cell subset with total T_{mem} cells that are more diverse. The lack of inclusion of specific T_{mem} cell populations will have influenced the genes and peaks identified as MAIT cell-specific. For example, *RORC*, *IL17A* and other type 17 genes showed increased expression and/or accessibility in MAIT cells compared with T_{mem} cells. However, it is unclear how the expression and accessibility of these genes in MAIT cells compares with Th17 cells.

Finally, due to cost and the limited availability of the technology at the time, bulk ATAC-seq was performed rather than single-cell ATAC-seq (Buenrostro et al., 2015b; Cusanovich et al., 2015). Hence, the size of each peak in my data reflects the average accessibility of the region across the analysed population, and it was not possible to investigate the presence of distinct MAIT cell subsets. However, at the protein level and at the transcriptional level (as demonstrated in Chapter 5),

MAIT cells are relatively homogeneous, so I am not expecting the use of bulk ATAC-seq to have had a major impact on the results.

4.7. Future plans

In this work, I have identified several TFs or families of TFs with potentially important roles in regulating MAIT cell phenotype and function. However, the role of these TFs requires experimental testing. This can be achieved using a combination of gene knockout mice and TF knockdown in human MAIT cells, including under resting conditions, following MAIT cell *in vitro* and *in vivo* activation, and in disease settings. A particularly interesting TF that I plan to explore further is STAT4, given its highly enriched motif activity in MAIT cells compared with T_{mem} cells (determined using IMAGE), despite comparable expression in the two cell types.

4.8. Conclusion

In conclusion, I have shown that MAIT cells and T_{mem} cells are transcriptionally and functionally distinct, with around half of their differentially expressed genes also exhibiting altered chromatin accessibility. In addition, I have identified AP-1 family members, RUNX2, NR1D1 and STAT4, as candidate TFs for regulating MAIT cell phenotype and function. While I was hoping to use TF footprinting to identify the binding sites and potential roles of these TFs, the limitations of the approach made the results from this analysis unreliable. However, I managed to obtain a list of predicted target genes using an alternative machine learning approach known as IMAGE. I plan to experimentally validate the role of some of these TFs, particularly STAT4.

4.9. Summary

- Characterised the chromatin accessibility landscape of $CD8^+$ MAIT cells relative to T_{mem} and naïve T cells.
- Identified AP-1 factors, RUNX2, NR1D1 and STAT4 as potential novel regulators of MAIT cell phenotype.
- Demonstrated the limitations of using ATAC-seq data for TF footprinting with the CENTIPEDE algorithm.

Table 4.1. TFs upregulated in expression in MAIT cells compared with T_{mem} cells.

Gene symbol	Ensembl gene ID	log_2FC	FDR
RORC	*ENSG00000143365*	7.274	7.9E-06
ZBTB16	*ENSG00000109906*	6.541	1.9E-06
CEBPD	*ENSG00000221869*	5.929	1.1E-05
HES4	*ENSG00000188290*	4.990	1.4E-04
NR4A3	*ENSG00000119508*	4.252	7.5E-04
NR1D1	*ENSG00000126368*	3.925	1.9E-06
MYC	*ENSG00000136997*	3.834	1.8E-03
RUNX2	*ENSG00000124813*	3.497	4.1E-04
EPAS1	*ENSG00000116016*	3.459	2.1E-03
GLIS3	*ENSG00000107249*	3.103	7.9E-04
ZNF516	*ENSG00000101493*	3.049	3.0E-04
RARG	*ENSG00000172819*	2.859	1.3E-04
SOX8	*ENSG00000005513*	2.688	1.9E-03
IRF8	*ENSG00000140968*	2.596	8.0E-03
NPAS2	*ENSG00000170485*	2.542	1.5E-04
MAFF	*ENSG00000185022*	2.508	9.9E-05
GATA6	*ENSG00000141448*	2.498	3.2E-03
ZNF165	*ENSG00000197279*	2.489	1.4E-03
FOSL2	*ENSG00000075426*	2.415	6.1E-04
AEBP1	*ENSG00000106624*	2.391	9.3E-04
EGR3	*ENSG00000179388*	2.289	5.3E-03
ZNF331	*ENSG00000130844*	2.149	6.7E-04
NFIX	*ENSG00000008441*	2.148	5.5E-04
SATB1	*ENSG00000182568*	2.138	2.3E-04
CSRNP1	*ENSG00000144655*	2.102	2.0E-04
HLF	*ENSG00000108924*	2.067	1.9E-03
PROX2	*ENSG00000119608*	2.063	1.9E-02
PBX4	*ENSG00000105717*	2.007	1.9E-03
NR4A2	*ENSG00000153234*	1.976	5.0E-04
BACH2	*ENSG00000112182*	1.955	4.0E-04
CREM	*ENSG00000095794*	1.941	7.8E-04
RARA	*ENSG00000131759*	1.857	5.4E-04
ATF5	*ENSG00000169136*	1.798	3.2E-04
HIC1	*ENSG00000177374*	1.700	4.6E-04
BHLHE40	*ENSG00000134107*	1.679	9.7E-03
SREBF1	*ENSG00000072310*	1.678	2.8E-04
ZNF541	*ENSG00000118156*	1.636	2.8E-03
NFKB1	*ENSG00000109320*	1.608	6.3E-03
MAFG	*ENSG00000197063*	1.583	1.7E-04
NFKB2	*ENSG00000077150*	1.575	9.2E-04
BCL6	*ENSG00000113916*	1.559	5.0E-04
CXXC5	*ENSG00000171604*	1.530	4.9E-02
SKIL	*ENSG00000136603*	1.472	3.7E-04
RELB	*ENSG00000104856*	1.442	4.8E-04
NR1D2	*ENSG00000174738*	1.365	4.7E-05
XBP1	*ENSG00000100219*	1.355	1.5E-04
CAMTA1	*ENSG00000171735*	1.349	1.0E-03
ZBED3	*ENSG00000132846*	1.348	8.6E-04

FOS	*ENSG00000170345*	1.307	2.4E-04
MXI1	*ENSG00000119950*	1.284	3.8E-03
ATF3	*ENSG00000162772*	1.253	3.6E-02
FOXO1	*ENSG00000150907*	1.251	4.0E-04
NR4A1	*ENSG00000123358*	1.222	3.6E-02
MAF	*ENSG00000178573*	1.218	3.2E-04
ZNF703	*ENSG00000183779*	1.210	2.5E-02
ZEB1	*ENSG00000148516*	1.202	4.4E-03
ZIK1	*ENSG00000171649*	1.202	1.4E-02
NFIL3	*ENSG00000165030*	1.185	6.9E-03
ZNF629	*ENSG00000102870*	1.142	1.2E-02
ZNF358	*ENSG00000198816*	1.140	1.6E-03
PHF1	*ENSG00000112511*	1.140	7.0E-05
ELK1	*ENSG00000126767*	1.133	2.0E-03
RORA	*ENSG00000069667*	1.115	2.3E-04
KLF16	*ENSG00000129911*	1.099	7.7E-04
HIC2	*ENSG00000169635*	1.087	3.3E-04
ZSCAN18	*ENSG00000121413*	1.083	4.4E-03
FOXJ2	*ENSG00000065970*	1.054	1.6E-04
ZNF771	*ENSG00000179965*	1.049	1.7E-02
CEBPZ	*ENSG00000115816*	1.047	3.4E-04
ZNF576	*ENSG00000124444*	1.021	1.9E-02
ELK3	*ENSG00000111145*	1.017	7.7E-04
NR1H2	*ENSG00000131408*	1.013	6.2E-04

Table 4.2. TFs downregulated in expression in MAIT cells compared with T_{mem} cells.

Gene symbol	Ensembl gene ID	log_2FC	FDR
MEIS1	*ENSG00000143995*	-7.211	4.4E-04
ZNF365	*ENSG00000138311*	-6.508	1.5E-05
BNC2	*ENSG00000173068*	-6.223	4.5E-04
ZNF683	*ENSG00000176083*	-5.623	9.0E-05
PRDM5	*ENSG00000138738*	-5.297	7.9E-04
SETBP1	*ENSG00000152217*	-4.945	4.3E-05
KLF7	*ENSG00000118263*	-4.893	3.4E-05
TSHZ2	*ENSG00000182463*	-4.576	8.2E-03
ZNF433	*ENSG00000197647*	-4.184	1.1E-03
IRF4	*ENSG00000137265*	-4.009	8.2E-05
CXXC4	*ENSG00000168772*	-3.658	4.4E-03
LEF1	*ENSG00000138795*	-3.578	3.4E-05
ZNF80	*ENSG00000174255*	-3.528	4.8E-04
HHEX	*ENSG00000152804*	-3.199	2.6E-03
KLF4	*ENSG00000136826*	-3.105	3.9E-03
ARID5B	*ENSG00000150347*	-3.015	1.5E-04
YBX3	*ENSG00000060138*	-2.884	6.2E-05
ASCL2	*ENSG00000183734*	-2.822	1.4E-03
FOXD1	*ENSG00000251493*	-2.771	7.5E-04
ZNF441	*ENSG00000197044*	-2.763	5.7E-05
TSHZ3	*ENSG00000121297*	-2.717	3.8E-03
ZEB2	*ENSG00000169554*	-2.529	7.2E-04
ZNF660	*ENSG00000144792*	-2.523	1.2E-03
ZNF322	*ENSG00000181315*	-2.506	2.2E-03
TRPS1	*ENSG00000104447*	-2.495	1.4E-04
ZNF442	*ENSG00000198342*	-2.451	3.9E-03
TFCP2L1	*ENSG00000115112*	-2.450	4.1E-02
E2F5	*ENSG00000133740*	-2.442	8.0E-04
ZSCAN2	*ENSG00000176371*	-2.330	9.6E-05
ZNF439	*ENSG00000171291*	-2.298	8.1E-04
JAZF1	*ENSG00000153814*	-2.282	4.2E-04
ZNF563	*ENSG00000188868*	-2.278	1.2E-02
ZNF607	*ENSG00000198182*	-2.240	1.4E-03
PLAG1	*ENSG00000181690*	-2.227	3.5E-04
ZBTB37	*ENSG00000185278*	-2.162	2.9E-04
ZKSCAN7	*ENSG00000196345*	-2.132	3.8E-03
ZFP3	*ENSG00000180787*	-2.087	2.3E-04
ZNF69	*ENSG00000198429*	-2.014	3.6E-03
ZNF594	*ENSG00000180626*	-1.991	4.2E-03
ZNF726	*ENSG00000213967*	-1.939	4.9E-04
ZBTB32	*ENSG00000011590*	-1.936	2.4E-02
RAG1	*ENSG00000166349*	-1.936	2.9E-02
ZNF844	*ENSG00000223547*	-1.915	3.2E-03
ZSCAN20	*ENSG00000121903*	-1.906	2.9E-02
ZNF595	*ENSG00000272602*	-1.891	3.2E-03
ZNF678	*ENSG00000181450*	-1.870	5.9E-04
ZKSCAN3	*ENSG00000189298*	-1.801	6.7E-03
ZNF816	*ENSG00000180257*	-1.756	1.5E-03

ZNF626	*ENSG00000188171*	-1.723	1.8E-03
HOXC4	*ENSG00000198353*	-1.687	2.1E-02
MSC	*ENSG00000178860*	-1.673	1.3E-03
ZNF491	*ENSG00000177599*	-1.639	1.2E-02
SIX5	*ENSG00000177045*	-1.625	1.8E-03
CARF	*ENSG00000138380*	-1.609	8.5E-04
DPF3	*ENSG00000205683*	-1.600	4.4E-03
BATF	*ENSG00000156127*	-1.595	1.1E-03
ZNF618	*ENSG00000157657*	-1.569	2.5E-03
ZSCAN22	*ENSG00000182318*	-1.532	1.1E-02
ZFP69	*ENSG00000187815*	-1.522	7.1E-04
ZNF28	*ENSG00000198538*	-1.522	2.2E-03
IKZF4	*ENSG00000123411*	-1.513	7.6E-03
ZNF85	*ENSG00000105750*	-1.504	1.3E-03
ZNF585B	*ENSG00000245680*	-1.491	1.2E-03
ZBTB3	*ENSG00000185670*	-1.485	6.7E-03
ZNF599	*ENSG00000153896*	-1.484	3.0E-04
ZNF440	*ENSG00000171295*	-1.481	4.8E-03
ZNF559	*ENSG00000188321*	-1.477	2.4E-03
ZNF467	*ENSG00000181444*	-1.466	1.5E-02
ZNF512	*ENSG00000243943*	-1.450	1.1E-03
HIVEP3	*ENSG00000127124*	-1.442	7.5E-04
ZNF175	*ENSG00000105497*	-1.435	7.2E-03
ZSCAN12	*ENSG00000158691*	-1.420	1.0E-03
TFEB	*ENSG00000112561*	-1.389	2.1E-03
ZNF234	*ENSG00000263002*	-1.386	8.7E-03
ZFP82	*ENSG00000181007*	-1.377	6.1E-03
ZNF717	*ENSG00000227124*	-1.355	4.3E-03
HES6	*ENSG00000144485*	-1.354	2.2E-03
YY2	*ENSG00000230797*	-1.350	2.2E-02
ZNF691	*ENSG00000164011*	-1.342	8.8E-03
JRK	*ENSG00000234616*	-1.341	8.7E-04
ZBTB38	*ENSG00000177311*	-1.333	1.4E-04
ZNF619	*ENSG00000177873*	-1.314	2.0E-03
AC023509.3	*ENSG00000267281*	-1.308	5.4E-03
ZNF90	*ENSG00000213988*	-1.305	3.5E-03
ZSCAN10	*ENSG00000130182*	-1.305	2.0E-02
ZNF519	*ENSG00000175322*	-1.300	5.7E-03
TIGD2	*ENSG00000180346*	-1.293	1.4E-02
ZNF397	*ENSG00000186812*	-1.292	8.6E-04
ZNF253	*ENSG00000256771*	-1.277	1.3E-03
ARID3A	*ENSG00000116017*	-1.275	1.7E-03
ZNF792	*ENSG00000180884*	-1.272	2.7E-03
ZNF737	*ENSG00000237440*	-1.266	3.7E-03
SOX13	*ENSG00000143842*	-1.261	2.5E-03
ZNF680	*ENSG00000173041*	-1.244	3.2E-03
ZNF665	*ENSG00000197497*	-1.235	1.3E-02
SETDB2	*ENSG00000136169*	-1.221	8.7E-03
ZNF25	*ENSG00000175395*	-1.214	1.8E-03
KLF10	*ENSG00000155090*	-1.212	2.2E-03
ZNF91	*ENSG00000167232*	-1.203	8.8E-04

ZNF280B	*ENSG00000275004*	-1.203	2.0E-03
ZNF35	*ENSG00000169981*	-1.199	1.3E-02
ZFP14	*ENSG00000142065*	-1.198	3.5E-03
ZNF674	*ENSG00000251192*	-1.175	5.0E-03
ZNF573	*ENSG00000189144*	-1.174	1.4E-02
HOXB4	*ENSG00000182742*	-1.152	5.4E-03
CCDC17	*ENSG00000159588*	-1.147	1.6E-03
ZNF624	*ENSG00000197566*	-1.140	4.5E-03
ZNF70	*ENSG00000187792*	-1.136	1.5E-02
MTERF1	*ENSG00000127989*	-1.129	1.7E-03
ZNF613	*ENSG00000176024*	-1.113	2.4E-02
ZNF257	*ENSG00000197134*	-1.104	2.0E-02
THAP6	*ENSG00000174796*	-1.100	1.6E-03
KLF8	*ENSG00000102349*	-1.089	4.2E-02
IRF2	*ENSG00000168310*	-1.087	2.0E-03
ZNF396	*ENSG00000186496*	-1.084	4.1E-02
ZNF320	*ENSG00000182986*	-1.080	4.4E-02
ZNF718	*ENSG00000250312*	-1.080	8.5E-03
ZNF623	*ENSG00000183309*	-1.077	3.3E-03
ZNF701	*ENSG00000167562*	-1.074	7.0E-03
ZNF600	*ENSG00000189190*	-1.066	7.1E-04
TBX19	*ENSG00000143178*	-1.057	1.8E-03
ZNF507	*ENSG00000168813*	-1.051	1.3E-03
ZNF141	*ENSG00000131127*	-1.048	5.5E-04
ZKSCAN4	*ENSG00000187626*	-1.046	1.9E-02
SETDB1	*ENSG00000143379*	-1.043	1.7E-03
ZNF77	*ENSG00000175691*	-1.042	9.8E-03
ZBTB34	*ENSG00000177125*	-1.038	2.4E-03
ZNF260	*ENSG00000254004*	-1.029	5.2E-03
SOX4	*ENSG00000124766*	-1.016	2.3E-02
ZNF585A	*ENSG00000196967*	-1.012	1.3E-02
ZNF197	*ENSG00000186448*	-1.011	3.3E-03
CREB1	*ENSG00000118260*	-1.009	7.8E-03
ZNF33A	*ENSG00000189180*	-1.007	4.2E-03
ZKSCAN1	*ENSG00000106261*	-1.006	2.2E-03
ZNF793	*ENSG00000188227*	-1.004	5.0E-03

Table 4.3. List of genes that showed both increased expression and increased accessibility in MAIT cells compared with T_{mem} cells (ordered from genes with the largest to the smallest fold change in gene expression). Ensembl gene IDs are listed in brackets where no gene symbol was available.

RBM24	*MPZL2*	*PDE8A*	*ALOX5AP*
SLC4A10	*SPTLC3*	*H1FX*	*RGS2*
ADAM12	*DKK3*	*BACH2*	*LIPC*
MATN2	*GPAT3*	*PCSK5*	*XBP1*
NNMT	*LAYN*	*IVNS1ABP*	*RAB11FIP1*
IL23R	*ZNF516*	*USP6NL*	*CAMTA1*
RORC	*CELSR1*	*TRAF3IP2*	*AGAP1*
PLXDC2	*LAMA2*	*LGALS3*	*IRAK1*
CA2	*CDS1*	*CD28*	*SOBP*
CCL20	*NCR3*	*ADAMTS17*	*FAM241A*
PXDC1	*COL1A1*	*PLCB1*	*BRD1*
LTK	*TNFRSF25*	*RAD54B*	*FOS*
COL5A1	*PRR29*	*RCAN3*	*SLC25A33*
FSBP	*N4BP3*	*NINJ1*	*TRIM47*
MPPED2	*PFKFB3*	*HOOK1*	*FLNB*
MCAM	*CCR2*	*RARA*	*SLC41A2*
ZBTB16	*FAM43A*	*ABCB1*	*GTF2F2*
LINGO4	*COL6A1*	*CHODL*	*CHD1*
ME1	*DUSP16*	*ATF7IP2*	*PLXNC1*
CTNND2	*GAS2L1*	*ERN1*	*MTFP1*
ARL14EPL	*RTN4R*	*SCN1B*	*CMIP*
RASAL1	*SPEG*	*VAV2*	*PATJ*
MAST1	*NEO1*	*PER1*	*ST3GAL4*
P2RY14	*MMP24*	*DUSP2*	*RAB8B*
COL5A3	*CHDH*	*SETD7*	*ZCCHC14*
IL4I1	*SAMD12*	*TRAF4*	*TTC39C*
TRIQK	*TUFT1*	*NCR3LG1*	*MAPKAPK2*
CEBPD	*TSPAN18*	*GALC*	*FOXO1*
WNT11	*LATS2*	*(ENSG00000284526)*	*GCH1*
CCR6	*EDA*	*PRNP*	*SLC7A6*
B3GALT2	*SOX8*	*RETREG1*	*LPCAT3*
TANC1	*SLC16A3*	*MACC1*	*POMGNT2*
TSPAN15	*FGF9*	*ZFP36L2*	*GNG7*
ITPKA	*PMFBP1*	*GPR35*	*ZNRD2*
DAB2IP	*OSM*	*HIC1*	*CABYR*
TPBG	*SLC17A3*	*CDH23*	*WHRN*
ABCA1	*ARHGAP32*	*BHLHE40*	*POLR1E*
IL17RE	*RELT*	*VIPR1*	*MIGA2*
TMIGD2	*LAPTM4B*	*ABTB2*	*MYLIP*
PTPN13	*IRF8*	*IFNGR1*	*MAF*
ESPN	*NRIP3*	*PTTG1IP*	*MAN1A1*
SCRN1	*APOL3*	*ELL*	*PDCL3*
TLE1	*DDIT4*	*CHD7*	*ZNF703*
SCART1	*PLB1*	*GPR150*	*INSR*
CCR1	*NPAS2*	*GPR171*	*SLC2A14*
BLK	*ABCG1*	*NANS*	*ZEB1*

KIT	*USP43*	*ARL4C*	*CXCR4*
AQP3	*CHKA*	*TXK*	*NUDT15*
SCGB3A1	*PLXND1*	*CLIC5*	*WWC2*
KLRB1	*MAFF*	*SLC19A1*	*GADD45G*
COL4A3	*CTSH*	*EML5*	*CASS4*
ROR2	*GATA6*	*DOLPP1*	*EEPD1*
IL12RB2	*JAML*	*GNA11*	*RBM38*
HK2	*PLPP2*	*UBE2S*	*NFIL3*
GREM2	*TNFAIP3*	*NFKB1*	*CYB5A*
IL18RAP	*TGFA*	*RALGPS2*	*SLC37A1*
DAB1	*SSTR3*	*SEC14L1*	*MEGF6*
SUSD4	*NTRK1*	*NRIP1*	*CD2AP*
CD40LG	*SYNJ2*	*SPTY2D1*	*SRSF3*
PARVA	*LONRF3*	*ZFP36L1*	*WDFY2*
RAB6B	*PELI2*	*CAMK4*	*TLE3*
SPON1	*FOSL2*	*IL7R*	*PPP1CB*
FLT4	*GRID2IP*	*EFNB1*	*ERGIC1*
B3GNT8	*IQCA1L*	*USP36*	*LFNG*
GAB1	*WDR86*	*BCL6*	*TIPARP*
NTN4	*RBFOX2*	*CERK*	*LONRF1*
COLQ	*ADGRA3*	*GLRX*	*ODF2L*
FERMT2	*FAM167A*	*PIM2*	*SH3PXD2A*
DTX1	*STARD5*	*CYB561*	*PRR5-ARHGAP8*
DPP4	*LUZP1*	*TSPAN6*	*SVIL*
CPNE7	*SOCS3*	*MAEA*	*DYNLT3*
BRSK1	*SGK1*	*CXXC5*	*PIM1*
SIPA1L2	*EGR3*	*LGALS3BP*	*GRAMD1B*
VLDLR	*PHACTR2*	*FAM184A*	*USP45*
NR1D1	*PDE4D*	*ID2*	*RORA*
CYP2E1	*SMARCD3*	*CPNE2*	*NEDD4L*
ELOVL4	*IL23A*	*ARHGEF7*	*SCARB2*
PRSS22	*MPZL3*	*ARL4A*	*PTP4A1*
CXCR6	*SPOCK2*	*MFSD2A*	*MAPK8IP3*
CPLX1	*CYTH3*	*SIPA1L1*	*MFGE8*
MYC	*ZDBF2*	*PBXIP1*	*MID1IP1*
MED12L	*ANK1*	*MICU3*	*DEGS1*
LST1	*HPGD*	*ATG2A*	*FHL3*
KCNQ3	*TBXAS1*	*CCSAP*	*RPS6KA3*
LRP12	*CR1*	*GPRASP1*	*RRM1*
PROSER2	*SPRED2*	*TOB1*	*RPLP1*
SCD	*GPR146*	*DUSP1*	*CMTM8*
USP2	*CKAP4*	*SPATS2L*	*CTDP1*
BTBD11	*ABCD1*	*SLC12A2*	*KIAA1324*
FURIN	*GPR65*	*SYTL2*	*CSNK1E*
FES	*RASSF2*	*TNFSF14*	*CPD*
ABCG2	*(ENSG00000285476)*	*TSPYL2*	*UBAP1*
RUNX2	*SATB1*	*EML4*	*MGAT5*
SPRY1	*MBOAT1*	*OXNAD1*	*TPT1*
SLC7A5	*LTBP3*	*FLT3LG*	*TMTC2*
KIF5C	*SOAT2*	*DBN1*	*SLC35F2*
SNX9	*SLAMF1*	*BTG1*	*PDCD4*

EPAS1	*PLD6*	*ARHGAP8*	*RAB33B*
AFF3	*PMEPA1*	*ECE1*	*HDAC9*
GNA15	*AREG*	*PABPC1*	*SPTBN1*
SIK1	*HLF*	*SPART*	*CD6*
NEU4	*CEP112*	*ACSL1*	*DDX3X*
SIK1B	*MT3*	*SESN1*	*ELK3*
PLEKHA7	*SH3RF3*	*EXT1*	*TMEM117*
BANK1	*PIANP*	*LSR*	*ZDHHC18*
IRS2	*RTL5*	*NCOA7*	*GNA13*
PLD1	*PBX4*	*TPD52*	*CDK17*
SLC15A3	*RASGRF2*	*FAAH2*	*HSPA9*
THBS1	*TNS4*	*TMEM184B*	
AMPD2	*LRIG1*	*MICAL2*	
GLIS3	*IL18R1*	*B4GALT1*	

Table 4.4. List of genes that showed both decreased expression and decreased accessibility in MAIT cells compared with T_{mem} cells (ordered from genes with the largest to the smallest fold change in gene expression).

SLC26A5	*SELL*	*NUGGC*	*RBM44*
VCAM1	*CDHR3*	*ID3*	*ZDHHC21*
DGKK	*FGL2*	*F2R*	*SWAP70*
NUAK1	*GCNT1*	*KIAA0825*	*ZNF397*
MEIS1	*C1orf21*	*CRTAM*	*CCM2*
ARHGEF28	*CENPF*	*HDDC2*	*C18orf54*
SGCD	*HHEX*	*PKIA*	*CTNNA1*
LRRN1	*MAL*	*PTPN12*	*ADAMTS1*
EFNA5	*GOLM1*	*USP28*	*KIF7*
LILRB1	*DACT1*	*FAM131B*	*IDH2*
BTBD3	*VWA5A*	*FBN1*	*MCF2*
ZNF365	*DST*	*SAMHD1*	*SUSD1*
PRSS23	*RAB27B*	*TJP2*	*DUSP22*
PALLD	*ZC2HC1A*	*STX7*	*ARID3A*
FCGR3A	*KLF4*	*CPNE8*	*MANEA*
RTKN2	*BMF*	*ATP8A2*	*GPR137B*
BNC2	*C5AR2*	*LARGE1*	*ZNF737*
FCRL6	*CTBP2*	*TSPYL5*	*ARHGEF3*
FGFBP2	*PRR5L*	*TCHP*	*THEMIS*
ST8SIA6	*TSPAN2*	*ECRG4*	*ATP7B*
ADGRG3	*NEXMIF*	*C4orf50*	*ETNK1*
MYO3B	*ARID5B*	*HOXC4*	*SERPINI1*
SESTD1	*GATM*	*ATXN7L1*	*KATNAL2*
GPR141	*TTLL7*	*EDARADD*	*ACSL5*
CX3CR1	*EVI5*	*MSC*	*WDR92*
ADGRG1	*GPR153*	*PARD6G*	*OASL*
SLC1A7	*CEP78*	*SHCBP1*	*ITGA4*
CFAP97D2	*CD8B*	*ARHGAP25*	*ADGRG5*
ZNF683	*STK33*	*GRK5*	*GIMAP4*
MGAM	*RGS12*	*CBR4*	*SLC25A25*
NME8	*KLRD1*	*SNX18*	*NATD1*
CMKLR1	*ZC3H12D*	*SIGLEC9*	*REELD1*
TTN	*PTTG1*	*CD84*	*SLC2A9*
DAB2	*AKR1C3*	*MINPP1*	*CD58*
SULT1B1	*MLF1*	*TPST2*	*SETDB2*
PDGFRB	*SLC4A8*	*BPGM*	*BRIP1*
FCRL3	*CCL3*	*HVCN1*	*KLF10*
DSC1	*PLOD1*	*SLC25A42*	*LDAH*
PRDM5	*PPM1L*	*DMXL2*	*KCTD11*
PROK2	*RIC3*	*CENPO*	*B4GALT3*
HLA-DQA1	*FBLN2*	*METTL3*	*STAM2*
RCAN2	*ZNF441*	*PLPP6*	*ZNF91*
FZD6	*CDCA7L*	*BATF*	*PRR11*
B3GAT1	*GFOD1*	*RDH10*	*ZFP14*
GZMH	*TMCC3*	*NUMBL*	*NUDT6*
CCDC141	*CHST2*	*CCL4*	*RNF112*
RHOBTB3	*GJB6*	*FYB1*	*ITK*

MLC1	*BEND5*	*LPAR6*	*MEAK7*
LRFN2	*TNIP3*	*SAMSN1*	*CMTM7*
SETBP1	*ZWINT*	*MBP*	*PRKACB*
MKI67	*TGFBI*	*ZNF618*	*FAM185A*
ENC1	*ZEB2*	*LYPLAL1*	*CTSC*
KLF7	*ZNF660*	*FAM178B*	*STK39*
ADRB1	*OSBPL5*	*RMI1*	*FAM102B*
SH3RF2	*VNN2*	*THNSL2*	*TBCEL*
LAIR2	*TRPS1*	*KBTBD7*	*PTPN18*
FGR	*CDCA7*	*SYCP3*	*MRPS6*
GCSAM	*GAS7*	*SESN3*	*KATNAL1*
NREP	*TFCP2L1*	*LIMD1*	*MMAA*
TNFRSF9	*BCL7A*	*ZNF28*	*ALG10B*
DOCK5	*E2F5*	*DPP8*	*RNF146*
TSHZ2	*ENPP5*	*NAPB*	*SH3TC1*
PRR4	*SPRED1*	*ARSD*	*PLEK*
NCR1	*GPR174*	*GNRH1*	*TRIM32*
HLA-DRB1	*DNAH10*	*MCM3*	*TPBGL*
PTPRO	*CTLA4*	*APOBEC3H*	*FBXO32*
ITGAX	*KLHL29*	*ABHD2*	*IGIP*
KLRC4	*KCNK12*	*C5orf30*	*C17orf67*
TENM1	*RCBTB2*	*FADS2*	*SLC25A15*
KLRC4-KLRK1	*CYSLTR2*	*MCOLN2*	*RBM43*
LGR6	*KCTD1*	*MAP3K20*	*DYRK1B*
APBA2	*LIMA1*	*ZNF440*	*MS4A1*
ARRDC4	*CRPPA*	*SLC4A7*	*BARD1*
CD93	*ZNF439*	*FAR2*	*RESF1*
IFIT3	*CDYL2*	*EPB41L4A*	*LUC7L3*
ME3	*JAZF1*	*FAM126A*	*DISP1*
ST6GALNAC3	*ZNF563*	*CXCR3*	*DPH6*
GZMB	*LRRCC1*	*HIVEP3*	*SH3BP5*
C14orf132	*PLCG2*	*FIG4*	*STMN1*
SNTG2	*TTC24*	*ASXL2*	*TAFA1*
SELP	*CCL4L2*	*RPE*	*NEXN*
ZNF433	*ZNF607*	*PYHIN1*	*CEP57L1*
KRT72	*TPRG1*	*ENG*	*POC5*
HLA-DRA	*SLC35G2*	*YAE1*	*UBE2F*
TIGIT	*TTC16*	*PTPN6*	*TMEM181*
KIAA1671	*PDGFD*	*MFSD4A*	*IRF2*
CLEC11A	*TRIO*	*NOTCH1*	*SNRNP35*
B4GALT6	*IFNLR1*	*DNAI2*	*MSL3*
IRF4	*FRMPD3*	*FBXO30*	*CENPE*
VANGL1	*GNLY*	*ARHGAP18*	*KIF21A*
BIRC7	*CLIC6*	*NINJ2*	*SYNE1*
ITGAM	*BMERB1*	*TM2D1*	*GNS*
PODN	*COL6A3*	*ENPP4*	*TMEM254*
GNAL	*ZBTB37*	*CENPJ*	*DNAL1*
AKAP5	*NUDT11*	*FBXO48*	*MMS22L*
CST3	*IL13*	*AGBL3*	*GGACT*
GPA33	*TIMD4*	*TFEB*	*GLE1*
COLGALT2	*MAMLD1*	*GSAP*	*WNK1*

FMN1	*KLRF1*	*ALDH1B1*	*C9orf40*
CSGALNACT1	*CA13*	*SLC16A7*	*MFSD4B*
PDLIM1	*ACKR3*	*ZFP82*	*FBXO36*
DHCR24	*GTSE1*	*ANXA4*	*SAMD3*
CXXC4	*METTL7A*	*UBXN2B*	*GAREM2*
KALRN	*GCNT2*	*LYST*	*PRKCB*
LEF1	*RDH5*	*CDKL1*	*MTRF1*
CIITA	*ATP10A*	*COTL1*	*SSBP3*
PHGDH	*ITGB1*	*ITGAL*	*SERINC5*
ZNF80	*ZNF69*	*ZNF717*	*TMEM80*
ST8SIA1	*TOX*	*TMPRSS13*	*HIST1H4H*
NIPAL2	*RAB33A*	*RTN4IP1*	*LNPEP*
BFSP1	*SPECC1*	*ACP5*	*ETAA1*
C12orf42	*NCAPG*	*SNAPC3*	*TRIM66*
SLC14A1	*ACVR2A*	*SLFN5*	*KBTBD8*
CD8B2	*RGS1*	*ADIPOR2*	*SEPTIN11*
AK5	*VIPR2*	*SLC30A6*	*KRAS*
MTSS1	*CMC1*	*ZNF619*	*SOX4*
TSPAN33	*ZNF844*	*C2orf76*	*MED15*
SLCO4C1	*ECT2*	*KCNQ5*	*DYRK2*
CD38	*C1orf112*	*FANCL*	*IKBKE*
SDK2	*RAP2A*	*PTER*	*MPP6*
FAM49A	*PRAG1*	*ZNF90*	*CREB1*
HLA-DOA	*KLRK1*	*ZNF519*	*CD47*
TBC1D4	*JAKMIP2*	*VAV3*	*STRADB*
DTHD1	*ZNF678*	*SLF1*	*AFAP1L2*
SOGA1	*GREB1*	*YPEL2*	*CPPED1*
IFNG	*GXYLT1*	*UCP3*	*CHST12*

5. Characterisation of the single-cell transcriptional and TCR repertoire heterogeneity of human MAIT cells in the blood and liver

5.1. Introduction

5.1.1. MAIT cell phenotypic and functional heterogeneity

MAIT cells comprise a relatively homogeneous population with regard to their phenotype and function. They are largely $CD8^+$ and exhibit characteristic expression of surface molecules (e.g. CD161, CD26), cytokine and chemokine receptors (e.g. IL-18R, IL-12R, CCR5, CCR6), and TFs (e.g. T-bet, RORγt, PLZF). Upon activation via their TCR or with cytokines such as IL-12 and IL-18, MAIT cells produce type 1 and type 17 cytokines (e.g. IFNγ, TNFα, IL-17), upregulate cytotoxic molecules (e.g. granzyme B, perforin) and exhibit cytotoxic activity.

Despite this distinctive phenotype, there is evidence for phenotypic and functional heterogeneity within the MAIT cell population. While the MAIT cell surface immunoproteome is relatively homogeneous, a survey of 332 surface proteins identified 21 proteins, enriched for receptors related to innate immunity, with variable expression levels on MAIT cells (Dias et al., 2017). Differential expression

of three of these proteins, namely the NK cell receptors CD56, CD84 and CD94, was associated with altered IFNγ production following cytokine stimulation.

As well as diversity in surface protein expression, MAIT cells show heterogeneity in cytokine production. For example, using mass cytometry, Lepore et al. identified subsets of MAIT cells producing distinct combinations of six cytokines following PMA and ionomycin stimulation (Lepore et al., 2014). Additionally, despite uniformly high expression of the type 17 TF RORγt (Dusseaux et al., 2011; Leeansyah et al., 2015), only a small percentage (< 5%) of MAIT cells produce IL-17 upon TCR or PMA and ionomycin stimulation (Dusseaux et al., 2011; Tang et al., 2013). However, this fraction can be increased by culture with cytokines, such as IL-7 or IL-1β + IL-23 (Tang et al., 2013).

MAIT cell phenotype is partly related to coreceptor expression, although to a much lesser degree than for conventional T cells. MAIT cells are predominantly $CD8^+$ (70-90%), with a smaller subset of DN cells (10-20%) and a minor population of $CD4^+$ cells (~1%) (Gherardin et al., 2018b). Multiple studies have demonstrated an equivalent phenotype and function for $CD8^+$ and DN MAIT cells, including similar expression of cytokine receptors and TFs, and comparable upregulation of cytokines and cytotoxic molecules following activation (Gherardin et al., 2018b; Kurioka et al., 2017). However, a recent study by Dias and colleagues identified a number of distinctions between these populations (Dias et al., 2018). While some of their findings are inconsistent with previous studies, the authors provide compelling evidence for at least some differences between $CD8^+$ and DN MAIT cells. More specifically, compared with $CD8^+$ MAIT cells, DN MAIT cells displayed a distinct transcriptional profile, an enhanced propensity for apoptosis (as previously shown in (Kurioka et al., 2017)), and a more restricted TCR repertoire. As DN MAIT cells

could be derived from $CD8^+$ MAIT cells upon TCR-dependent activation *in vitro*, the authors hypothesised that $CD8^+$ and DN MAIT cells represent distinct states of differentiation.

Due to their rarity, $CD4^+$ MAIT cells are relatively poorly characterised, though some distinguishing properties have been identified. For example, compared with $CD8^+$ and DN MAIT cells, circulating $CD4^+$ MAIT cells show altered expression of TFs (e.g. PLZF, Eomes), chemokine receptors (e.g. CCR4, CXCR6), and NK cell receptors (e.g. NKG2A, NKG2D, CD56), as well as differential cytokine production (e.g. increased IL-2 secretion following PMA and ionomycin stimulation) (Gherardin et al., 2018b; Kurioka et al., 2017; Leeansyah et al., 2015).

MAIT cells show tissue-specific phenotypic and functional features. Gut and liver MAIT cells are transcriptionally distinct from those in the blood, displaying a more activated and tissue-resident gene expression profile (Salou et al., 2018; Slichter et al., 2016; Tang et al., 2013). Consistent with an activated phenotype, liver MAIT cells show increased cell surface expression of activation markers, for example CD38 and HLA-DR, relative to blood MAIT cells (Tang et al., 2013). MAIT cell cytokine production varies between tissues. MAIT cells in the female genital tract are skewed towards type 17 functions compared with blood MAIT cells, exhibiting increased production of IL-17 and IL-22, and decreased production of IFNγ (Gibbs et al., 2016). In adipose tissue, around 15% of MAIT cells can produce IL-10 (Carolan et al., 2015), a cytokine not produced by MAIT cells from other tissues.

5.1.2. MAIT cell TCR repertoire heterogeneity

MAIT cells express a semi-invariant *TRAV1-2*/*TRAJ33* (Vα7.2-Jα33) TCRα chain paired with a TCRβ repertoire biased towards use of *TRBV6* and *TRBV20* (Vβ13 and Vβ2) (Tilloy et al., 1999). Nevertheless, some heterogeneity has been described in the TCR repertoire. In addition to *TRAJ33*, two alternative *TRAJ* segments, *TRAJ12* and *TRAJ20* (Jα12 and Jα20), are used at moderate frequency by MAIT cells (estimated 8-31% *TRAJ12* or *TRAJ20* MAIT cells per donor based on six donors, with a maximum of 43 cells analysed per donor (Reantragoon et al., 2013)), with non-canonical *TRAJ* chains expressed only rarely (Gold et al., 2010; 2014; Lepore et al., 2014; Reantragoon et al., 2013). Aside from the use of different *TRAJ* chains, the MAIT cell TCRα repertoire shows limited variability. The CDR3α region is of essentially fixed length and has a highly conserved sequence, with minor junctional variability (Gherardin et al., 2018b; Held et al., 2015; Porcelli et al., 1993; Reantragoon et al., 2013; Tilloy et al., 1999). *TRBV6* and *TRBV20* are preferentially incorporated into the TCRβ chain, however a wide array of *TRBV* segments are used, resulting in a CDR3β region with variable length and sequence (Gold et al., 2014; Lepore et al., 2014; Porcelli et al., 1993; Reantragoon et al., 2013; Tilloy et al., 1999). Thus, the MAIT cell TCRβ repertoire is much more diverse than the TCRα repertoire.

Increasing evidence suggests that MAIT cell TCR diversity could have functional implications. Use of *TRAJ33* and *TRAJ12* was shown to vary between tissues, suggesting *TRAJ* usage may modify TCR affinity or specificity, although sample size was limited (Lepore et al., 2014). MAIT cells responding to stimulation with different pathogens *in vitro* show distinct TCR repertoires, implying a degree of pathogen or ligand specificity in the TCR (Gold et al., 2014). During human *S*. Paratyphi A

infection, the distribution of MAIT cell clonotypes changes, with some clonotypes expanding and others contracting (Howson et al., 2018). Cell lines transduced with TCRs harbouring the expanded or contracted TCRβ chains paired with a common TCRα chain, show comparatively greater or lesser activation (as measured by CD69 upregulation), respectively, upon stimulation with riboflavin metabolites or bacterial supernatant in the presence of APCs (Howson et al., 2018). This is in line with prior studies showing that MAIT cells with different *TRBV* segments (Dias et al., 2017) or CDR3β sequences (Eckle et al., 2014) exhibit varied activation *in vitro* following stimulation with bacteria or riboflavin ligands, respectively.

In addition to TCR variation within the conventional MAIT cell population, rare MR1/5-OP-RU^{+}Vα7.2^{-} T cells are present in peripheral blood at a frequency of 0.01-0.17% of all T cells (Gherardin et al., 2016; 2018b; Koay et al., 2019b; Meermeier et al., 2016). These cells fall outside my definition of MAIT cells, as described in Chapter 1, which requires expression of a semi-invariant *TRAV1-2* TCR. However, they are important to mention as they will be detected alongside MAIT cells when using the MR1/5-OP-RU tetramer. MR1/5-OP-RU^{+}Vα7.2^{-} T cells have been divided into phenotypically MAIT-like and non-MAIT-like subsets (0.12% and 0.62% of MR1/5-OP-RU tetramer^{+} T cells, respectively) (Koay et al., 2019b). The MAIT-like subset principally recognises MR1/5-OP-RU, while a wider variety of molecules can be detected by the non-MAIT-like subset (in addition to MR1/5-OP-RU), with ligand specificity varying between clones. Example ligands include the folate-derivates MR1/6-FP and MR1/Ac-6-FP that are inhibitory for classical MAIT cells (Gherardin et al., 2016), as well as unidentified molecules from *Streptococcus pyogenes*, a riboflavin auxotroph (Meermeier et al., 2016). There is also a small population MR1-reactive T cells that do not recognise MR1/5-OP-RU (Gherardin et al., 2016; Lepore et al., 2017), however these will not be discussed here.

5.1.3. MAIT cells in the liver

MAIT cells are highly enriched in human liver, comprising 5-50% of liver T cells (Dusseaux et al., 2011; Jeffery et al., 2016; Jo et al., 2014; Tang et al., 2013). Within the liver, they are predominantly localised near bile ducts in the portal tracts, with significantly fewer MAIT cells found in the parenchyma (Jeffery et al., 2016).

Comparison of the expression of 394 immune genes revealed highly similar transcriptional programmes between blood and liver MAIT cells, with only 18 differentially expressed genes (FC ≥ 2) (Tang et al., 2013). These included *IFNG* and *IRF4*, which were upregulated in liver MAIT cells, and *SELL* (encodes CD62L) and *IL8*, which showed increased expression in the blood. Despite the overall similarity of blood and liver MAIT cells, recent RNA-seq data shows that these populations are transcriptionally distinct at a genome-wide level, with liver MAIT cells showing increased expression of genes associated with T cell activation (e.g. *TNF*, *PDCD1*), TCR signalling (e.g. *FOS*, *JUN*), and tissue-residency (e.g. *RUNX3*, *CD69*) (Salou et al., 2018). As mentioned earlier, liver MAIT cells also show increased cell surface expression of the activation markers CD38 and HLA-DR compared with blood MAIT cells (Tang et al., 2013).

There is conflicting evidence as to the relative functional capacity of blood and liver MAIT cells. In one study, liver MAIT cells showed increased production of type 1 cytokines and IL-17A compared with blood MAIT cells following stimulation with PMA and ionomycin (although only the difference in IL-17A production appears significant) (Tang et al., 2013). In a second study, TLR8 stimulation induced a similar fraction of IFNγ-producing MAIT cells in blood and liver (Jo et al., 2014), supporting comparable type 1 cytokine production by blood and liver MAIT cells.

Thus, further evidence is required to establish the extent of functional differences between blood and liver MAIT cells.

5.1.4. Summary and unanswered questions

Overall, heterogeneity can be identified in the phenotype, function and TCR repertoire of human MAIT cells, both within and between different tissues. However, whether such diversity reflects the presence of multiple transcriptionally distinct MAIT cell subsets, is currently unknown. Phenotypic and functional variation across tissues could reflect tissue-specific changes in the activation or differentiation state of the entire MAIT cell population, or the enrichment of functionally distinct MAIT cell subsets in particular tissues. With regard to the TCR repertoire, prior studies have performed paired TCRαβ analysis on only a small number of MAIT cells (< 200 clones (Reantragoon et al., 2013)) and little is known about the sharing of TCRs between blood and different tissues, as well as between donors. Moreover, whether the resting transcriptional profile of individual MAIT cells is related to their specific TCR usage, has not been investigated. Given the particular enrichment of MAIT cells in the liver and our limited understanding of their role at this site, the liver provides an ideal setting in which to address these questions.

5.2. Aim

To characterise heterogeneity in the transcriptional profile and TCR repertoire of MAIT cells in human blood and liver at a single-cell level.

5.3. Key questions

1. How does the transcriptional profile of MAIT cells vary within and between different tissues?
2. How does the TCR repertoire of MAIT cells vary between tissues and individuals?
3. Is the transcriptional profile of MAIT cells associated with their TCR usage?

5.4. Experimental design (Experiment 1)

Combined scRNA-seq and scTCR-seq of sorted MAIT cells from matched human blood (n = 3) and liver (n = 4) using the droplet-based Chromium Single Cell Immune Profiling Solution from 10x Genomics. For two of the donors, blood and liver T_{mem} cells were also analysed.

5.5. Results

5.5.1. Gene expression

MAIT cells are transcriptionally distinct from conventional T_{mem} cells

MAIT cells and T_{mem} cells were sorted as $CD3^+MR1/5\text{-}OP\text{-}RU^{++}$ (± $V\alpha7.2^+$) and $CD3^+MR1/5\text{-}OP\text{-}RU^-CCR7^-$ T cells, respectively, from matched blood and liver (Fig. 5.1A,B). Sort purity was > 99% (Fig. 5.1C) and MR1/6-FP tetramer controls showed low background staining (Fig. 5.1D). Published data (Billerbeck et al., 2010; Salou et al., 2018; Turtle et al., 2011) and my bulk RNA-seq data (as discussed in Chapter 4) show that MAIT cells are transcriptionally distinct from conventional T_{mem} cells. This knowledge could therefore be used to validate my scRNA-seq data.

Clustering analysis was performed on each of the four samples containing a mixture of MAIT cells and T_{mem} cells. MAIT cell clusters were distinct from those of T_{mem} cells both in the blood and the liver (Fig. 5.2A,B). As the majority of MAIT cells use *TRAV1-2*/*TRAJ33* TCRα chains, expression or lack of expression of these TCR segments allowed me to confidently assign clusters to MAIT cells or T_{mem} cells, respectively (Fig. 5.2C,D). Although some T_{mem} cells expressed *TRAV1-2* or *TRAJ33*, coexpression of these segments demarcated the MAIT cell clusters. The distinct clustering of MAIT cells and T_{mem} cells was not driven by differential TCR expression, as genes encoding TCR segments were removed from the gene expression matrix prior to clustering analysis. Cluster assignment was confirmed by the enriched expression of classical MAIT cell markers, including *KLRB1*, *DPP4*, *RORC* and *ZBTB16*, in the MAIT cell clusters relative to the T_{mem} cell clusters (Fig. 5.2E,F).

MAIT cells and T_{mem} cells differentially expressed 107-154 genes in the four samples, with 25 and 29 genes upregulated and downregulated, respectively, in MAIT cells from all samples (Table 5.1). All but one of the upregulated genes and 21/29 of the downregulated genes were differentially expressed between MAIT cells and T_{mem} cells in my bulk RNA-seq data (Table 5.1), providing strong validation for the quality of my scRNA-seq data.

Blood and liver MAIT cells are transcriptionally distinct

Having validated the quality of my scRNA-seq data, I decided to focus on the MAIT cell population, the main subset of interest for this study. I therefore performed a combined analysis of MAIT cells from the seven samples (three blood, four liver). In total, 20,654 good quality MAIT cells were obtained. Clustering analysis of all MAIT cells identified 14 transcriptionally distinct clusters (Fig. 5.3A), which were confirmed

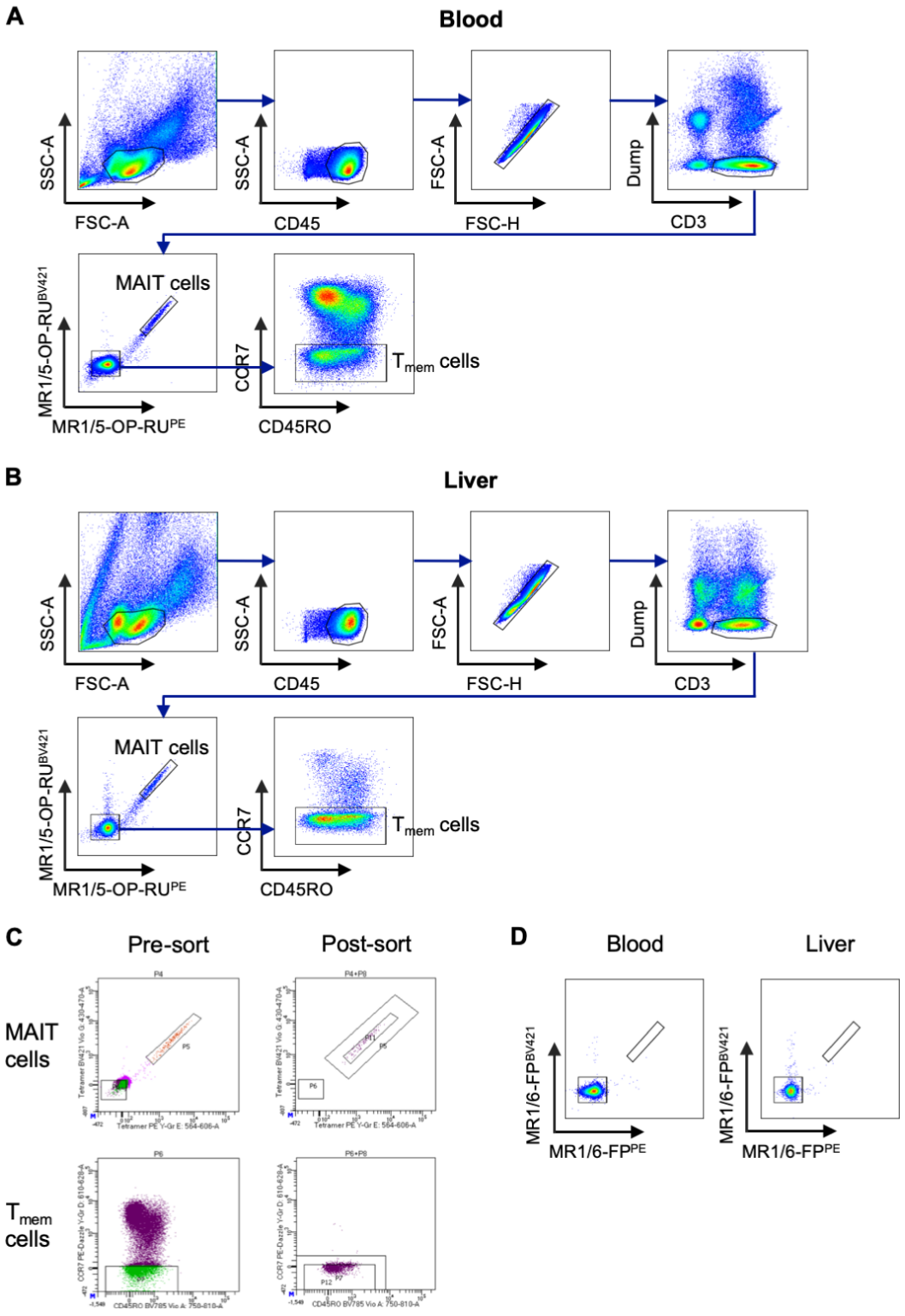


Figure 5.1. Sorting MAIT cells and T_{mem} cells from human blood and liver. (A,B) Representative gating strategy for sorting MAIT cells and T_{mem} cells from matched human blood (A) and liver (B) for scRNA-seq and scTCR-seq. **(C)** FACS plots for a representative sample showing MAIT and T_{mem} cells pre- and post-sorting. Sort purity was > 99%. **(D)** FACS plots for a representative blood and liver sample showing background staining with the control MR1/6-FP tetramer.

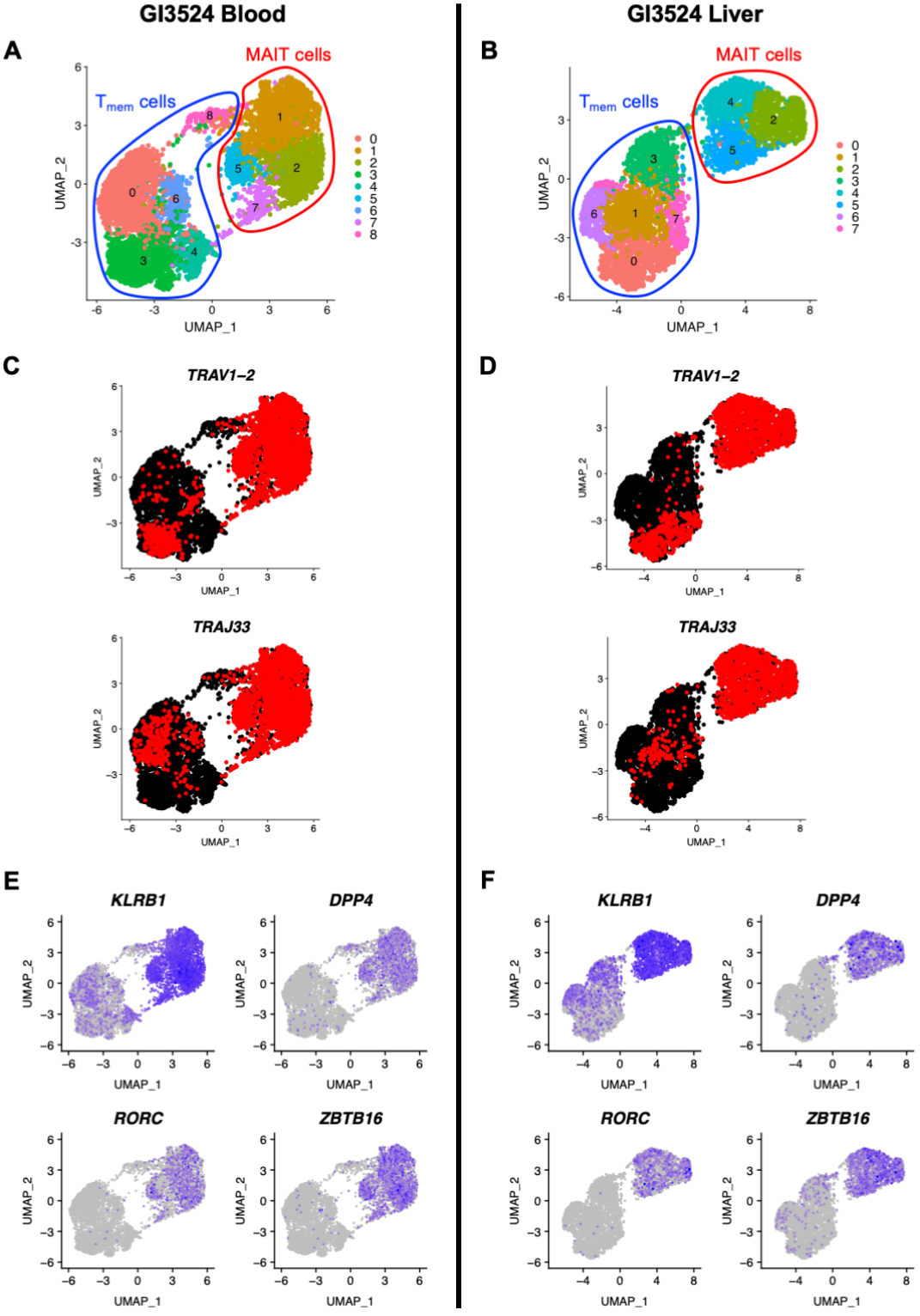


Figure 5.2. MAIT cells and T_{mem} cells are transcriptionally distinct. (A,B) Unsupervised clustering analysis of representative blood (A) and liver (B) samples, visualised by UMAP. **(C,D)** MAIT cell clusters were identified through their predominant use of a *TRAV1-2*/*TRAJ33* TCRα chain. MAIT cell clusters are 1, 2, 5 and 7 in GI3524 Blood, and 2, 4 and 5 in GI3524 Liver. **(E,F)** Enriched expression of *KLRB1*, *DPP4*, *RORC* and *ZBTB16* validates the classification of MAIT cell clusters.

to be MAIT cells by their predominant use of a *TRAV1-2*/*TRAJ33* TCRα chain and expression of classic MAIT cell markers (e.g. *KLRB1* and *DPP4*) (Fig. 5.3B).

Blood and liver MAIT cells occupied different regions of the UMAP (Fig. 5.3C), with individual clusters comprised predominantly of cells from a single tissue site (Fig. 5.3D). Differential expression analysis between blood and liver MAIT cells identified 186 differentially expressed genes, with the majority of these genes (136) upregulated in the liver (Fig. 5.3E; Table 5.2). Upregulated genes encoded tissue-residency markers (e.g. *CD69*, *PRDM1*), TCR-induced TFs (e.g. AP-1 family members, *REL*), effector molecules (e.g. *IFNG*, *TNF*, *PRF1*), chemokines and chemokine receptors (e.g. *CCL3*, *CCL4*, *CXCR6*), anti-proliferative molecules (e.g. *BTG1*, *BTG2*) and heat-shock proteins (e.g. *HSPA1A*, *HSPA1B*, *HSPA8*). Whilst some genes such as *CD69* and *CXCR6* showed uniformly higher expression in liver compared with blood MAIT cells, other genes such as *GNLY* and *CCL3* were enriched in specific blood or liver clusters (Fig. 5.3F).

To establish whether liver MAIT cells showed a more tissue-resident gene expression profile compared with blood MAIT cells, I performed GSEA using a mouse T_{RM} cell gene signature published by Milner and colleagues (Milner et al., 2017). The gene signature comprises 121 genes upregulated in $CD8^+$ T_{RM} cells from multiple tissues (small intestinal epithelium, kidney, skin, lung, brain) compared with circulating splenic memory T cells. Liver MAIT cells showed highly enriched expression of the T_{RM} gene signature (Fig. 5.4A), indicating a tissue-resident phenotype. The 75 leading edge genes included *CD69*, *TNF*, *HAVCR2* (encodes TIM-3), *ICOS* and *FOS*. Enrichment of tissue-residency genes in liver MAIT cells was confirmed using a second T_{RM} gene signature (Fig. 5.4B) (Beura et al., 2019). This was derived from mouse $CD4^+$ T_{RM} cells as opposed to $CD8^+$ T_{RM} cells and

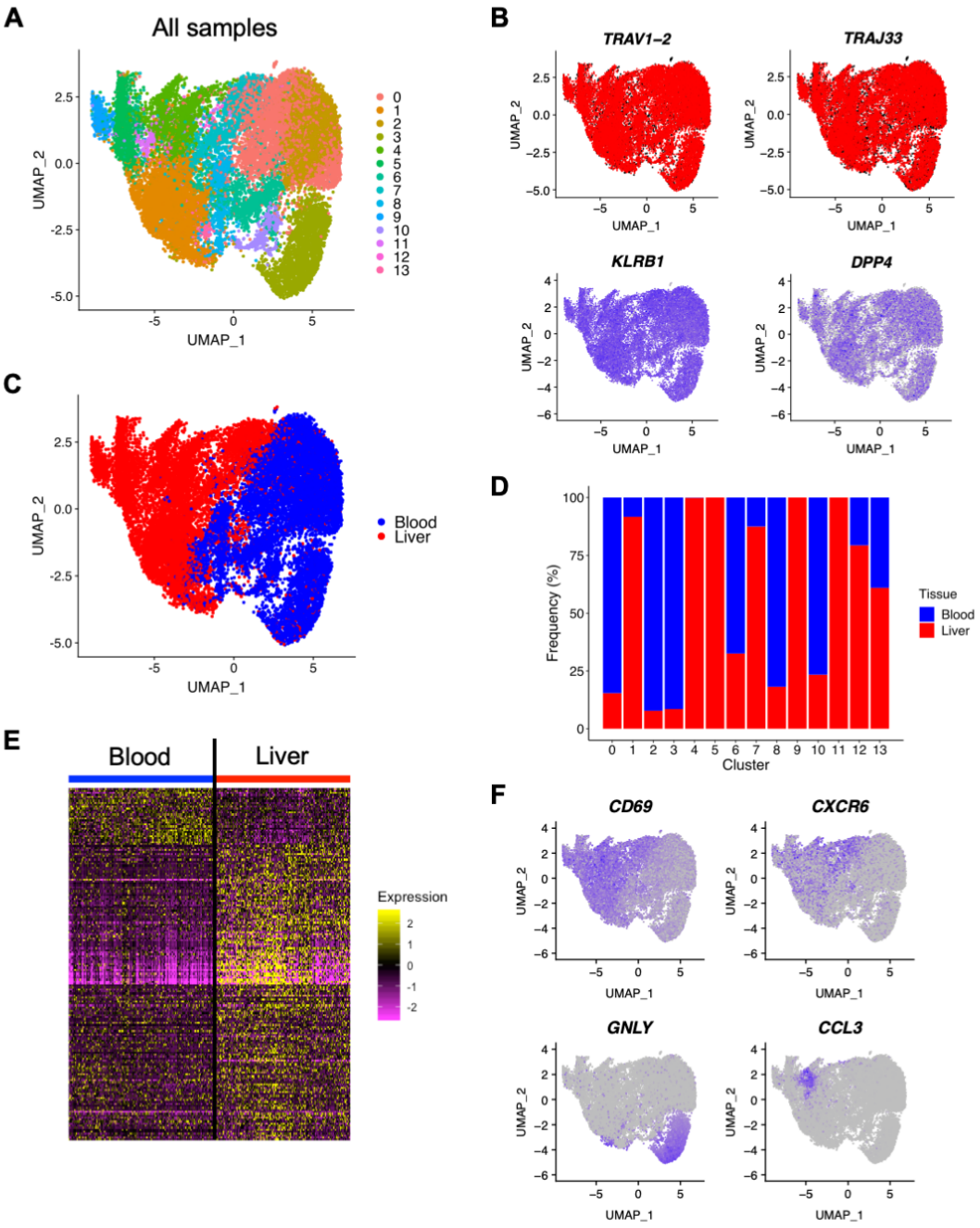


Figure 5.3. MAIT cells in the blood and liver are transcriptionally distinct. **(A)** Clustering analysis of MAIT cells from all samples, as visualised by UMAP. **(B)** UMAP plots showing expression of the TCR gene segments *TRAV1-2* and *TRAJ33* (top) and the genes *KLRB1* and *DPP4* (bottom). **(C)** Localisation of blood (blue) and liver (red) MAIT cells on the UMAP shown in (A). **(D)** Bar plot displaying the percent contribution of blood and liver MAIT cells to each of the 14 MAIT cell clusters in (A). **(E)** Heatmap of the genes differentially expressed between blood and liver MAIT cells. **(F)** UMAP plots showing the expression of example genes upregulated in liver (*TNF*, *CCL4*, *CD69*) or blood (*GNLY*) MAIT cells.

included 648 genes upregulated in $CD4^+$ T_{RM} cells from multiple tissues (female reproductive tract, small intestinal epithelium, small intestinal lamina propria) relative to splenic $CD4^+$ T_{CM} and T_{EM} cells. $CD4^+$ T_{RM} and $CD8^+$ T_{RM} cells show overlapping gene signatures (Beura et al., 2019). Given that both blood and liver MAIT cells predominantly express CD8, enrichment of the $CD4^+$ T_{RM} gene signature in liver compared with blood MAIT cells suggests that the tissue-residency characteristics of liver MAIT cells are not specific to $CD8^+$ T cells.

To gain further insight into the genes upregulated in liver MAIT cells compared with blood MAIT cells, I performed pathway enrichment analysis. The most upregulated pathways were related to interleukin signalling and the cellular stress response (Fig. 5.4C). This suggests that MAIT cells in the liver are more activated than those in the blood, probably due to the continual exposure of liver cells to microbial- and food-derived products from the gut.

MAIT cell clusters within blood or liver

MAIT cells comprise a variable number of clusters within individual samples

As I was interested in exploring the relationship between the clonal origin and transcriptional profile of MAIT cells, it was necessary to also analyse each of the samples individually. Moreover, given the variable contribution of each donor to the 14 clusters of MAIT cells (Fig. 5.5A), I reasoned that separate analysis would allow me to better characterise the extent of MAIT cell heterogeneity within each sample.

A variable number of MAIT cell clusters were identified per sample, ranging from four to eight (Fig. 5.5B). For matched blood and liver samples, the number of

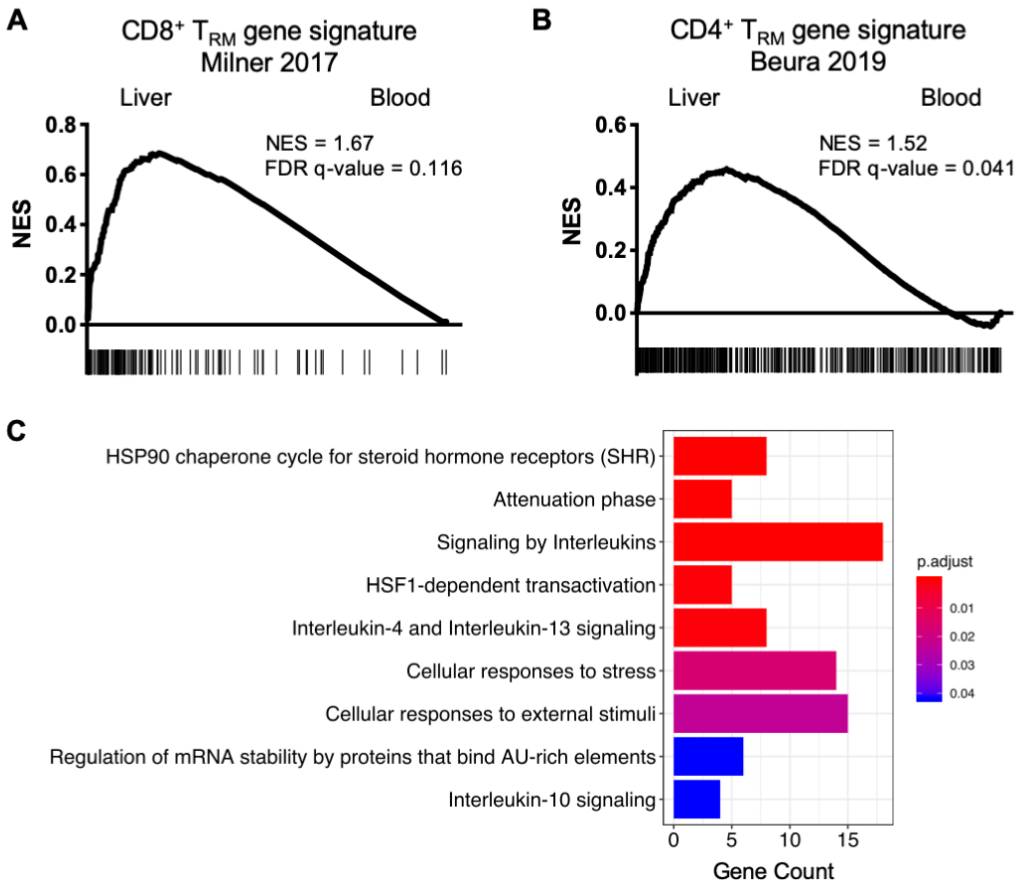


Figure 5.4. Liver MAIT cells exhibit an activated and tissue-resident phenotype. (A,B) GSEA on liver compared with blood MAIT cells using the T_{RM} gene signatures from (Milner et al., 2017) (A) and (Beura et al., 2019) (B). NES = normalised enrichment score. **(C)** Reactome pathway enrichment analysis of the genes upregulated in liver MAIT cells compared with blood MAIT cells.

clusters in the liver was always equal to or more than the number of clusters in the blood.

MAIT cell subsets could not be distinguished by coreceptor expression

Despite recently reported transcriptional differences (Dias et al., 2018), *CD8*$^+$ and DN MAIT cells did not cluster separately, either within individual samples (Fig. 5.6A,B) or in the combined analysis of all samples (Fig. 5.6C). Moreover, the fractions of DN, *CD8A*$^+$*CD8B*$^+$ and *CD8A*$^+$ cells were largely comparable across the 14 clusters of MAIT cells from the blood and liver (Fig. 5.6D). DN MAIT cells were defined as cells that lacked expression of *CD4*, *CD8A* and *CD8B*. This definition is imprecise and likely overestimated the fraction of DN cells, as cells in which neither *CD4* nor *CD8A*/*CD8B* expression was detected could have been bona fide DN MAIT cells, or CD8$^+$/CD4$^+$ MAIT cells for which coreceptor expression was not detected for technical reasons. Consistent with this hypothesis, 13.8% of all cells from the seven samples were defined as DN by scRNA-seq, while prior FACS data from two of the donors suggested a DN frequency of 2.7-6.3% (data not shown). Despite this limitation of the scRNA-seq data, if DN MAIT cells were a fundamentally distinct population from CD8$^+$ MAIT cells, I would have expected to observe a distinct cluster that was entirely lacking in *CD4* and *CD8A*/*CD8B* expression. Differential expression analysis between DN and *CD8A*$^+$*CD8B*$^+$/*CD8A*$^+$ MAIT cells identified five differentially expressed genes in the blood and eight in the liver (Table 5.3 and 5.4), of which only *CD8A* and *CD8B* (downregulated in blood and liver DN MAIT cells) were consistent with published bulk RNA-seq data (Dias et al., 2018).

CD4$^+$ MAIT cells were present at a low frequency, comprising 161 cells in total and 11-54 cells per donor (as measured by scRNA-seq). These *CD4*$^+$ MAIT cells did not form distinct clusters and were spread throughout the MAIT cell population (Fig.

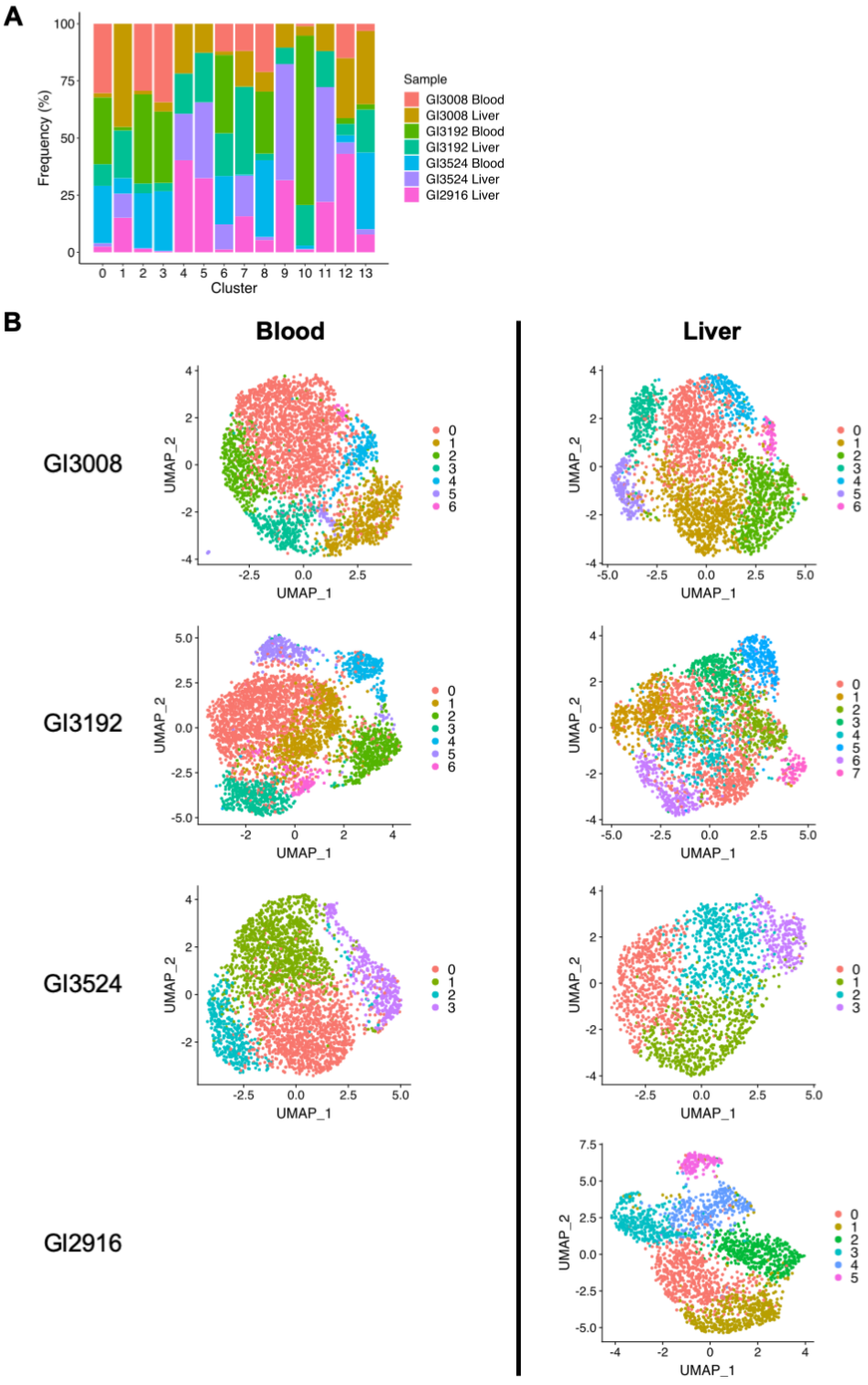


Figure 5.5. MAIT cells comprise a variable number of clusters in the blood and liver. **(A)** Bar chart showing the percent contribution of each sample to the 14 clusters of MAIT cells from the blood and liver. **(B)** Clustering analysis of MAIT cells from each sample, as visualised by UMAP.

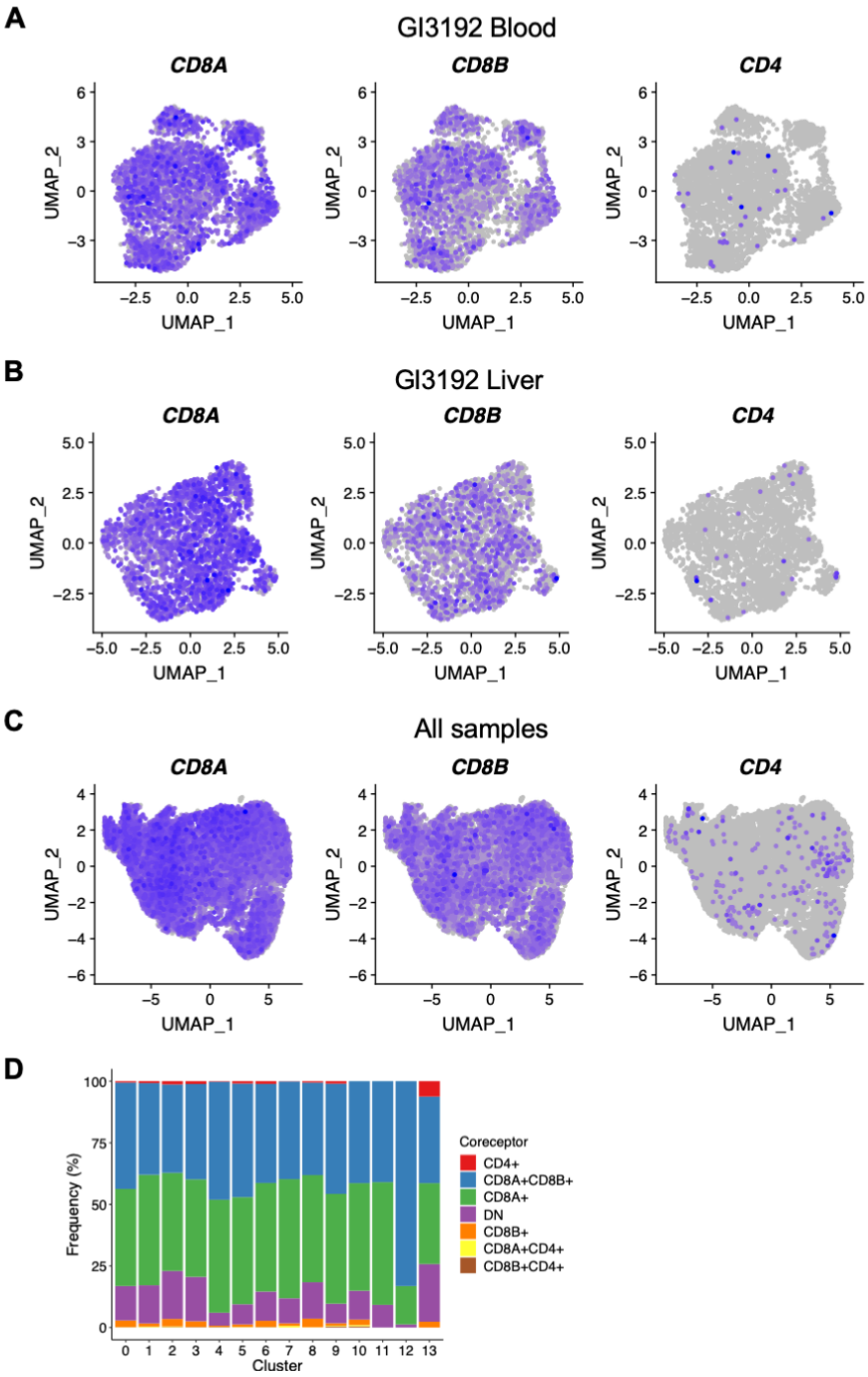


Figure 5.6. MAIT cell subsets are not defined by coreceptor expression. (A-C) UMAP plots of representative blood (A) and liver (B) samples, or all samples (C), showing the expression of *CD8A*, *CD8B* and *CD4*. **(D)** Bar graph showing the percentage of cells expressing the indicated coreceptor(s) within each of the 14 clusters of MAIT cells from the blood and liver. DN = negative for *CD8A*, *CD8B* and *CD4*.

5.6A-D). Nevertheless, 30 and 41 genes were differentially expressed between *CD4*$^+$ and *CD8A*$^+$*CD8B*$^+$/*CD8A*$^+$ MAIT cells in the blood and liver, respectively (Table 5.5 and 5.6). Of these, 22 genes were differentially expressed in both the blood and liver, including *CCL5*, *KLRB1* and *PRF1*. At the protein level, circulating CD4$^+$ MAIT cells show reduced expression of several TFs, including PLZF (*ZBTB16*) and Eomes, and altered expression of surface molecules, such as CCR4, CD56 and NKG2A (*KLRC1*), relative to CD8$^+$ and DN MAIT cells (Gherardin et al., 2018b; Kurioka et al., 2017; Leeansyah et al., 2015). Of these, *ZBTB16* and *KLRC1* were differentially expressed at the mRNA level in the liver but not in the blood (Table 5.5 and 5.6).

Overall, relatively few genes were differentially expressed between DN and *CD8*$^+$ MAIT cells, and between *CD4*$^+$ and *CD8*$^+$ MAIT cells. Furthermore, *CD8*$^+$, DN and *CD4*$^+$ MAIT cells did not comprise transcriptionally distinct subsets at a genome-wide level. These preliminary findings need to be confirmed in studies with better methods for DN MAIT cell identification and with samples comprising a larger number of CD4$^+$ MAIT cells.

Cluster markers are associated with cell signalling and activation

Cluster markers were defined as genes differentially expressed between each cluster and all other cells. Clusters showed modest transcriptional differences, with a median of seven genes upregulated per cluster in the blood and 30 in the liver. Cluster markers were similar across donors and tissues and included activation markers (e.g. *CD69*, *CD40LG*), chemokines (e.g. *CCL3*, *CCL4*, *CCL20*), cytokines (e.g. *IFNG*, *TNF*), cytotoxic molecules (e.g. *GZMB*, *GNLY*), TCR-induced TFs (e.g. AP-1 family members) and NK cell-associated molecules (e.g. *KLRD1*, *NKG7*). In some instances, expression of a marker was essentially unique to a particular

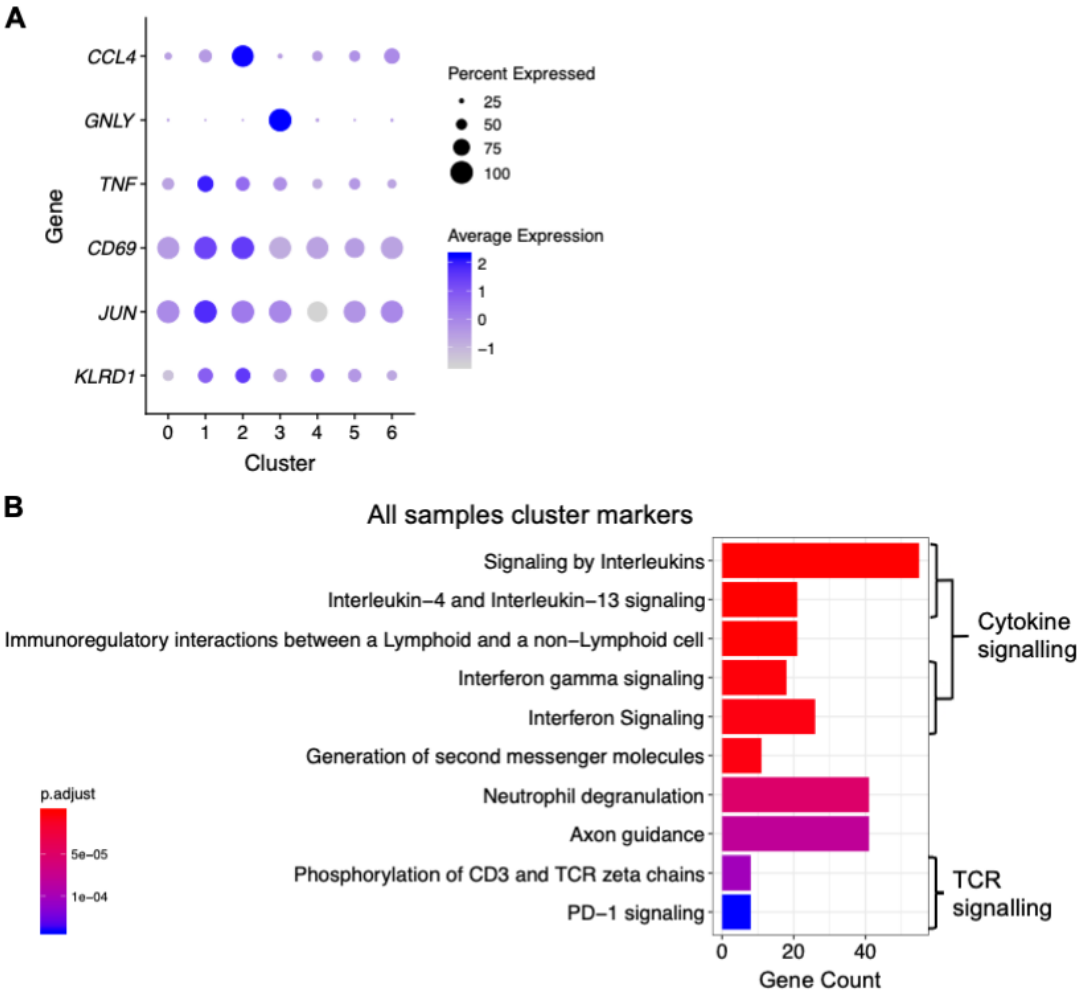
cluster (e.g. *CCL4* in Fig. 5.7A), but in the majority of cases variation in marker expression between clusters was more subtle (e.g. *CD69* in Fig. 5.7A).

To further characterise the cluster markers, I performed pathway enrichment analysis on the cluster markers combined from all samples (460 unique genes; Table 5.7). The top 10 most upregulated Reactome pathways were predominantly related to cytokine and TCR signalling, as well as cell-cell interaction (Fig. 5.7B). Likewise, enriched gene ontology (GO) terms were associated with the cellular response to stimulation, signalling and cell activation (Fig. 5.7C). Given that blood and liver MAIT cells were transcriptionally distinct (Fig. 5.3 and 5.4), I wanted to establish whether the pathways differentially expressed between clusters were similar in the two tissues. Consistent with the merged analysis, separate pathway enrichment analysis of blood (114 genes) and liver (398 genes) cluster markers confirmed enrichment for cytokine and TCR signalling pathways in both tissues (Fig. 5.8A,B), although cytokine pathways were dominant in the liver.

To confirm that cluster markers were specifically associated with the activation of MAIT cells, cluster markers were compared with the genes differentially expressed between stimulated and unstimulated MAIT cells (Fig. 5.9A). Lists of differentially expressed genes were obtained from three published RNA-seq datasets in which MAIT cells were activated under various conditions (Hinks et al., 2019; Lamichhane et al., 2019b; Leng et al., 2019).

The stimulation conditions from all papers were grouped into the following four categories:

1. 6 h TCR stimulation – 5-A-RU stimulation in (Lamichhane et al., 2019b) and 5-OP-RU stimulation in (Hinks et al., 2019)



C

Top 15 enriched GO terms	p value
1. Immune system process	3.97×10^{-56}
2. Immune response	1.96×10^{-50}
3. Response to stimulus	1.64×10^{-46}
4. Cell activation	2.23×10^{-39}
5. Regulation of immune system process	1.34×10^{-38}
6. Response to organic substance	2.09×10^{-38}
7. Cellular response to cytokine stimulus	3.99×10^{-38}
8. Cell surface receptor signalling pathway	3.16×10^{-37}
9. Cytokine-mediated signalling pathway	9.46×10^{-37}
10. Leukocyte activation	5.11×10^{-36}
11. Response to cytokine	6.94×10^{-36}
12. Response to stress	9.92×10^{-36}
13. Response to chemical	1.05×10^{-35}
14. Defence response	1.51×10^{-35}
15. Signal transduction	1.75×10^{-35}

Figure 5.7. MAIT cell clusters differentially express pathways associated with cell signalling and activation. (A) Dot plot of a representative sample (GI3008 Liver) showing the expression of example genes with variable expression between clusters. **(B)** Reactome pathway enrichment analysis of the cluster markers combined from all samples (top 10 enriched pathways are shown). **(C)** Top 15 enriched GO terms amongst the cluster markers from all samples.

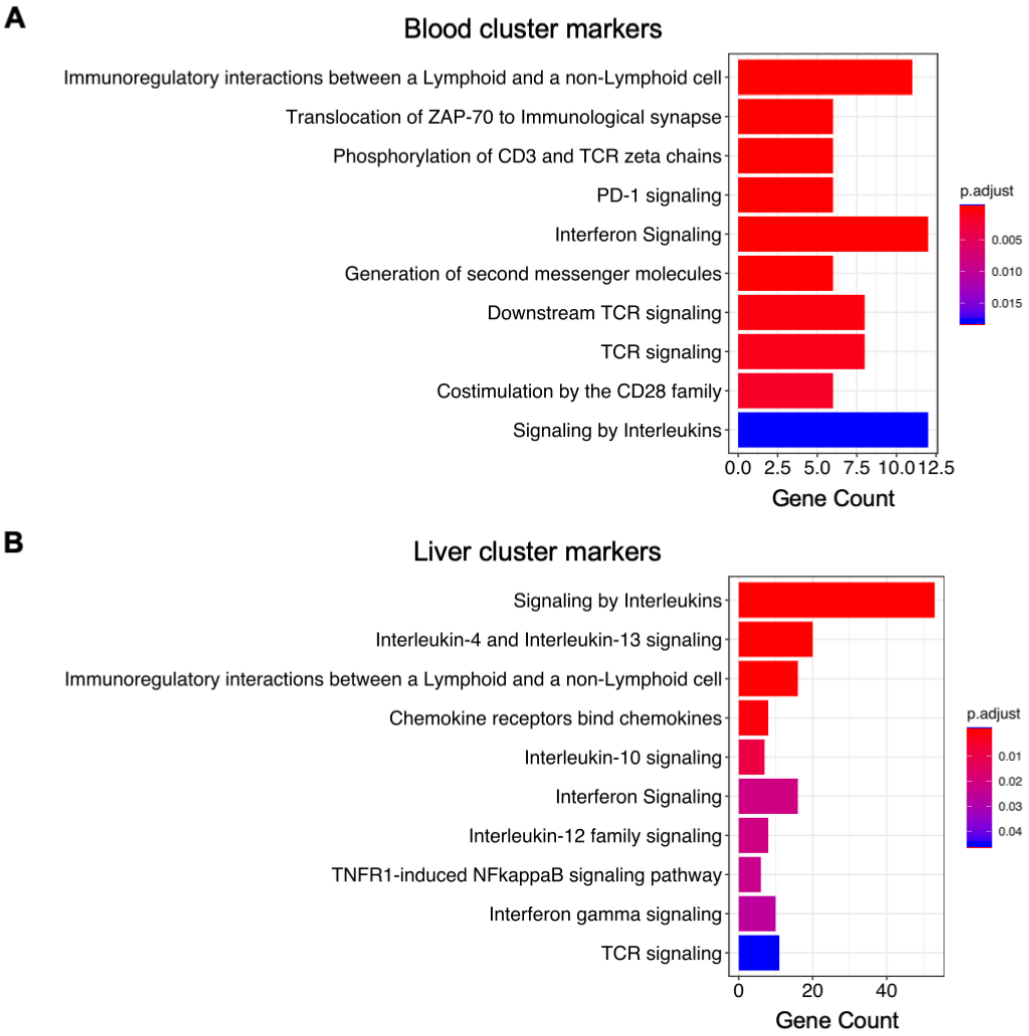


Figure 5.8. Blood cluster markers are enriched for both TCR and cytokine signalling pathways, while cytokine signalling pathways are predominantly enriched in the liver. **(A,B)** Reactome pathway enrichment analysis of the cluster markers combined from blood (A) or liver (B) samples. Plots show selected enriched pathways.

2. 24 h TCR stimulation – αCD3αCD28 beads in (Leng et al., 2019)
3. *E. coli* stimulation – 24 h in (Lamichhane et al., 2019b)
4. Cytokine stimulation – 24 h with either IL-12 + IL-18 in (Lamichhane et al., 2019b) or IL-12 + IL-18 + IL-15 + TL1A in (Leng et al., 2019)

In total, 63% of cluster markers (289/460 genes) were differentially expressed in at least one of the activation conditions (Fig. 5.9B). Of the activation-associated cluster markers, 84% were differentially expressed in response to TCR stimulation (6 h or 24 h), 42% were modulated by *E. coli* stimulation and 50% by cytokine stimulation, while 14% were differentially expressed in all conditions (Fig. 5.9C; Table 5.7). This suggests that variation in gene expression between MAIT cell clusters is largely driven by differential activation at the time of sample collection. Both TCR and cytokine signals appear to contribute to this, although pathway enrichment analysis suggests a dominant role for cytokine signals in the liver (Fig. 5.8B).

Summary

Overall, MAIT cells in the liver showed an activated and tissue-resident gene expression profile relative to blood MAIT cells. Combined analysis of all blood and liver samples identified 14 transcriptionally distinct MAIT cell clusters, while individual samples comprised 4-8 clusters. *CD8*$^+$, DN and *CD4*$^+$ MAIT cells were distributed throughout the clusters, suggesting that coreceptor expression does not define transcriptionally distinct subsets of MAIT cells. Cluster markers were enriched for genes associated with both TCR- and cytokine-mediated activation. Therefore, MAIT cell clusters may represent different states of activation, rather than truly distinct MAIT cell subsets.

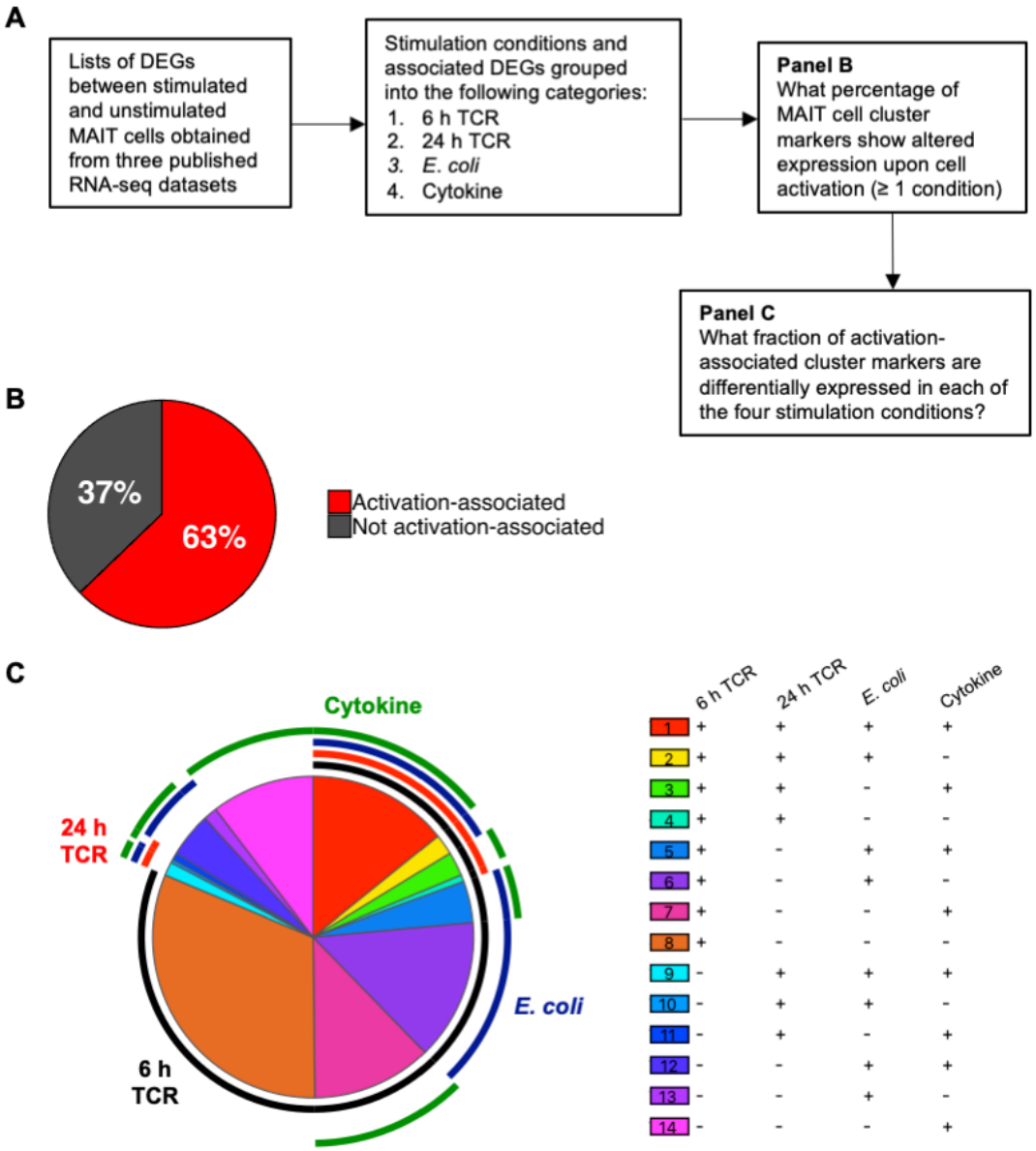


Figure 5.9. Cluster markers are associated with MAIT cell activation. (A) Flow chart describing the comparison of MAIT cell cluster markers with the genes differentially expressed between stimulated and unstimulated MAIT cells, obtained from three published bulk RNA-seq datasets (Hinks et al., 2019; Lamichhane et al., 2019b; Leng et al., 2019). Stimulation conditions were grouped into four categories: 6 h TCR stimulation, 24 h TCR stimulation, *E. coli* stimulation and cytokine stimulation. DEGs = differentially expressed genes. **(B)** Pie chart showing the fraction of cluster markers that were differentially expressed in at least one stimulation condition ("activation-associated"). **(C)** Pie chart showing the fraction of activation-associated cluster markers that were differentially expressed in different permutations of the stimulation conditions. The arcs around the pie chart demonstrate the total fraction of activation-associated cluster markers that were differentially expressed following 6 h TCR stimulation (81%), 24 h TCR stimulation (22%), *E. coli* stimulation (42%) or cytokine stimulation (50%).

5.5.2. TCR usage

I next turned to the TCR data to address my second key question, namely how does the TCR repertoire of MAIT cells vary between tissues and individuals. While published studies have characterised the general features of the MAIT cell TCR repertoire, data was either generated from bulk sorted MAIT cells and thus lacked paired TCRαβ data, or was obtained from single-cell sequencing of less than 200 MAIT cells (Gherardin et al., 2018b; Held et al., 2015; Lepore et al., 2014; Reantragoon et al., 2013). As far as I am aware, my dataset of over 20,000 MAIT cells with scTCR-seq data is the largest generated so far, and thus provides a unique opportunity to describe the MAIT cell TCR repertoire in detail.

The oligoclonality of MAIT cells is comparable with that of T_{mem} cells

Nucleotide clonotypes were defined as cells with identical CDR3α and CDR3β nucleotide sequences. When a cell expressed two TCRα and/or two TCRβ chains, matching of all CDR3 sequences was required. MAIT cells were highly oligoclonal (Fig. 5.10A). Oligoclonality was largely comparable across tissues and donors, with the highest similarity observed for matched tissue MAIT cells. The degree of MAIT cell oligoclonality was similar to that of T_{mem} cells (Fig. 5.10B). MAIT cell clonotypes defined using only the CDR3α nucleotide sequence (TCRα only clonotypes; Fig. 5.10C) were more oligoclonal than those defined using only the CDR3β nucleotide sequence (TCRβ only clonotypes; Fig. 5.10D). Namely, to account for the TCRs expressed by 80% of cells, a significantly higher percentage of TCRβ only clonotypes were required relative to TCRα only clonotypes (mean 24.7% and 12.3% for TCRβ only and TCRα only clonotypes, respectively; Fig. 5.10E).

The MAIT cell TCRα chain was highly conserved across donors and tissues. Across all donors and samples, *TRAJ33*, *TRAJ12* and *TRAJ20* were used by an average of 90%, 5% and 3%, respectively, of *TRAV1-2* MAIT cell TCRs (Fig. 5.11A). All other *TRAJ* segments (26 in total) were each used at an average frequency of < 0.5%. *TRAJ* gene usage showed no significant difference between blood and liver (Fig. 5.11B).

Essentially all *TRAV1-2* CDR3α sequences were 12 amino acids in length (Fig. 5.11C) and the CDR3α sequence was highly conserved across donors and tissues (Fig. 5.11D). All *TRAJ33*, *TRAJ12* and *TRAJ20* CDR3α sequences included the conserved tyrosine 95 residue (position 8 of the CDR3α) (Fig. 5.11E), which forms a hydrogen bond with the ribityl moiety of 5-OP-RU and is required for MAIT cell activation (Corbett et al., 2014; Eckle et al., 2014; Patel et al., 2013; Reantragoon et al., 2012). IMGT/JunctionAnalysis (Giudicelli and Lefranc, 2011; Monod et al., 2004) of 1,000 CDR3α nucleotide sequences from *TRAV1-2*/*TRAJ33* MAIT cell TCRs, and all *TRAV1-2*/*TRAJ12* (876 sequences) and *TRAV1-2*/*TRAJ20* (469 sequences) TCRs, showed that the majority of residues in the CDR3α region were germline-encoded (data not shown). Amino acid usage at position 11 and 12 of the CDR3α region differed depending on *TRAJ* segment usage (Fig. 5.11E). While *TRAJ33* and *TRAJ12* CDR3α sequences showed some sequence variation at positions 3 and 4, positions 4-6 were variable in *TRAJ20* sequences (Fig. 5.11E), although certain amino acids were still favoured at these sites. This variability was due to limited P and N nucleotide additions.

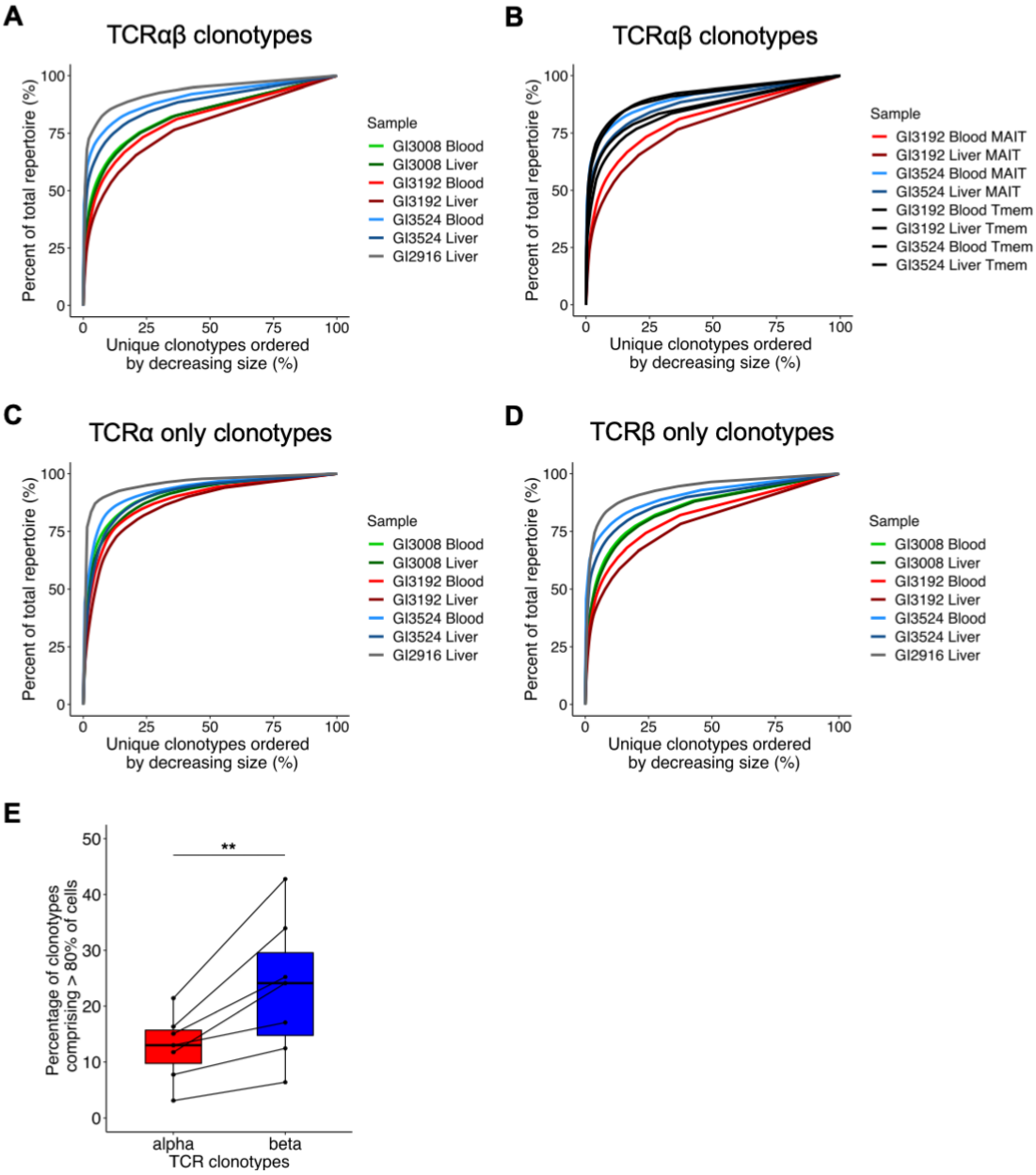


Figure 5.10. MAIT cells are highly oligoclonal in blood and liver. (A-D) Line graphs illustrating the oligoclonality of the MAIT cell (A,C,D) and T_{mem} cell (B) TCR repertoire in blood and liver. Clonotypes were defined using both the CDR3α and CDR3β nucleotide sequences (A,B), the CDR3α sequence only (C), or the CDR3β sequence only (D). **(E)** Box plot showing the percentage of unique TCRα only (red) or TCRβ only (blue) clonotypes comprising > 80% of cells in each sample. Lines connect data from a single individual. Two-tailed paired sample t-test. ** $p < 0.01$.

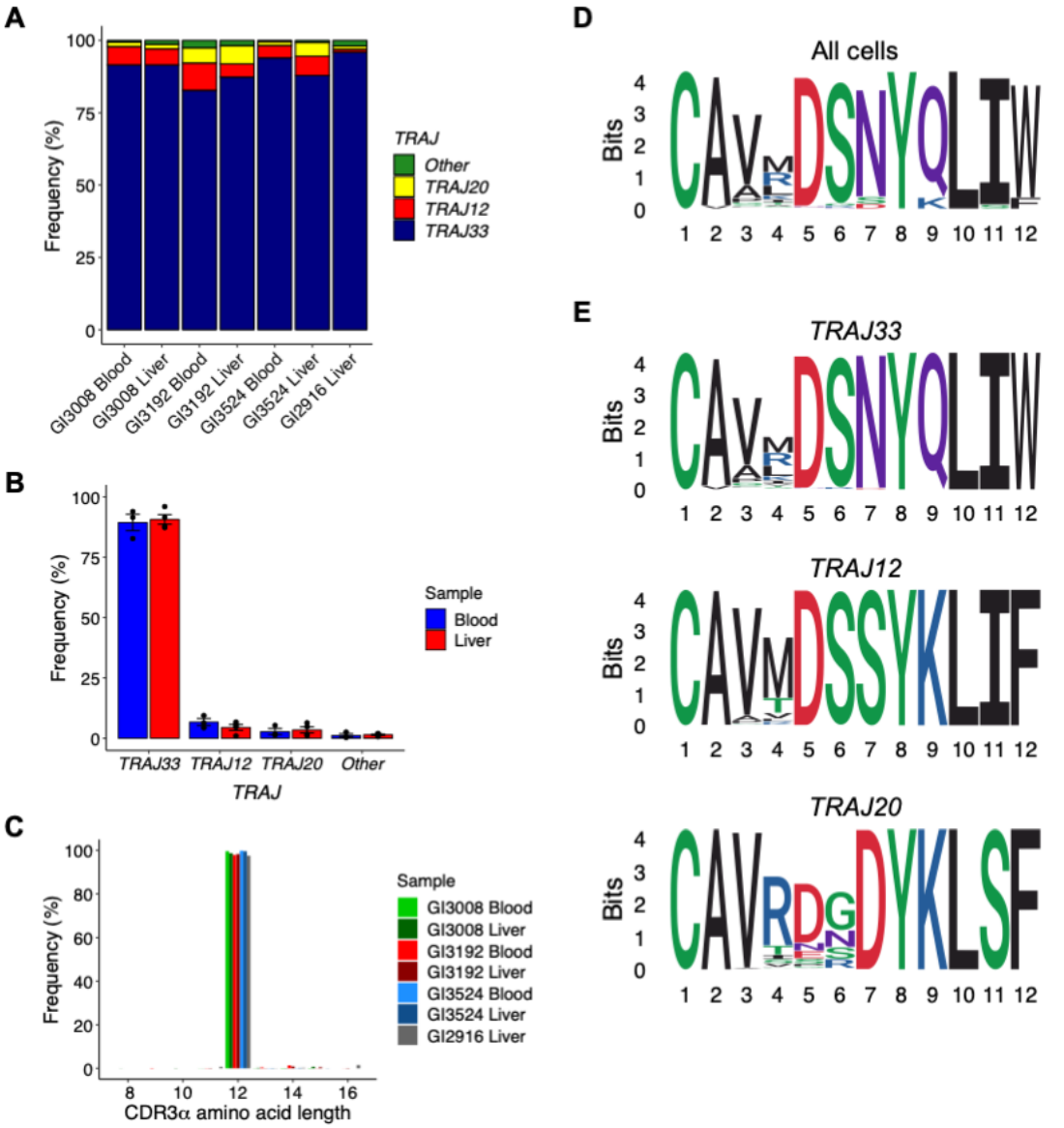


Figure 5.11. *TRAJ* gene usage by, and CDR3α amino acid length and sequence of, *TRAV1-2* MAIT cells in blood and liver. (A) Percentage use of *TRAJ33*, *TRAJ12*, *TRAJ20* and other *TRAJ* segments amongst *TRAV1-2* TCRs in each sample. **(B)** Bar chart showing the average percentage of MAIT cells from the blood and liver using each *TRAJ* segment. *TRAJ* usage showed no significant difference between blood and liver (comparison of frequency in blood and liver for each *TRAJ* segment using two-sample t-test). **(C)** CDR3α amino acid length in each sample. **(D,E)** Sequence logos generated from *TRAV1-2* CDR3α amino acid sequences of length 12, combined from all samples. Plots show all CDR3α sequences (D) or CDR3α sequences split by *TRAJ* gene usage (E). The overall height of the stacked letters at each position indicates the sequence conservation, while the relative abundance of each amino acid is indicated by the height of individual letters within the stack. Acidic bases are shown in red, basic residues in blue, hydrophobic in black and polar in green.

A transcriptionally distinct non-*TRAV1-2* MR1/5-OP-RU$^+$ T cell population was not identified

A small population of T cells have been identified that bind the MR1/5-OP-RU tetramer but express a *TRAV* chain other than *TRAV1-2* (Gherardin et al., 2016; 2018b; Koay et al., 2019b; Meermeier et al., 2016). The majority of these cells exhibit a non-MAIT-like phenotype (Koay et al., 2019b), but their transcriptional profile has not been studied at a genome-wide level. Therefore, I examined whether these cells could be identified in my data and whether they were transcriptionally distinct from MAIT cells.

There were some sorted MR1/5-OP-RU$^+$ T cells that expressed a *TRAV* chain other than *TRAV1-2*. These non-*TRAV1-2* MR1/5-OP-RU$^+$ T cells used diverse *TRAV* segments (Fig. 5.12A,B) and were dispersed throughout the clusters of *TRAV1-2* MAIT cells (Fig. 5.12C,D). The frequency of non-*TRAV1-2* cells amongst total MR1/5-OP-RU$^+$ T cells was in the range 3.6-7.4% (mean 5.1%). This is much higher than the expected frequency of this population based on the literature (median 0.78% of MR1/5-OP-RU$^+$ T cells) (Gherardin et al., 2018b; Koay et al., 2019b). While MAIT cells from donors GI3192 and GI3524 were sorted as MR1/5-OP-RU^{++} T cells, MAIT cells from the other two donors (GI3008, GI2916) were sorted as MR1/5-OP-RU^{++}Vα7.2$^+$ T cells and thus should all have expressed *TRAV1-2*. However, the frequency of non-*TRAV1-2* MR1/5-OP-RU$^+$ T cells per donor was independent of sorting strategy (Fig. 5.12E). Along with their higher than expected frequency, this suggested that the majority of non-*TRAV1-2* cells were in fact *TRAV1-2* cells that had lost expression of their *TRAV1-2* chain for technical reasons, with the detected TCRα chain being a second TCRα chain simultaneously expressed by these cells. This is conceivable as dropouts, where a gene is not

detected despite being expressed, are present at a high frequency in scRNA-seq data (Lafzi et al., 2018).

If this hypothesis is correct, you would expect to see an increased frequency of cells with a single as opposed to two TCRα chains within the non-*TRAV1-2* population relative to the *TRAV1-2* population (as all analysed cells had one or two TCRα chains). This trend was observed in the majority of samples (Fig. 5.12F,G), although the difference was not statistically significant when all samples were analysed together (Fig. 5.12H). Additionally, you would expect some CDR3α and CDR3β sequences from non-*TRAV1-2* TCRs to match with CDR3α and CDR3β sequences from *TRAV1-2* TCRs. On average, 46% (range 23-74%) of non-*TRAV1-2* MR1/5-OP-RU$^+$ T cells had both CDR3α and CDR3β sequences that were shared with TCRs expressed by *TRAV1-2* MAIT cells. Together, these observations support my hypothesis that the majority of non-*TRAV1-2* MR1/5-OP-RU$^+$ T cells detected were likely MAIT cells for which *TRAV1-2* expression was not detected. Thus, I am unable to comment on the transcriptional profile of non-*TRAV1-2* MR1/5-OP-RU$^+$ T cells.

Biased but not exclusive use of *TRBV6-1*, *TRBV6-4* and *TRBV20-1* gene segments by MAIT cells in blood and liver

MAIT cells showed biased use of *TRBV6-1*, *TRBV6-4* and *TRBV20-1* gene segments (Fig. 5.13A), as previously described (Lepore et al., 2014; Reantragoon et al., 2013; Tilloy et al., 1999). There was no significant difference in *TRBV* usage between blood and liver (Fig. 5.13A). Looking at the individual samples, *TRBV6-1*, *TRBV6-4* and *TRBV20-1* were all enriched (to a greater or lesser extent) in the three donors with matched blood and liver samples (Fig. 5.13B-D). In contrast, the *TRBV*

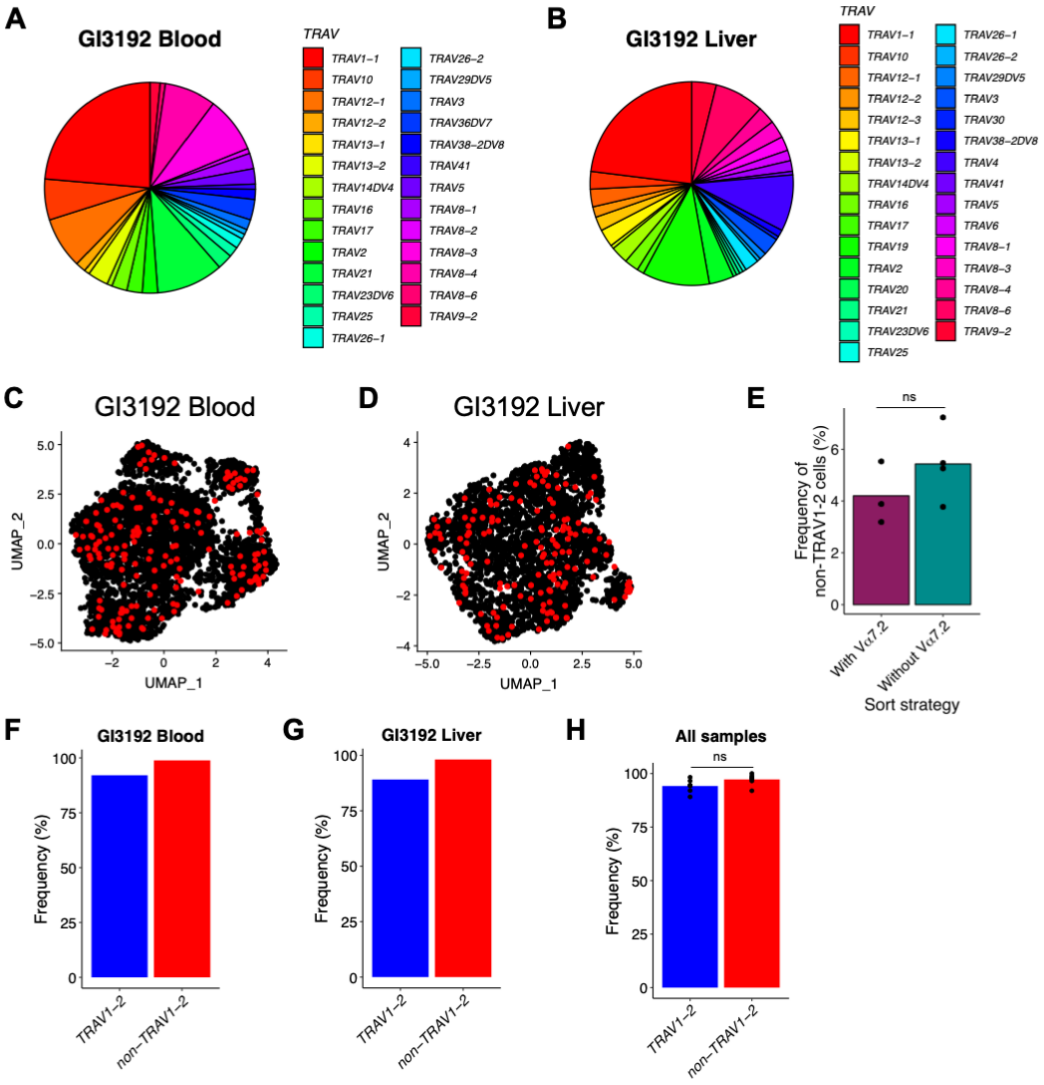


Figure 5.12. The majority of the MR1/5-OP-RU$^+$ non-*TRAV1-2* T cells identified are likely *TRAV1-2* MAIT cells for which *TRAV1-2* expression was not detected for technical reasons. Plots show a representative blood (A,C,F) and liver (B,D,G) sample, or all samples (E,H). **(A,B)** Pie charts showing *TRAV* segment usage amongst non-*TRAV1-2* MR1/5-OP-RU$^+$ T cells. **(C,D)** UMAP plots showing the location of non-*TRAV1-2* T cells amongst the MAIT cell clusters. **(E)** Bar chart showing the percentage of non-*TRAV1-2* T cells in each sample depending on whether or not Vα7.2 was used for sorting (alongside MR1/5-OP-RU). Two sample t-test. **(F,G)** Bar charts showing the percentage of *TRAV1-2* MAIT cells and non-*TRAV1-2* T cells expressing one TCRα chain within a representative blood (F) and liver (G) sample. **(H)** Bar chart showing the average percentage of *TRAV1-2* MAIT cells and non-*TRAV1-2* T cells expressing one TCRα chain. Black dots indicate individual samples. Statistical comparison was performed using a paired t-test. ns = not statistically significant.

repertoire of GI2916 liver MAIT cells was highly skewed towards *TRBV6-4*, with *TRBV6-1* and *TRBV20-1* used infrequently (Fig. 5.13E). This bias was largely driven by a single *TRBV6-4* clonotype, which comprised 49% of MAIT cells in this sample. Despite their biased *TRBV* repertoire, a variety of *TRBV* segments were used by MAIT cells in all samples (Fig. 5.13B-E). In GI3524, the most commonly used *TRBV* segment was actually *TRBV27* (Fig. 5.13D). Analogous to GI2916, this enrichment was largely driven by a single clonotype comprising 42% and 27% of MAIT cells in GI3524 blood and liver samples, respectively.

TRBV usage may exhibit some bias dependent on *TRAJ* usage

Despite limited variability in the MAIT cell TCRα chain, *TRAJ* gene usage had a predictable impact on CDR3α sequence, as demonstrated above (Fig. 5.11E). However, whether differential *TRAJ* usage is associated with altered segment usage within the TCRβ chain, has not been addressed. Therefore I examined *TRBV* segment usage by *TRAJ33*, *TRAJ12* and *TRAJ20* MAIT cell TCRs.

Compared with *TRAJ33*, *TRAJ12* and *TRAJ20* TCRα chains showed significantly increased pairing with *TRBV6-4* TCRβ chains (Fig. 5.14A,B). In contrast, *TRAJ33* showed significantly increased pairing with *TRBV20-1* TCRβ chains relative to both *TRAJ12* and *TRAJ20* (Fig. 5.14A,C). Pairing with *TRBV6-1* TCRβ chains was comparable across *TRAJ* segments (Fig. 5.14D).

Overall, the data suggest some preferential pairing of MAIT cell TCRα and TCRβ chains contingent on *TRAJ* and *TRBV* usage. However, due to the low percentage of *TRAJ12* and *TRAJ20* sequences within the MAIT cell TCR repertoire, the number of cells within these groups was very small in some samples. For example, in the blood of GI3524, only 96 and 30 *TRAV1-2* TCRs used *TRAJ12* and *TRAJ20*,

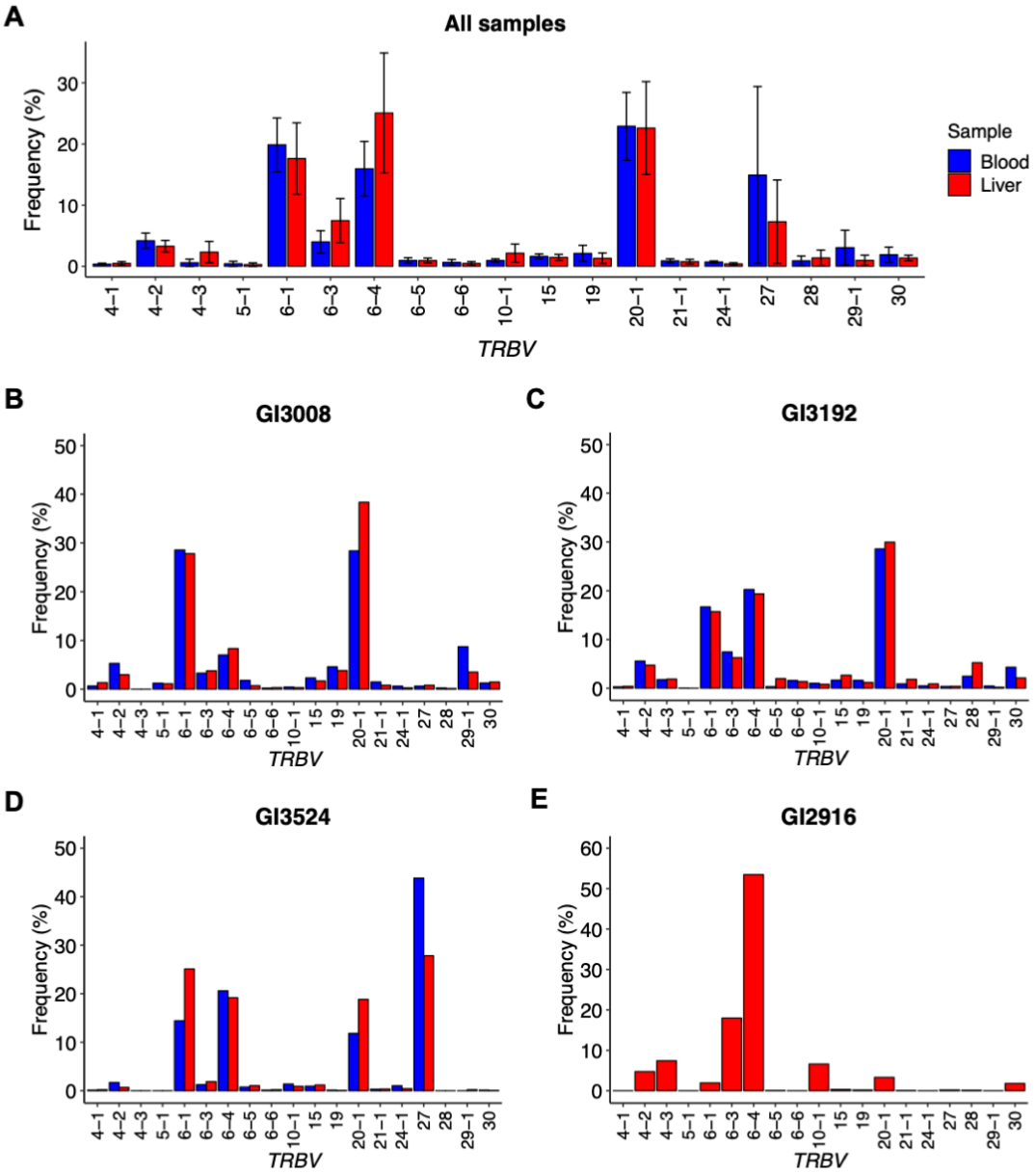


Figure 5.13. MAIT cells show biased but not exclusive use of *TRBV6-1*, *TRBV6-4* and *TRBV20-1* in blood and liver. (A) Bar chart showing the mean ± SEM percentage of MAIT cells using the indicated *TRBV* gene segments in the blood and liver. There was no significant difference in *TRBV* usage between blood and liver MAIT cells (paired t-test for each *TRBV* segment). **(B-E)** Bar charts showing the percentage of MAIT cells using the indicated *TRBV* gene segments in the blood and liver of donors GI3008 (B), GI3192 (C) and GI3524 (D), and in the liver of donor GI2916 (E). Blood samples are shown in blue and liver samples in red. Bar charts include *TRBV* segments with a frequency > 1% in any one sample.

respectively, compared with 2,085 cells for *TRAJ33*. Considering the low number of cells and the relatively small number of donors examined, further studies are required to validate the identified associations between *TRAJ* and *TRBV* segment usage.

CDR3β length shows considerable variability and is largely independent of *TRBV* or *TRAJ* segment usage

In contrast with the highly restricted length of the MAIT cell CDR3α region, the CDR3β length was highly variable (Fig. 5.15A). Across all samples, approximately 25% of CDR3β sequences were each of 14, 15 and 16 amino acids in length, with most of the remaining sequences being of length 12, 13, 17 or 18 amino acids (Fig. 5.15A). In donors GI3524 and GI2916, CDR3β lengths were heavily skewed towards 16 and 14 amino acids, respectively, driven predominantly by large individual clones within these samples.

I next examined whether CDR3β length varied depending on *TRBV* segment usage, focussing on *TRBV6-1*, *TRBV6-4* and *TRBV20-1*. CDR3β length was independent of *TRBV* segment usage, except for a significant increase in the frequency of 13 amino acid CDR3β sequences amongst *TRBV6-4* TCRs relative to *TRBV20-1* TCRs (Fig. 5.15B). However, there was considerable variability between samples in the distribution of CDR3β lengths for each *TRBV* segment.

Finally, I assessed whether MAIT cell TCRs with different *TRAJ* segments and therefore distinct CDR3α sequences showed preferential pairing with TCRβ chains of a particular CDR3β length. While the frequency of *TRAJ12*, *TRAJ20* and *TRAJ33* TCRs was similar at most CDR3β lengths, some significant differences were

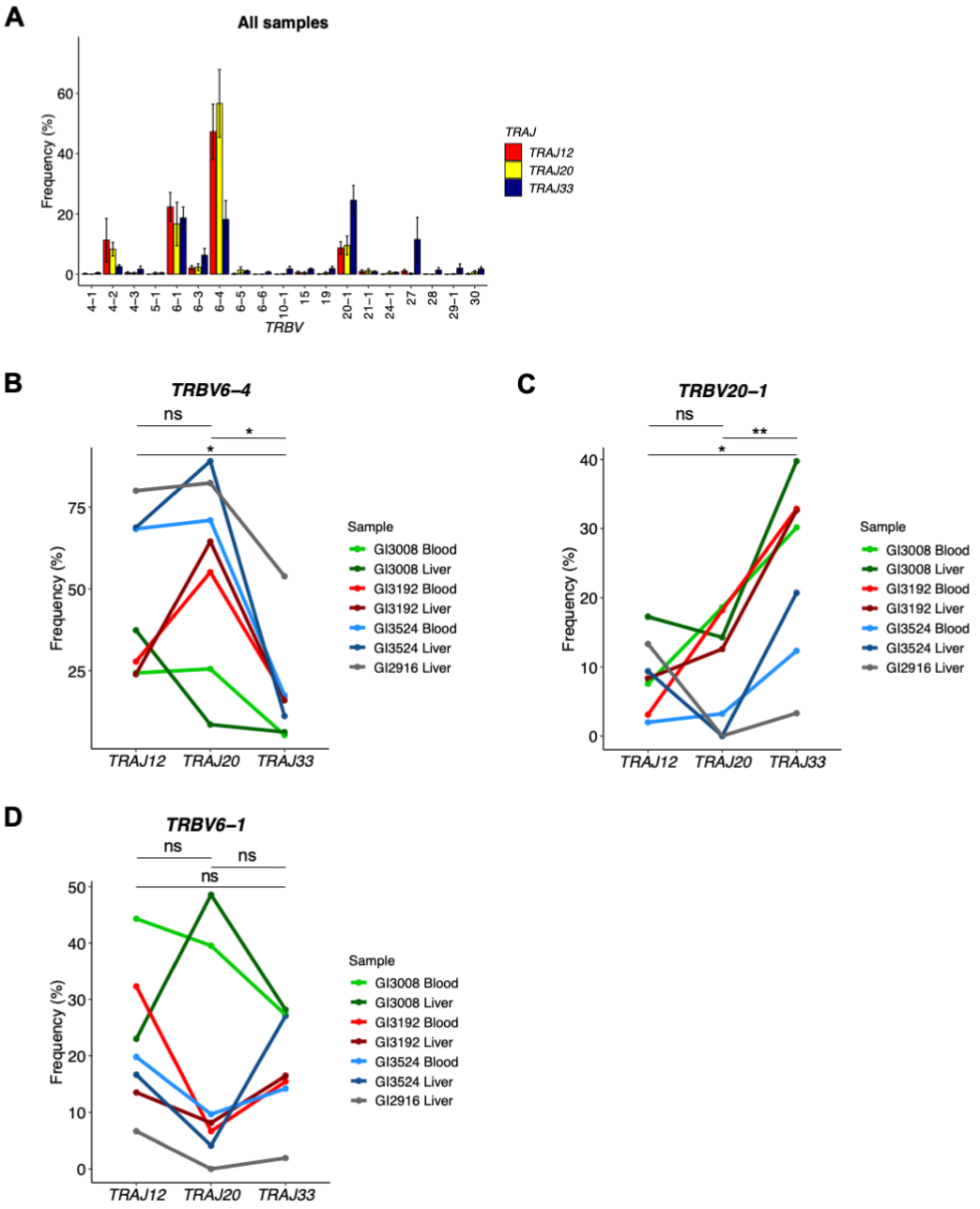


Figure 5.14. *TRBV* segment usage by *TRAV1-2* MAIT cell TCRs containing *TRAJ12*, *TRAJ20* or *TRAJ33*. (A) Bar chart showing the mean ± SEM percentage of *TRAJ12* (red), *TRAJ20* (yellow) and *TRAJ33* (blue) *TRAV1-2* MAIT cells using each *TRBV* gene segment across all samples. Bar chart includes *TRBV* segments with a frequency > 1% in any one sample. **(B-D)** Line plots showing the frequency of *TRAJ12*, *TRAJ20* and *TRAJ33 TRAV1-2* TCRs using *TRBV6-4* (B), *TRBV20-1* (C) and *TRBV6-1* (D) in each sample. Repeated measures one-way ANOVA with the Greenhouse-Geisser correction followed by Tukey's multiple comparisons test with individual variances computed for each comparison in (B-D). * $p < 0.05$, ** $p < 0.01$. ns = not statistically significant.

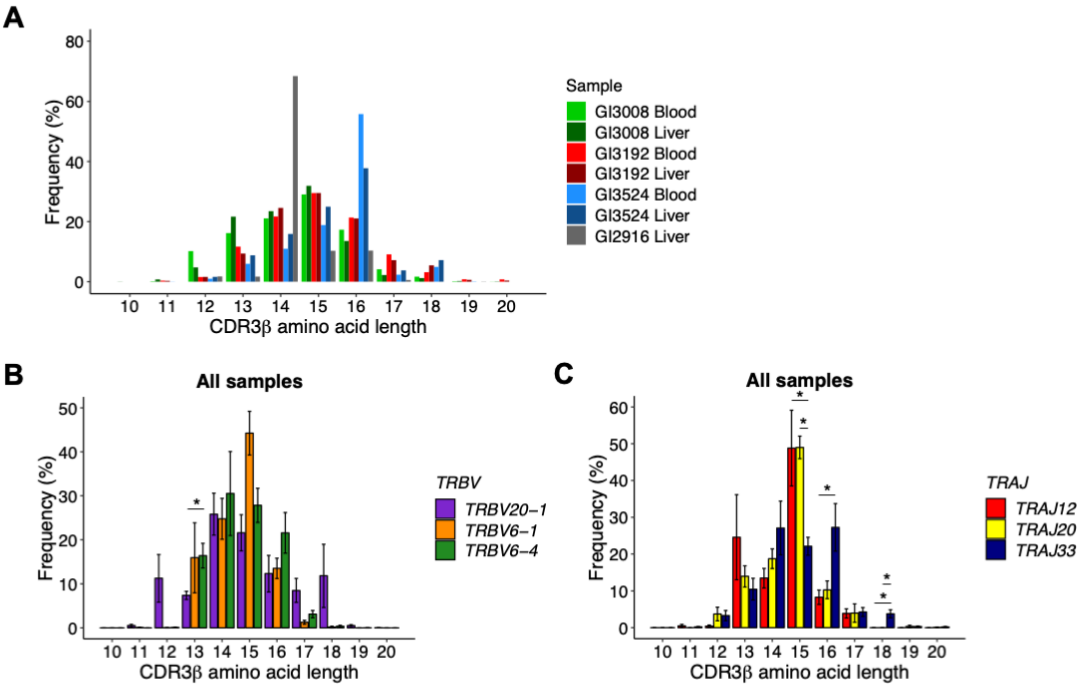


Figure 5.15. The length of MAIT cell CDR3β sequences was highly variable, but largely independent of *TRBV* and *TRAJ* segment usage. (A) Frequency of CDR3β sequences of different lengths in each sample. **(B)** Average frequency of CDR3β sequences of different lengths for *TRBV6-1* (orange), *TRBV6-4* (green) and *TRBV20-1* (purple) TCRs from all samples. **(C)** Average frequency of CDR3β sequences of different lengths for *TRAJ12* (red), *TRAJ20* (yellow) and *TRAJ33* (blue) TCRs from all samples. A minor percentage of MAIT cells had CDR3β lengths < 10 and > 20 amino acids (down to minimum of 8 and up to maximum of 26), which are not shown in these graphs. Error bars in (B,C) = mean ± SEM. Repeated measures two-way ANOVA with the Greenhouse-Geisser correction followed by Tukey's multiple comparisons test with individual variances computed for each comparison in (B,C). * $p < 0.05$.

identified. Compared with *TRAJ33*, the frequency of 15 amino acid CDR3β sequences was significantly increased amongst *TRAJ12* and *TRAJ20* TCRs, while 16 amino acid CDR3β sequences were significantly enriched in *TRAJ33* TCRs relative to *TRAJ12* TCRs (Fig. 5.15C). Finally, 18 amino acid CDR3β sequences were found at a higher frequency amongst *TRAJ33* relative to *TRAJ12* and *TRAJ20* TCRs (Fig. 5.15C).

Overall, CDR3β length was diverse and largely independent of *TRBV* and *TRAJ* segment usage, although some significant associations were identified. Given the small size of this study, whether the identified associations are specific to this dataset or more generally applicable, requires further investigation.

CDR3β sequence is highly diverse

Given the variable length of the CDR3β region, I initially restricted my analysis to 15 amino acid CDR3β sequences. In contrast with the highly conserved CDR3α sequence, MAIT cell CDR3β sequences showed considerable variability. While amino acids at the beginning and end of the CDR3β region (AA 1-4 and AA 12-15), which derive from *TRBV* and *TRBJ* gene segments, respectively, were relatively conserved, the centre of the CDR3β region showed little sequence conservation (Fig. 5.16A-D). This was predominantly due to junctional diversity created by the deletion and addition of nucleotides at the V-D and D-J junctions of the TCRβ chain. Similar results were obtained for 14 and 16 amino acid CDR3β sequences (data not shown).

MAIT cell TCR clonotypes are highly shared between matched blood and liver

The features of the MAIT cell TCRα and TCRβ chain were highly similar across matched blood and liver samples, suggesting there could be a large degree of repertoire sharing between MAIT cells in these two tissue sites. This was confirmed by the comparison of nucleotide clonotypes between blood and liver MAIT cells within each donor. Over 70% of MAIT cells from matched blood and liver belonged to a shared TCR clonotype (Fig. 5.17A-C). For example, in donor GI3192, 3,777 cells comprising 342 clonotypes were shared between the blood and liver, with 769 cells (564 clonotypes) and 559 cells (414 clonotypes) having a clonotype specific to the blood and liver, respectively. Similarly high clonal overlap was found between matched blood and liver T_{mem} cells (Fig. 5.17D,E).

Functional TCRαβ clonotypes show minor overlap between donors, but TCRα chains are highly shared

Given the restricted TCRα chain repertoire and high oligoclonality of MAIT cells, I investigated whether MAIT cell TCRs were shared across donors. For this analysis, clonotypes were defined based on the amino acid sequence of the CDR3α and CDR3β regions as opposed to the nucleotide sequence, as it is the protein structure that determines the functionality of the TCR. Specifically, functional TCR clonotypes were defined as cells that matched in at least one CDR3α and one CDR3β amino acid sequence.

Analysis was limited to the three donors with matched blood and liver samples. There was little sharing of functional TCR clonotypes between donors (Fig. 5.18A). Only one functional TCR clonotype was found in all three donors, with 98% of MAIT

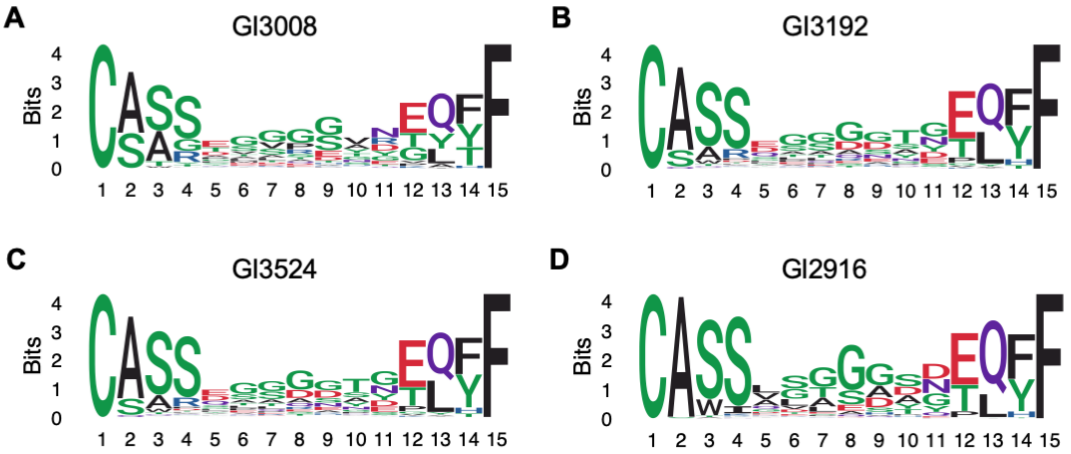


Figure 5.16. MAIT cell CDR3β sequences are highly diverse. (A-D) Sequence logos generated from all *TRAV1-2* CDR3β amino acid sequences of length 15 from donors GI3008 (A), GI3192 (B), GI3524 (C) and GI2916 (D). The overall height of the stacked letters at each position indicates the sequence conservation, while the relative abundance of each amino acid is indicated by the height of individual letters within the stack. Acidic bases are shown in red, basic residues in blue, hydrophobic in black and polar in green.

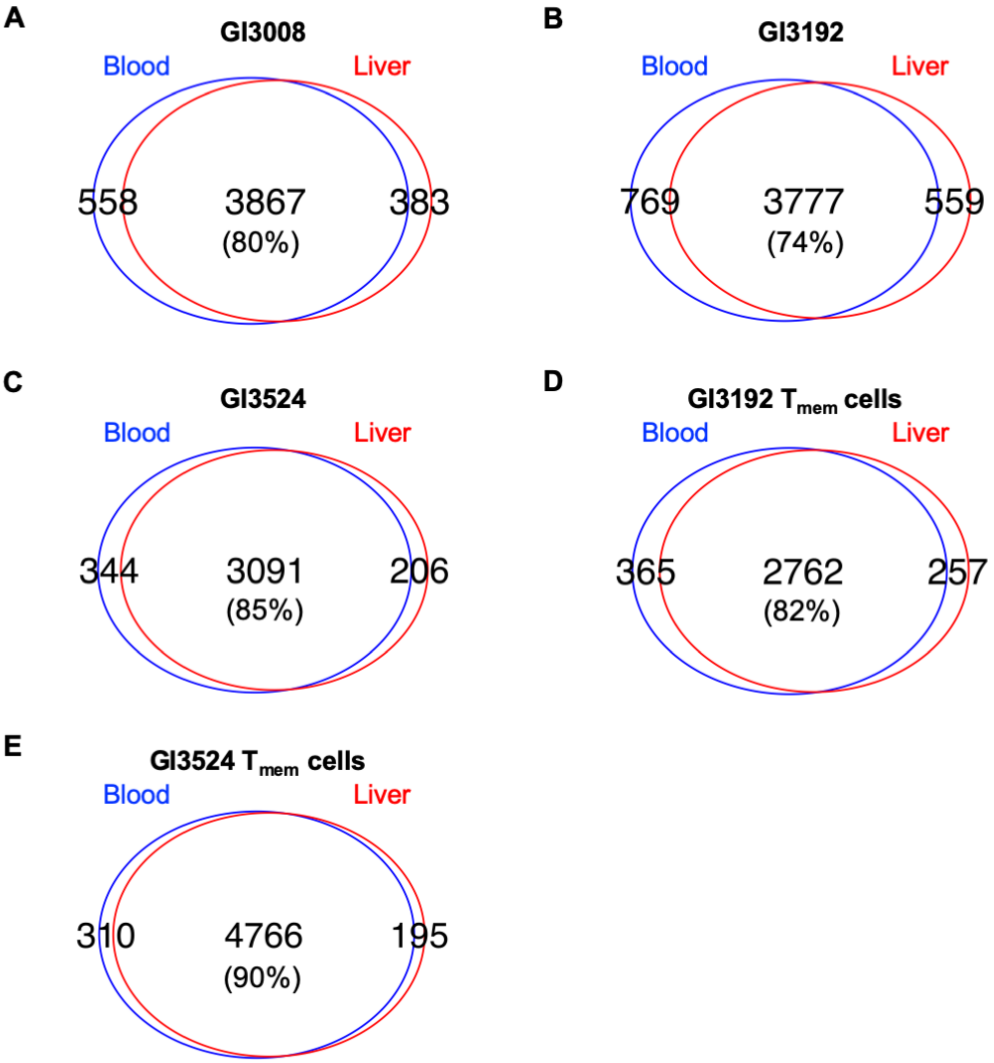


Figure 5.17. MAIT cell TCR clonotypes are highly shared between blood and liver, similar to T_{mem} cells. (A-D) Venn diagrams illustrating the overlap of nucleotide clonotypes between matched blood and liver MAIT cells (A-C), and matched blood and liver T_{mem} cells (D,E), in each donor. The numbers in the Venn diagrams indicate the number of cells with a clonotype either specific to the blood or liver, or shared between the blood and liver.

cells from each donor belonging to a clonotype that was unique to that donor. While minor sharing was identified for MAIT cell TCRs, T_{mem} cell functional TCR clonotypes showed no overlap in the two donors analysed (Fig. 5.18B).

As demonstrated, the MAIT cell TCRα chain comprises *TRAV1-2* paired with *TRAJ33*/*12*/*20* and has a restricted CDR3α length and sequence. The MAIT cell TCRβ chain shows greater diversity, despite biased use of particular *TRBV* segments. Based on this knowledge, I hypothesised that functional clonotypes defined based on the TCRα chain only (hereafter functional TCRα clonotypes) might show high overlap between donors. Consistent with this hypothesis, the three donors shared 44 functional TCRα clonotypes comprising 85% (11,524/13,544 cells) of total MAIT cells (Fig. 5.18C). In contrast, the TCRα chain repertoire of T_{mem} cells was largely distinct between donors (Fig. 5.18D). Nine functional TCRα clonotypes were shared between the two donors, comprising < 1% (71/8,771 cells) of all T_{mem} cells. Thus, distinct from conventional T_{mem} cells, MAIT cell TCRα chains are highly shared between individuals and it is the TCRβ chain which governs the uniqueness of individual MAIT cell TCR repertoires.

Summary

In summary, the MAIT cell TCRα chain was highly conserved, showing restricted *TRAV* and *TRAJ* usage, and defined CDR3α length and sequence. In contrast, the TCRβ chain showed considerable variability. While there was a bias in *TRBV* usage, CDR3β length and sequence was diverse. My data provide some evidence for the preferential pairing of specific TCRα and TCRβ chains, but this requires further study. The TCR repertoire of matched blood and liver MAIT cells was highly shared, but there was little sharing of TCRs between donors. Individuality in the

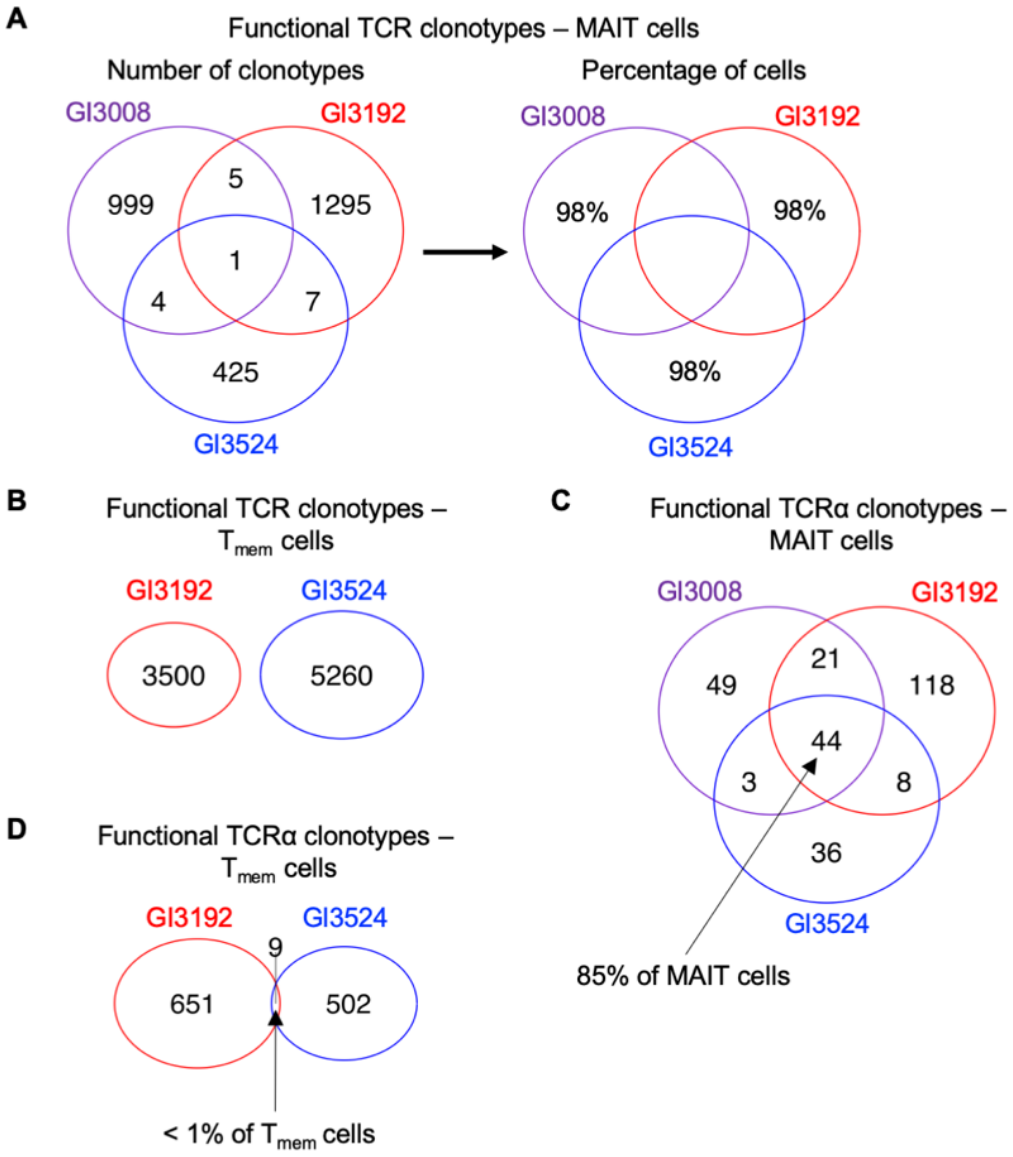


Figure 5.18. MAIT cell functional TCRαβ clonotypes are largely donor-specific, while functional TCRα clonotypes are highly shared between donors. (A) Venn diagrams illustrating the overlap of MAIT cell functional TCR clonotypes (paired TCRαβ) between donors (left) and the corresponding percentage of cells from each donor that have a donor-specific clonotype (right). **(B)** Overlap of T_{mem} cell functional TCR clonotypes between donors. **(C,D)** Overlap of MAIT cell (C) and T_{mem} cell (D) functional TCRα clonotypes between donors. The numbers in the Venn diagrams in (B-D) indicate the number of clonotypes that are donor-specific or shared between donors.

MAIT cell TCR repertoire was driven by diverse TCRβ usage, as TCRα chains showed a high degree of overlap between donors.

5.5.3. Gene expression and TCR usage

Having independently examined the transcriptome and TCR repertoire of blood and liver MAIT cells, I next wanted to investigate my third key question, namely whether MAIT cell transcriptional heterogeneity is associated with TCR usage.

TCR clonotypes are unequally distributed between clusters

The largest TCR clonotypes from each sample (named clonotype 1, 2, 3 etc.) were visualised on their respective UMAP, alongside a UMAP showing the identified MAIT cell clusters. TCR clonotypes were unequally distributed between clusters. The degree of bias in clonotype distribution varied between clonotypes and samples, and was not necessarily concordant between matched blood and liver MAIT cells. In some instances, individual TCR clonotypes were almost exclusively limited to specific clusters (Fig. 5.19A,B). For example, clonotype 1 from the liver of GI2916 specifically localised to clusters 0, 1 and 2, while clonotype 2 predominantly localised to cluster 3. In other cases, the biased localisation of clonotypes to clusters was more subtle (Fig. 5.19C). There were also instances where clonotypes did not appear to preferentially localise to particular clusters (Fig. 5.19D).

To more broadly visualise the localisation of clonotypes across clusters, I generated bar charts showing the largest 20 clonotypes from each sample and the percentage of cells from each clonotype that were localised within the various clusters (Fig. 5.20A-G). These graphs demonstrate the differences between clonotypes within individual samples in terms of their extent of preference for particular clusters, as

well as differences between donors in the number of clonotypes with strongly biased localisation.

Activation-associated genes and NK cell receptors show clonotype-specific expression

To establish why MAIT cell clonotypes may be biased towards particular clusters, I performed differential expression analysis between the clonotypes. Clonotype markers were defined as genes differentially expressed between each clonotype and all other cells. There were 351 unique clonotype markers across all samples. These included activation-induced molecules (e.g. *CD69*), effector cytokines (e.g. *IFNG*, *TNF*), chemokines (e.g. *CCL3*, *CCL4*, *CCL5*), NK cell receptors (e.g. *KLRD1*, *KLRF1*), and cytotoxicity-related genes (e.g. *GZMK*, *NKG7*) (Fig. 5.21A,B; Table 5.8). Of the 351 clonotype markers, 62% overlapped with cluster markers (Fig. 5.21C; Table 5.8). This suggested that MAIT cell clonotypes may differ in their level of activation, as observed for clusters.

To test this hypothesis, I performed pathway enrichment analysis on the clonotype markers. As with the cluster markers, this analysis was performed separately for blood and liver samples, so that I could establish whether differences between clonotypes were comparable across the two tissues. Consistent with clonotypes varying in their activation state, clonotype markers from blood (158 genes) and liver (285 genes) samples were enriched for pathways associated with cytokine and TCR signalling (Fig. 5.22A,B). The most enriched pathway in both tissues predominantly comprised NK cell receptors and was also highly enriched amongst cluster markers. Comparison of the clonotype markers with genes differentially expressed upon MAIT cell activation (as performed for cluster markers in Fig. 5.9A,B) revealed an

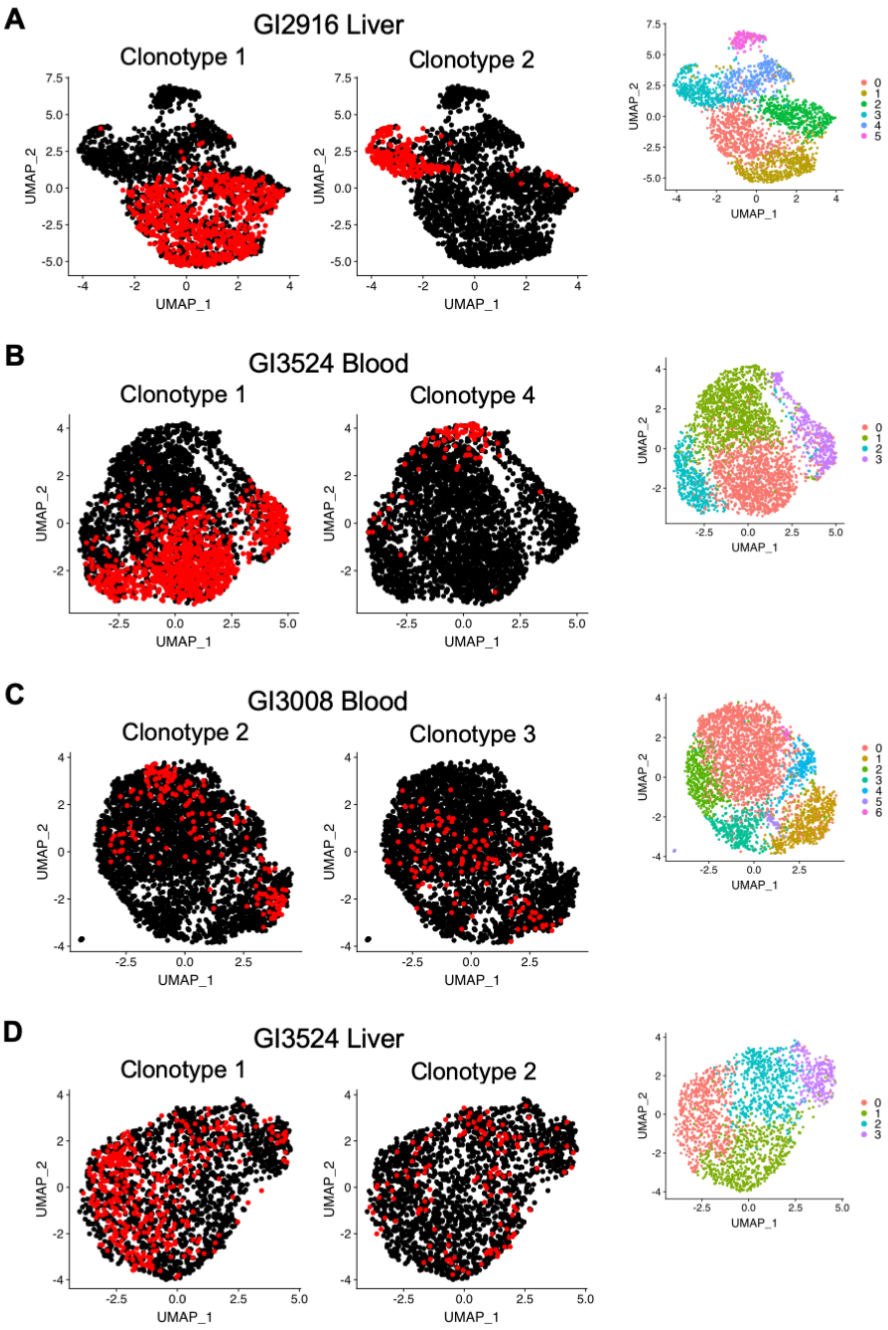


Figure 5.19. TCR clonotypes show biased localisation to particular MAIT cell clusters. Degree of bias varies between clonotypes and samples. (A-D) Individual clonotypes (shown in red) visualised on their corresponding UMAP. Previously identified transcriptional clusters are shown in the UMAP on the right. **(A,B)** Examples of clonotypes almost exclusively localised to particular clusters. **(C)** Examples of clonotypes with some bias in cluster localisation. **(D)** Examples of clonotypes that do not appear to preferentially localise to particular clusters.

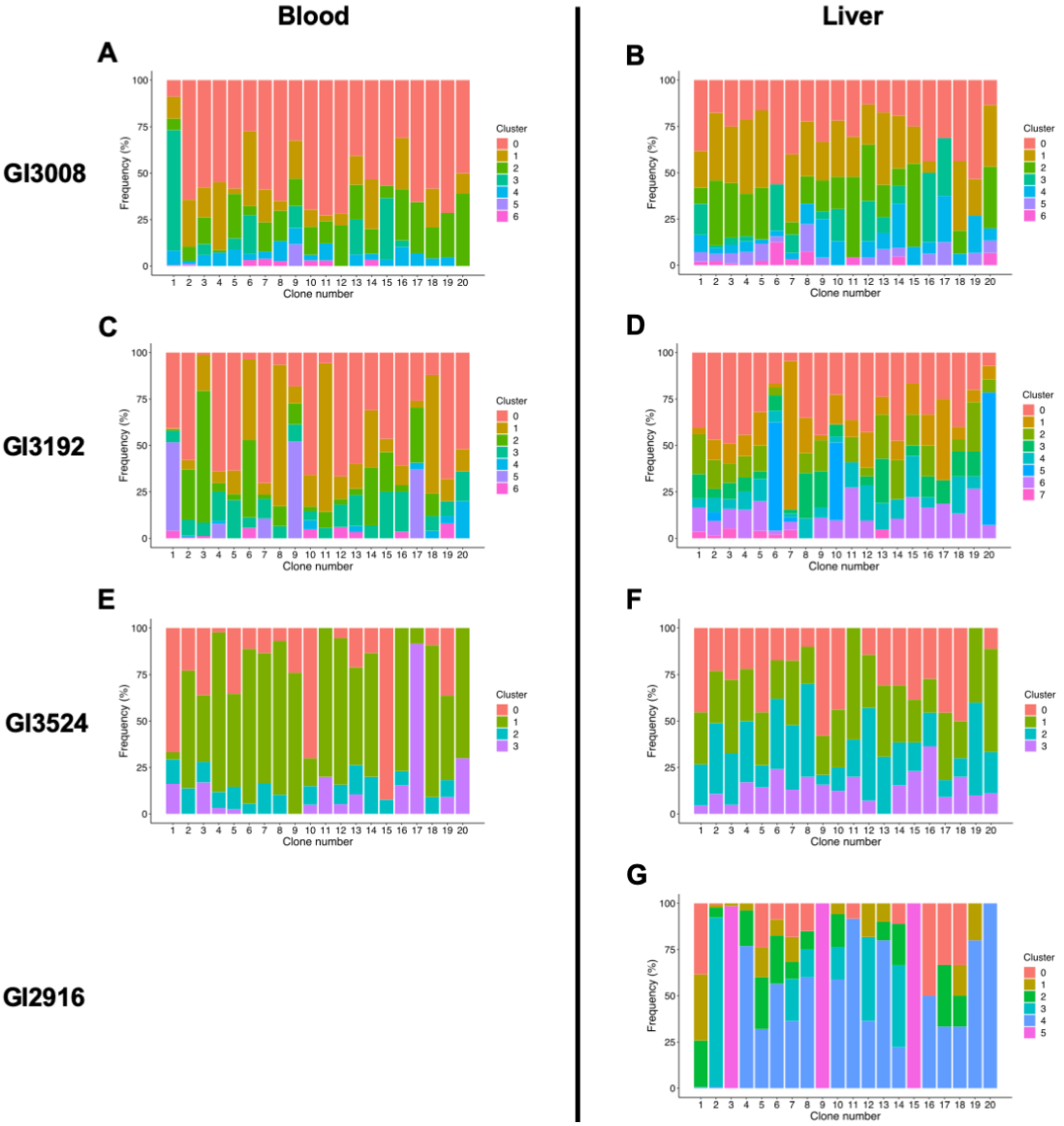


Figure 5.20. Localisation of MAIT cell clonotypes within transcriptional clusters. (A-D) Graphs showing the percentage of cells from each of the top 20 clonotypes from a sample found within each transcriptional cluster from the same sample. Blood samples are shown on the left and liver samples on the right for donors GI3008 (A,B), GI3192 (C,D) and GI3524 (E,F). The liver sample from donor GI2916 is shown in (G).

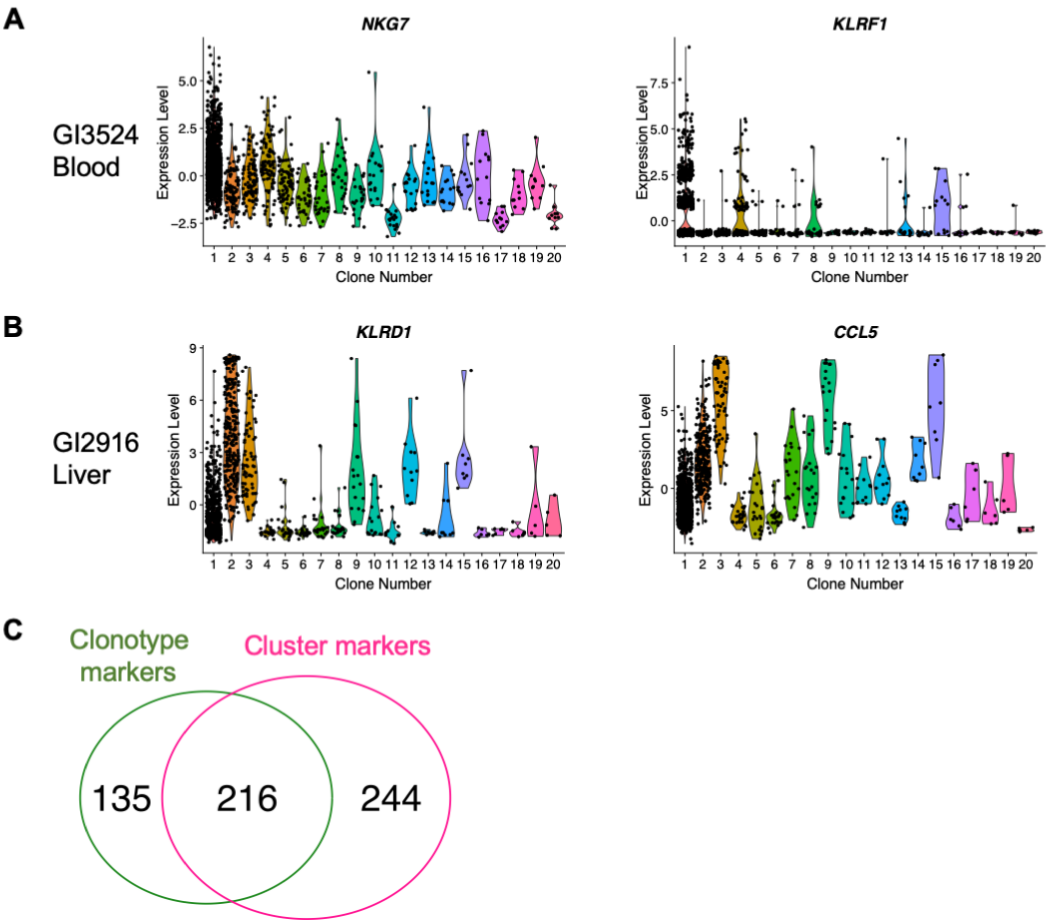


Figure 5.21. Clonotype markers are similar to cluster markers. (A,B) Expression of example clonotype markers in a representative blood (A) and liver (B) sample. The largest 20 clonotypes are shown for each sample. **(C)** Venn diagram showing the overlap between clonotype markers and cluster markers.

overlap of 64% and 59% for blood and liver, respectively (Fig. 5.22C,D), again supporting the hypothesis that clonotypes vary in their activation state.

Summary

To summarise, MAIT cell clonotypes showed biased localisation to particular MAIT cell clusters, although the degree of bias varied between clonotypes and samples. The genes and pathways altered in expression between clonotypes were similar to those that varied between clusters, and included cytokine and TCR signalling molecules, and NK cell receptors. Therefore, clonal origin contributes to phenotypic heterogeneity within the MAIT cell population. Altered expression of cytokine and NK cell receptors on MAIT cells is associated with differential responsiveness to cytokine stimulation (Dias et al., 2017). Thus, my findings suggest a potential association between MAIT cell clonality and responsiveness to TCR or cytokine stimulation, a hypothesis which I investigated in Chapter 6.

5.6. Discussion

Despite evidence for heterogeneity in the phenotype, function and TCR repertoire of human MAIT cells, whether they comprise multiple transcriptionally distinct subsets was previously unknown. Moreover, there was little data on the extent of TCR repertoire sharing between tissues and individuals, or regarding potential associations between the transcriptome and TCR usage of individual MAIT cells. Using single-cell transcriptomics, I have shown that matched blood and liver MAIT cells were highly clonally related, but displayed tissue-driven transcriptional changes. In contrast, the MAIT cell TCR repertoire was essentially unique to each individual. While MAIT cells within the blood and liver comprised multiple

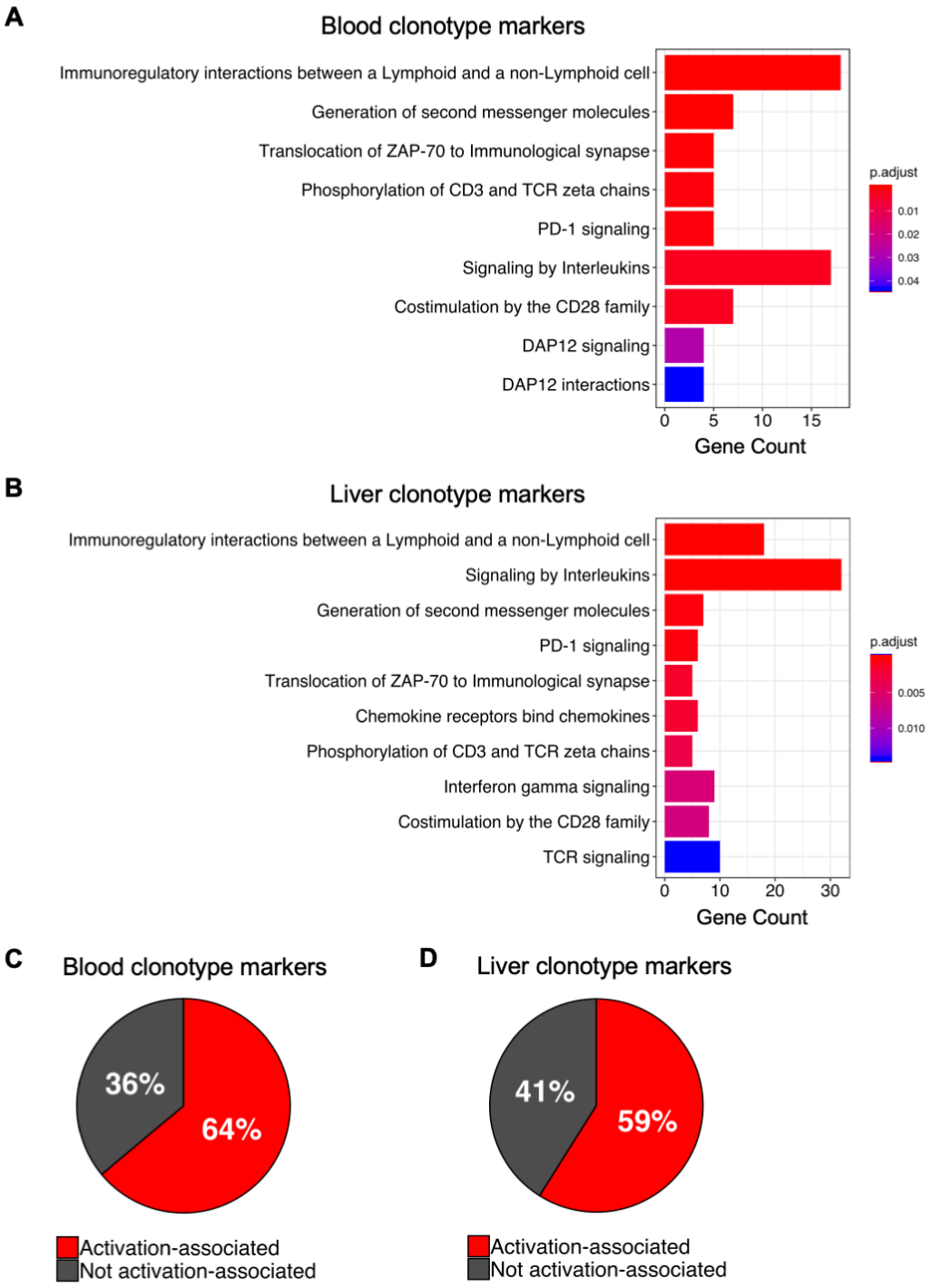


Figure 5.22. Clonotype markers are enriched for TCR and cytokine signalling pathways. (A,B) Reactome pathway enrichment analysis of the clonotype markers combined from blood (A) or liver (B) samples. Plots show selected enriched pathways. **(C,D)** Pie charts showing the overlap of clonotype markers from the blood (C) or liver (D) with the genes differentially expressed between stimulated (6 h TCR, 24 h TCR, *E. coli* or cytokine) and unstimulated MAIT cells in three published bulk RNA-seq datasets (Hinks et al., 2019; Lamichhane et al., 2019b; Leng et al., 2019). Figure 5.9A,B explains how the same comparison was performed for cluster markers.

transcriptionally distinct clusters, these appeared to reflect altered activation states, in part related to MAIT cell clonal origin, rather than fundamentally different populations.

Liver MAIT cells were transcriptionally distinct from blood MAIT cells (Fig. 5.3C-F), suggesting phenotypic and/or functional specialisation in tissues, in agreement with previous studies of MAIT cells in the female genital tract (Gibbs et al., 2016), oral mucosa (Sobkowiak et al., 2018) and adipose tissue (Carolan et al., 2015). Rather than the presence of unique or dedicated MAIT cell subsets in the liver, transcriptional changes were generally observed throughout the MAIT cell population, with liver MAIT cells appearing as a more activated, tissue-resident version of blood MAIT cells. This supports a model whereby MAIT cells undergo environmentally-driven adaptation following tissue entry, as opposed to a situation where tissue-specialised MAIT cell subsets preferentially home to particular tissue sites. Whether MAIT cell phenotypic and functional adaptations are associated with underlying epigenetic changes, and whether or not these are reversible, is currently unknown.

Published gene signatures associated with T cell tissue-residency were enriched in liver MAIT cells (Fig. 5.4A,B), consistent with bulk RNA-seq data (Salou et al., 2018). In mice, almost all liver MAIT cells are functionally resident, as demonstrated through parabiosis experiments (Salou et al., 2018). Taken together, these data suggest that human liver MAIT cells are likely to be functionally resident. Nevertheless, MAIT cells are present in human thoracic duct lymph at a similar frequency to matched peripheral blood, and MAIT cells in the lymph and blood show similar TCRα and TCRβ usage (Voillet et al., 2018). This implies that at least some human tissue MAIT cells can recirculate, although it is unknown if any of these

derive from the liver. The extent of MAIT cell functional residency in human liver could be investigated in a transplant setting using HLA mismatching to distinguish recipient- and donor-derived MAIT cells (Zuber et al., 2015).

The key questions of why MAIT cells are so abundant in the liver and what their major function is in this tissue site, are beginning to be addressed. MAIT cells express chemokine receptors associated with liver trafficking, including CCR5 and CXCR6 (Dusseaux et al., 2011; Oo et al., 2010). Following intravenous transfer of mature thymocytes into B6-MAIT^{Cast} recipients, MAIT cells preferentially home to the liver and lung relative to conventional T cells (Salou et al., 2018). Once in the liver, MAIT cells appear to have a profibrogenic role. In coculture assays *in vitro*, MAIT cells stimulate the proliferation of hepatic stellate cells and hepatic myofibroblasts, and increase their production of proinflammatory cytokines and profibrogenic factors (Böttcher et al., 2018; Hegde et al., 2018). Moreover, in an *in vivo* model of liver fibrogenesis, MR1$^{-/-}$ mice that lack MAIT cells exhibit reduced fibrosis, while fibrosis is enhanced in MAIT cell-enriched Vα19 transgenic mice (Hegde et al., 2018). Thus, in healthy liver, MAIT cells may play a role in wound healing and tissue repair. In support of this hypothesis, mouse skin MAIT cells show enriched expression of tissue repair-associated genes under homeostatic conditions (relative to conventional CD4^{+} T cells), and genes related to tissue repair are upregulated in mouse and human MAIT cells upon TCR-dependent activation (Constantinides et al., 2019; Hinks et al., 2019; Lamichhane et al., 2019b; Leng et al., 2019). Furthermore, MAIT cells in mouse skin (activated/expanded in response to *S. epidermidis* or 5-OP-RU) can enhance wound healing *in vivo* (Constantinides et al., 2019). In contrast, MAIT cells may be profibrogenic and thus pathogenic in chronic inflammatory liver disease. Whether this wound healing or profibrogenic activity is their main role in the liver, remains unknown.

My data indicate that human MAIT cells within tissues are fundamentally a single population, with phenotypic and functional heterogeneity reflecting differences in their activation state. This highlights a major difference between the MAIT cell population in humans and mice. In mice, MAIT cells comprise transcriptionally and functionally distinct MAIT1 and MAIT17 subsets that differentiate in the thymus (Koay et al., 2016; Legoux et al., 2019b; Rahimpour et al., 2015; Salou et al., 2018). MAIT1 cells express T-bet and type 1 cytokines, while MAIT17 cells, the major subset in mice, express RORγt and IL-17. In contrast, human MAIT cells express both T-bet and RORγt, and have mixed type 1 and type 17 functionality (Dusseaux et al., 2011; Koay et al., 2016; Le Bourhis et al., 2010; Leeansyah et al., 2015; Tang et al., 2013). The lack of distinct MAIT cell subsets identified in this study using single-cell transcriptional analysis provides further evidence for differences in the subset composition of mouse and human MAIT cells. Given the high evolutionary conservation of the MAIT cell TCR and MR1 (their MHC molecule), it is unclear why MAIT cells have evolved to comprise multiple specialised subsets in mice and one multifunctional population in humans. Nevertheless, it highlights the need to be cautious when directly translating MAIT cell discoveries from mouse to human.

Despite prior evidence for differences in human MAIT cell phenotype and function depending on coreceptor expression (Dias et al., 2018; Gherardin et al., 2018b; Kurioka et al., 2017; Leeansyah et al., 2015), *CD4*$^+$ and DN MAIT cells did not cluster separately from the major population of *CD8*$^+$ MAIT cells in my scRNA-seq data (Fig. 5.6). This is consistent with scRNA-seq data on MAIT cells from various mouse tissues, including the skin and lung (Constantinides et al., 2019). Previous studies have identified a relatively small number of proteins that are differentially expressed between CD4$^+$ and CD8$^+$/DN human MAIT cells, including the TFs PLZF (*ZBTB16*) and Eomes, and the NK cell receptors CD56, NKG2A (*KLRC1*) and

NKG2D (Gherardin et al., 2018b; Kurioka et al., 2017; Leeansyah et al., 2015). However, the genome-wide transcriptional profile of these populations has not been compared. Only 30 and 41 genes were differentially expressed between *CD4*$^+$ and *CD8A*$^+$*CD8B*$^+$/*CD8A*$^+$ MAIT cells from the blood and liver, respectively, in my scRNA-seq data (Table 5.5 and 5.6), of which only two liver genes (*ZBTB16*, *KLRC1*) overlapped with the differentially expressed proteins identified previously. Thus, further studies are necessary to validate the specific differences in gene and protein expression between CD4$^+$ and CD8$^+$/DN MAIT cells. Overall, current data indicates minor differences between CD4$^+$ and CD8$^+$/DN MAIT cells, which are not sufficient to drive distinct clustering of these populations.

DN MAIT cells are transcriptionally distinct from CD8$^+$ MAIT cells at a genome-wide level, as demonstrated by bulk RNA-seq (Dias et al., 2018). Thus, it is unclear why DN MAIT cells did not cluster separately to *CD8*$^+$ MAIT cells by scRNA-seq. It may be that the differences between resting DN and *CD8*$^+$ MAIT cells were comparatively smaller than the variability in activation-associated gene expression within the MAIT cell population. Alternatively, the genes differentially expressed between DN and *CD8*$^+$ MAIT cells may include a large fraction of lowly expressed genes that would be poorly detected by scRNA-seq due to the high frequency of dropouts (Bacher and Kendziorski, 2016). In my data, only five and eight genes were differentially expressed between *CD8A*$^+$*CD8B*$^+$/*CD8A*$^+$ and DN MAIT cells in the blood and liver, respectively (Table 5.3 and 5.4). This is likely an underestimation of the true differences in gene expression, because the DN population defined by scRNA-seq is a mixed population comprising true DN MAIT cells and CD8$^+$/CD4$^+$ MAIT cells for which *CD4* or *CD8A*/*CD8B* expression was not detected due to dropout. Thus, preliminary data suggests that DN and CD8$^+$ MAIT cells do not comprise transcriptionally distinct subsets, but further scRNA-seq

studies with more cells and a better defined DN population are required to confirm these findings.

With regard to the MAIT cell TCR repertoire, observed *TRAJ*/*TRBV* segment usage and CDR3α sequences were largely consistent with previous studies (Gherardin et al., 2018b; Held et al., 2015; Lepore et al., 2014; Reantragoon et al., 2013; Tilloy et al., 1999). There was some association between *TRAJ* and *TRBV* usage (Fig. 5.14). *TRAJ12* and *TRAJ20* TCRα chains showed statistically increased pairing with *TRBV6-4* relative to *TRAJ33*, while *TRAJ33* TCRα chains showed significantly increased pairing with *TRBV20-1* relative to *TRAJ12* and *TRAJ20*. *TRBV6-4* usage dominated the *TRAJ12* and *TRAJ2*0 populations identified by Reantragoon and colleagues, although the total number of *TRAJ12* and *TRAJ20* sequences obtained was only 25 and 17, respectively (Reantragoon et al., 2013). Additionally, all six *TRAJ12* clones generated by Lepore and colleagues paired with *TRBV6-1* or *TRBV6-4* (Lepore et al., 2014). Together, these data suggest an association between *TRAJ* and *TRBV* usage. However, given the considerable variability in *TRBV* usage between donors, the relatively small number of *TRAJ12* and *TRAJ20* MAIT cells analysed (15-287 and 17-161 *TRAV1-2*/*TRAJ12* and *TRAV1-2*/*TRAJ20* sequences analysed per donor, respectively), and the modest effect sizes obtained, larger studies will be required to verify the suggested preferences in *TRAJ*/*TRBV* pairing.

Blood and liver MAIT cells showed similar *TRAJ* and *TRBV* usage (Fig. 5.11B and 5.13A). Moreover, MAIT cells were highly clonally related in matched tissues, with over 70% of cells from the blood and liver belonging to a shared TCR clonotype (Fig. 5.17). This may even underestimate the degree of clonal overlap, as the sharing of smaller clonotypes may not have been detected due to sampling effects.

Previous studies showed variation in TCRα (Lepore et al., 2014) and TCRβ (Zumwalde et al., 2018) segment usage between MAIT cells in the blood and different tissues. However, matched tissues were not used in these studies. The high clonal overlap I have identified in matched blood and liver, combined with the lack of clonal sharing between individuals, illustrates the importance of using matched tissues when comparing MAIT cell TCR segment usage across tissues.

To my knowledge, this is the first study to unequivocally demonstrate the uniqueness of the paired MAIT cell TCRαβ repertoire across individuals. More specifically, 98% of MAIT cells from each donor belonged to a clonotype that was unique to that donor (Fig. 5.18A). A previous study showed a low overlap of CDR3β clonotypes between six donors using RNA deep sequencing on bulk sorted $V\alpha7.2^{+}CD161^{hi}$ T cells (Lepore et al., 2014). Comparison of CDR3β clonotypes between paired donors showed that on average 2% of unique CDR3β clonotypes present in one donor were also found in the second donor. This frequency is very similar to what I obtained for TCRβ clonotypes from the three donors with matched blood and liver samples (mean 2.3%, range 1.1-5.4%). However, in the previous study, these shared clonotypes made up on average 7.3% of the total TCRβ repertoire from each donor, while my single-cell data suggests that on average 2.6% (range 1.6-4.3%) of MAIT cells from one donor belonged to a clonotype that was shared with a second donor. The differences between the two studies could reflect differences in the donors studied or the techniques used for analysis. Bulk analyses, as used in the previous study, can be biased by the preferential amplification of particular CDR3β sequences (Rosati et al., 2017).

Although there was little overlap of paired TCRαβ clonotypes between donors, similarity in the antigen specificity of MAIT cells between individuals can be

explained by their highly shared TCRα chain repertoire (Fig. 5.18C). Unlike most conventional αβ T cells, the MAIT cell TCRα chain has a dominant role in antigen binding. More specifically, tyrosine 95 of the CDR3α region binds to the ribityl moiety of riboflavin metabolites and is essential for MAIT cell activation (Corbett et al., 2014; Eckle et al., 2014; Patel et al., 2013; Reantragoon et al., 2012). In contrast, no individual Vβ residue is required for MAIT cell activation (Reantragoon et al., 2012). The TCRβ chain appears to have a more minor role in modulating the affinity of the TCR for MR1-bound ligand and the degree of MAIT cell activation by riboflavin metabolites (Eckle et al., 2014; Howson et al., 2018), thus explaining why high variability is tolerated in the MAIT cell TCRβ chain.

Previous studies have identified a small population of MR1/5-OP-RU$^+$ T cells that lack expression of *TRAV1-2*, comprising a median of 0.78% of MR1/5-OP-RU$^+$ T cells and 0.02% of all T cells (Gherardin et al., 2016; 2018b; Koay et al., 2019b; Meermeier et al., 2016). The majority of these cells have a non-MAIT-like phenotype (Koay et al., 2019b) and therefore would be expected to cluster separately from classical MAIT cells. While I did identify non-*TRAV1-2* MR1/5-OP-RU$^+$ T cells in my data, these were dispersed throughout the MAIT cell clusters (Fig. 5.12C,D). Given their higher than expected frequency and overlapping TCR repertoire with MAIT cells, the majority of these cells were likely MAIT cells for which *TRAV1-2* was not detected for technical reasons. Consequently, I am unable to provide any insight into the genome-wide transcriptional profile of non-*TRAV1-2* MR1/5-OP-RU$^+$ T cells.

In some instances, transcriptionally distinct MAIT cell clusters within the blood or liver in fact reflected the presence of different clonotypes (Fig. 5.19). This is the first time that the clonal origin of MAIT cells has been associated with their resting

transcriptional profile and hence their phenotype. It is unknown whether transcriptional differences between clonotypes arise during thymic development or in the periphery later in life. The latter hypothesis seems most plausible, given that clonotype markers were enriched for activation-associated genes.

Clonotype-dependent activation states may occur as a result of: (1) intrinsic differences in TCR affinity, (2) intrinsic differences in the expression of other molecules that affect a cell's capacity for activation (e.g. NK cell receptors) and/or (3) stochastic variation between clonotypes in the time since *in vivo* activation. There is evidence from the literature for the differential activation of MAIT cells dependent on TCR usage (model 1). For example, MAIT cells responding to different bacteria *in vitro* exhibit distinct TCR repertoires (Gold et al., 2014), and MAIT cells undergo clonotype-specific expansion and contraction *in vivo* during human *S.* Paratyphi A infection (Howson et al., 2018). However*, in vitro* studies indicate a relatively minor impact of TCRβ variation on ligand affinity and cell activation (Eckle et al., 2014; Howson et al., 2018). Consistent with the second model, MAIT cell clonotypes differentially expressed NK cell receptors such as *KLRD1* (encodes CD94), and increased expression of CD94, along with CD56 and CD84, is associated with enhanced MAIT cell activation following cytokine stimulation (Dias et al., 2017). Furthermore, clonotype markers were enriched for cytokine signalling pathways. Therefore, it is plausible that variation in clonotype activation results from inherent differences in cytokine responsiveness. Finally, clonotypes may stochastically vary in activation state (model 3). This could occur in the blood if the cells derive from different tissues, in which they may or may not have undergone recent activation. In the liver, MAIT cells from a particular clonotype could exhibit similar spatial localisation and thus be exposed to comparable stimuli, resulting in clonotype-specific activation states. Overall, it is

worth investigating whether TCR clonotypes show differential activation in response to TCR and/or cytokine stimuli, as I have done in Chapter 6.

5.7. Limitations

There are a few of limitations of my data that should be taken into consideration when evaluating the findings. Firstly, it is not possible to obtain liver resections from entirely healthy donors for obvious reasons. The healthiest achievable source of liver tissue is probably from brain dead organ donors; however these were not available to me. Therefore, liver tissue was obtained from donors that were undergoing a liver resection to remove a tumour, comprising one donor with metastatic CRC, two with primary non-cirrhotic HCC (combined with nonalcoholic steatohepatitis [NASH] in one donor) and one with a liver haemangioma. While the pathologist ensured that tissue was taken from outside the tumour margins, it cannot be guaranteed that this tissue was completely free from all signs of disease. Nevertheless, the transcriptional profile (as measured by RNA-seq) of MAIT cells from peritumoral liver tissue from HCC patients is similar to that of healthy liver MAIT cells (Duan et al., 2019). Moreover, liver MAIT cells obtained from > 10 cm distal to metastatic CRC express high levels of IFNγ in response to both TCR-dependent and -independent stimuli, and are significantly more functional than tumour-infiltrating MAIT cells (Shaler et al., 2017a). Given that the liver tissue used in this study was always obtained from a site distal to the tumour, the MAIT cells analysed should resemble healthy liver MAIT cells.

Matched blood was also obtained from the patients undergoing liver resection, so that I could examine TCR repertoire sharing between matched tissues. Given their

illnesses, the frequency or phenotype of their blood MAIT cells may show some differences relative to MAIT cells from completely healthy individuals. A reduced peripheral blood MAIT cell frequency has been demonstrated in HCC (Duan et al., 2019) and CRC (Ling et al., 2016; Won et al., 2016), although one study found no significant difference in circulating MAIT cell frequency between CRC patients and healthy controls (Sundström et al., 2015). Nonalcoholic fatty liver disease (NAFLD) encompasses a spectrum of diseases (Mark Benedict, 2017), including NASH. In NAFLD patients, circulating MAIT cell frequency is decreased and MAIT cells exhibit an activated phenotype, with altered cytokine production (Böttcher et al., 2018; Li et al., 2018). However, given the diversity of clinical symptoms amongst NAFLD patients in the previous studies, the relevance of the findings to the NASH patient analysed in this study is unclear.

As well as disease-specific effects, blood and liver MAIT cell phenotype could have been influenced by any treatments the patients had received for their current disease or any prior illnesses. The donor with metastatic CRC received eight cycles of chemotherapy that were completed five months prior to surgery. Shaler and colleagues found no significant effect of chemotherapy (FOLFOX or FOLFOX + Avastin) on the frequency or functionality of MAIT cells within liver tissue derived from a site distant to metastatic CRC (Shaler et al., 2017b), suggesting that this treatment should not have considerably impacted my results.

The donors used in this study were aged between 51 and 65, and this should be taken into consideration when interpreting the findings. MAIT cell frequency decreases with age after a peak at approximately 25 years, with the proportion of $CD8^+$ MAIT cells declining throughout life (Gherardin et al., 2018b; Lee et al., 2014; Novak et al., 2014). While little is known about potential MAIT cell functional

changes with age in adults, one study found a decreased ratio of IFNγ to IL-4 secreting MAIT cells in elderly (70-83 years) individuals relative to young (25-39 years) individuals (Lee et al., 2014). Given the described age-associated functional changes in the conventional T cell population (Goronzy and Weyand, 2017), MAIT cells likely undergo some functional changes with age. MAIT cell clonality increases during childhood, but whether any changes are observed throughout adulthood has not been investigated.

Due to the high cost of scRNA-seq and scTCR-seq, we were only able to sequence blood and liver MAIT cells from three donors, along with a liver sample from a fourth donor. For the general transcriptional and TCR repertoire analysis, my findings were similar across donors. However, the extent to which MAIT cell clonotypes showed biased localisation to specific transcriptional clusters varied widely between donors. Therefore, further studies will be necessary to assess the degree of association between the TCR usage and transcriptional profile of individual MAIT cells. I have assessed this relationship in a further three donors in Chapter 6.

Finally, I was able to gain an insight into the transcriptional profile and TCR repertoire of MAIT cells in the liver, but whether my findings are applicable to other tissue settings is unknown. It would be particularly interesting to examine MAIT cells from mucosal sites, such as the gut or lung, where they will be exposed to a huge number of riboflavin-synthesising commensal bacteria that could impact on their activation state.

5.8. Conclusion

Overall, my study is the largest single-cell analysis of human blood and tissue MAIT cells to date and provides important insights into the MAIT cell population. Consistent with their mixed expression of type 1 and type 17 molecules, human MAIT cells fundamentally comprise a single population. Previously identified phenotypic and functional heterogeneity within and between different tissues reflects variation in activation state that depends in part on both MAIT cell clonal origin and tissue localisation. Additionally, my findings demonstrate that the human MAIT cell TCR repertoire is highly shared between blood and liver, but essentially unique to individuals as a result of diverse TCRβ usage. This was unexpected, given the semi-invariant nature of the MAIT cell TCR.

Table 5.1. Genes upregulated and downregulated in MAIT cells compared with T_{mem} cells in all of the four samples (two matched blood and liver) containing both cell populations.

Upregulated in MAIT vs. T_{mem}	**Downregulated in MAIT vs. T_{mem}**
KLRB1	*ZEB2*
IL7R	*PFN1**
NCR3	*FCRL6*
AQP3	*ITGB2*
LTB	*TMSB10**
SLC4A10	*HLA-DRB5*
TMIGD2	*ADGRG1*
LST1	*HLA-DRB1*
DPP4	*HLA-DPA1*
JAML	*HLA-DPB1*
CTSA	*NKG7*
ZBTB16	*CCL5**
SPOCK2	*GZMB*
FOS	*FGFBP2*
ERN1	*GZMH*
CCR6	*ITGB1*
CD69	*LGALS1*
DDIT4	*SH3BGRL3**
ZFP36L1	*KLRD1*
HSP90AB1	*HLA-DQA2**
FLT3LG	*CST7**
NFKBIA	*LINC02446**
PDCD4	*CD8B*
CD7	*GNLY*
*GBP5**	*FYB1*
	GSTP1
	*LSP1**
	SPON2
	CMC1

* = not differentially expressed between MAIT and T_{mem} cells in my bulk RNA-seq data (see Chapter 4).

Table 5.2. Differentially expressed genes between MAIT cells in the blood and liver.

Gene symbol	Average $log_eFC^{\#}$	% blood	% liver	p value	Adjusted p value
GNLY	0.667	0.287	0.104	2E-169	3.1E-165
MT-ND5	0.566	1	0.994	0	0
MT-CYB	0.541	1	0.995	0	0
MT-ATP6	0.475	1	0.986	0	0
CD52	0.469	0.997	0.973	0	0
HIST1H1E	0.442	0.363	0.084	0	0
MT-CO2	0.434	1	0.999	0	0
MT-CO3	0.431	1	0.997	0	0
CNN2	0.423	0.919	0.691	0	0
MT-ND6	0.413	0.845	0.685	1.1E-200	1.6E-196
MT-ND4L	0.396	1	0.999	0	0
LIMD2	0.390	0.825	0.566	0	0
ARL4C	0.390	0.759	0.48	0	0
LYAR	0.386	0.678	0.362	2.5E-306	3.9E-302
MT-ND3	0.364	1	0.995	0	0
MT-ATP8	0.356	1	0.995	0	0
EMP3	0.353	0.836	0.571	0	0
RPL13A	0.346	1	0.994	0	0
MT-ND1	0.337	1	0.996	0	0
FLNA	0.336	0.749	0.51	3.9E-278	6E-274
NOSIP	0.334	0.608	0.39	6.9E-208	1.1E-203
CDC25B	0.325	0.44	0.154	0	0
LITAF	0.319	0.816	0.567	6.8E-254	1.1E-249
RPL6	0.308	1	1	0	0
RPSA	0.302	1	1	0	0
SH3BGRL3	0.299	0.999	0.996	0	0
P2RY8	0.294	0.463	0.189	0	0
RPL18A	0.291	1	1	0	0
NKG7	0.289	0.995	0.982	3.5E-190	5.4E-186
RPL14	0.288	1	1	0	0
HOPX	0.283	0.706	0.485	5.3E-214	8.3E-210
S100A10	0.282	0.886	0.745	3.3E-264	5.1E-260
UCP2	0.278	0.738	0.529	3.7E-144	5.7E-140
C12orf75	0.277	0.361	0.11	0	0
RPL13	0.273	1	1	0	0
RPL3	0.265	1	1	0	0
RPL18	0.260	1	1	0	0
RPS14	0.259	1	1	0	0
RPLP0	0.258	1	1	0	0
RPL29	0.257	1	1	0	0
AES	0.257	0.552	0.331	7.4E-198	1.1E-193
RPL19	0.256	1	1	0	0
RPL23A	0.253	1	1	0	0
RPS8	0.253	1	1	0	0
RPL28	0.252	1	1	0	0
S100A4	0.252	1	0.998	0	0
MYC	0.251	0.362	0.174	2.5E-190	3.9E-186

RIPOR2	0.251	0.58	0.347	3.4E-192	5.2E-188
RPS16	0.251	1	1	0	0
RPL7A	0.250	1	1	0	0
KMT2E-AS1	-0.250	0.11	0.317	3.6E-286	5.6E-282
SERTAD1	-0.255	0.095	0.285	1.1E-180	1.7E-176
RBM39	-0.255	0.731	0.836	1.3E-144	2E-140
STK17B	-0.255	0.325	0.488	7.75E-34	1.2E-29
DNAJB4	-0.256	0.042	0.229	0	0
ITK	-0.256	0.174	0.39	4.9E-219	7.6E-215
SNHG12	-0.258	0.163	0.38	2.2E-186	3.4E-182
XCL1	-0.258	0.021	0.106	2.33E-84	3.61E-80
PIM2	-0.258	0.36	0.503	4.9E-106	7.6E-102
SLAMF1	-0.259	0.277	0.485	7.5E-176	1.2E-171
RORA	-0.260	0.461	0.631	2.3E-125	3.6E-121
HSP90AB1	-0.261	0.879	0.924	3.64E-81	5.64E-77
HNRNPA2B1	-0.261	0.923	0.96	6.1E-191	9.5E-187
ANKRD28	-0.263	0.051	0.229	1.2E-90	1.86E-86
PHLDA1	-0.279	0.034	0.214	1.6E-277	2.5E-273
AKAP13	-0.280	0.358	0.567	9.5E-211	1.5E-206
DDX5	-0.283	0.947	0.975	3.1E-185	4.8E-181
OXNAD1	-0.285	0.288	0.505	1E-156	1.6E-152
SARAF	-0.287	0.98	0.99	0	0
MT2A	-0.288	0.687	0.698	3.22E-49	4.99E-45
AC007384.1	-0.288	0.112	0.321	0	0
ACTG1	-0.289	0.994	0.998	0	0
CD53	-0.289	0.651	0.802	2.7E-246	4.1E-242
YME1L1	-0.290	0.339	0.55	1.8E-248	2.8E-244
AC087239.1	-0.291	0.126	0.362	0	0
CD2	-0.292	0.657	0.804	2.8E-211	4.4E-207
TMEM107	-0.293	0.183	0.357	1.5E-253	2.3E-249
TMSB4X	-0.294	1	1	0	0
SNHG9	-0.301	0.25	0.43	4.5E-224	6.9E-220
EIF4A3	-0.303	0.076	0.289	1.3E-263	2E-259
FYN	-0.305	0.606	0.753	0	0
CD40LG	-0.305	0.295	0.487	4.1E-201	6.3E-197
BATF	-0.306	0.191	0.334	1.48E-58	2.29E-54
RUNX2	-0.307	0.133	0.39	0	0
CDKN1A	-0.307	0.027	0.213	5.1E-248	7.9E-244
CD55	-0.308	0.249	0.469	0	0
POLR2A	-0.308	0.229	0.452	0	0
ZC3H12A	-0.315	0.092	0.308	0	0
PNRC1	-0.316	0.318	0.546	2.2E-201	3.5E-197
IPCEF1	-0.317	0.202	0.452	9.4E-282	1.5E-277
CCNL1	-0.318	0.485	0.666	4.1E-196	6.3E-192
PRF1	-0.319	0.858	0.9	0	0
PFKFB3	-0.322	0.067	0.3	0	0
ELF1	-0.323	0.409	0.635	6.3E-273	9.7E-269
CBLB	-0.323	0.368	0.606	0	0
CLK1	-0.324	0.255	0.518	5.6E-301	8.7E-297
CITED2	-0.327	0.104	0.318	7.3E-236	1.1E-231
DDX3X	-0.329	0.344	0.557	8.9E-288	1.4E-283

LINC01871	-0.330	0.49	0.683	4.2E-196	6.6E-192
AKIRIN2	-0.332	0.227	0.467	3.2E-179	4.9E-175
RGS2	-0.334	0.154	0.34	3.5E-194	5.4E-190
IKZF3	-0.340	0.361	0.598	2.7E-252	4.1E-248
RSRP1	-0.342	0.543	0.715	2.5E-195	3.8E-191
CELF2	-0.344	0.517	0.732	2.6E-289	4E-285
SCML4	-0.346	0.226	0.484	3.5E-268	5.4E-264
BRD2	-0.348	0.449	0.667	8.7E-255	1.4E-250
HSPA8	-0.350	0.963	0.99	4.5E-240	7E-236
WDR74	-0.352	0.286	0.463	1.6E-290	2.4E-286
PTGER4	-0.356	0.156	0.393	3.7E-208	5.7E-204
CD8A	-0.360	0.798	0.873	6.1E-168	9.4E-164
CD96	-0.368	0.508	0.732	1.6E-291	2.5E-287
GLUL	-0.370	0.119	0.382	0	0
REL	-0.372	0.205	0.448	0	0
ILF3-DT	-0.373	0.261	0.493	0	0
PPP2R5C	-0.382	0.588	0.735	1.9E-255	3E-251
RHOH	-0.383	0.337	0.588	0	4.2E-306
PPP1R10	-0.393	0.161	0.483	0	0
NR4A2	-0.395	0.192	0.419	3.3E-250	5.1E-246
CRYBG1	-0.405	0.155	0.454	0	0
ZFP36L2	-0.407	0.678	0.798	2.9E-286	4.5E-282
ID2	-0.411	0.724	0.834	8.5E-146	1.3E-141
ZFP36L1	-0.413	0.686	0.796	1.5E-127	2.3E-123
AC245014.3	-0.420	0.382	0.488	3.6E-258	5.6E-254
TAGAP	-0.424	0.455	0.697	3.6E-292	5.5E-288
TUBA4A	-0.430	0.248	0.547	0	0
SYTL3	-0.439	0.229	0.541	0	0
MALAT1	-0.443	1	1	0	0
TXNIP	-0.447	0.971	0.981	1.3E-296	2E-292
Z93241.1	-0.448	0.235	0.429	0	0
ATF3	-0.452	0.013	0.251	0	0
SLA	-0.456	0.209	0.545	0	0
CYCS	-0.479	0.518	0.753	0	0
PMAIP1	-0.486	0.155	0.44	0	0
SRSF7	-0.494	0.608	0.773	7.9E-198	1.2E-193
PARP8	-0.506	0.557	0.828	0	0
MCL1	-0.516	0.253	0.562	0	0
CCL3L1	-0.519	0.038	0.205	0	0
PIK3IP1	-0.523	0.396	0.735	0	0
PRDM1	-0.527	0.337	0.656	0	0
NEAT1	-0.533	0.434	0.722	0	0
UBC	-0.539	0.972	0.992	0	0
AC007952.4	-0.540	0.409	0.624	0	0
IFNG	-0.541	0.01	0.189	0	0
SAT1	-0.553	0.479	0.686	0	0
TSPYL2	-0.563	0.229	0.563	0	0
FKBP5	-0.564	0.256	0.668	0	0
IFNGR1	-0.567	0.328	0.658	0	0
DDIT4	-0.573	0.586	0.81	0	0
DNAJA1	-0.573	0.302	0.649	0	0

CXCR6	-0.599	0.199	0.538	0	0
CD160	-0.624	0.138	0.549	0	0
IL7R	-0.626	0.987	0.996	0	0
ADGRE5	-0.636	0.473	0.797	0	0
GADD45B	-0.639	0.399	0.673	0	0
CYTIP	-0.660	0.547	0.854	0	0
SLC2A3	-0.669	0.391	0.71	0	0
DUSP2	-0.704	0.26	0.431	0	0
CXCR4	-0.715	0.698	0.901	0	0
KLRD1	-0.724	0.266	0.536	0	0
CCL4L2	-0.738	0.033	0.195	4.8E-257	7.5E-253
CCNH	-0.788	0.367	0.781	0	0
BTG1	-0.800	0.647	0.904	0	0
EGR1	-0.805	0.003	0.341	0	0
DNAJB1	-0.827	0.553	0.726	0	0
NFKBIZ	-0.853	0.123	0.599	0	0
RGS1	-0.874	0.021	0.249	1.1E-225	1.6E-221
CCL3	-0.880	0.016	0.198	0	0
SRGN	-0.892	0.904	0.991	0	0
BTG2	-0.899	0.346	0.71	0	0
TNFAIP3	-0.945	0.593	0.851	0	0
TSC22D3	-1.039	0.798	0.97	0	0
IER2	-1.042	0.689	0.845	0	0
PPP1R15A	-1.184	0.473	0.881	0	0
TNF	-1.238	0.079	0.551	0	0
FOSB	-1.342	0.251	0.856	0	0
JUNB	-1.398	0.715	0.885	0	0
KLF6	-1.400	0.671	0.943	0	0
CD69	-1.497	0.746	0.975	0	0
DUSP1	-1.523	0.493	0.822	0	0
HSPA1B	-1.564	0.009	0.442	0	0
ZFP36	-1.613	0.768	0.97	0	0
NFKBIA	-1.683	0.584	0.952	0	0
HSPA1A	-1.780	0.035	0.35	0	0
JUN	-1.905	0.707	0.953	0	0
FOS	-2.098	0.594	0.957	0	0
CCL4	-2.441	0.233	0.756	0	0

[#] Positive FC indicates increased expression in the blood.

Table 5.3. Differentially expressed genes between DN and *CD8A*$^+$*CD8B*$^+$/*CD8A*$^+$ MAIT cells in the blood.

Gene symbol	Average log_eFC[#]	% DN	% *CD8*$^+$	p value	Adjusted p value
FOS	0.265	0.653	0.583	3.12E-31	4.83E-27
CCL5	-0.302	0.915	0.966	6.05E-84	9.37E-80
TYROBP	-0.310	0.147	0.334	4.82E-45	7.47E-41
*CD8B**	-0.641	0	0.505	0	0
*CD8A**	-1.470	0	1	0	0

[#] Positive FC indicates increased expression in DN MAIT cells.
* Differentially expressed in blood and liver.

Table 5.4. Differentially expressed genes between DN and *CD8A*$^+$*CD8B*$^+$/*CD8A*$^+$ MAIT cells in the liver.

Gene symbol	Average log_eFC[#]	% DN	% *CD8*$^+$	p value	Adjusted p value
LTB	0.595	0.806	0.604	4.31E-11	6.68E-07
GNLY	0.472	0.219	0.088	2.55E-10	3.95E-06
CD160	-0.255	0.39	0.575	3.83E-10	5.93E-06
CCL3	-0.330	0.118	0.21	8.20E-09	1.27E-04
CCL4L2	-0.353	0.089	0.211	1.03E-08	1.60E-04
*CD8B**	-0.613	0	0.479	2.15E-265	3.33E-261
CCL4	-0.834	0.511	0.795	4.29E-30	6.64E-26
*CD8A**	-1.772	0	1	0	0

[#] Positive FC indicates increased expression in DN MAIT cells.
* Differentially expressed in blood and liver.

Table 5.5. Differentially expressed genes between *CD4*$^+$ and *CD8A*$^+$*CD8B*$^+$/*CD8A*$^+$ MAIT cells in the blood.

Gene symbol	Average log_eFC[#]	% *CD4*$^+$	% *CD8*$^+$	p value	Adjusted p value
*CD4**	0.792	1	0	1.07E-193	1.65E-189
*LTB**	0.767	0.854	0.723	2.25E-07	3.48E-03
S100A11	0.470	0.854	0.585	8.17E-12	1.27E-07
*LDHB**	0.454	0.939	0.773	1.03E-16	1.60E-12
CISH	0.411	0.427	0.175	1.47E-06	2.28E-02
PIM2	0.353	0.646	0.355	2.75E-08	4.26E-04
*ANXA1**	0.346	0.89	0.795	1.86E-07	2.89E-03
CMTM6	0.332	0.585	0.322	1.73E-07	2.69E-03
*FXYD5**	0.327	1	0.954	1.19E-12	1.84E-08
NOSIP	0.320	0.854	0.588	8.66E-10	1.34E-05
*ITGB1**	0.297	0.39	0.118	6.33E-10	9.80E-06
*CD5**	0.289	0.585	0.3	5.73E-08	8.88E-04
*CTSB**	0.279	0.39	0.126	2.77E-09	4.29E-05
*EMP3**	0.260	0.988	0.827	8.78E-08	1.36E-03
HLA-DPA1	-0.274	0.073	0.316	9.21E-09	1.43E-04
LCP1	-0.308	0.732	0.885	6.65E-07	1.03E-02
*TMIGD2**	-0.332	0.085	0.382	2.61E-06	4.04E-02
*PRF1**	-0.341	0.695	0.867	2.78E-07	4.31E-03
*HCST**	-0.357	0.927	0.983	5.44E-09	8.43E-05
*KLRB1**	-0.386	0.939	0.997	4.12E-19	6.39E-15
*KLRD1**	-0.400	0	0.281	2.11E-17	3.26E-13
*GZMM**	-0.439	0.561	0.867	5.73E-13	8.88E-09
*CST7**	-0.568	0.793	0.967	1.87E-20	2.90E-16
*CD8B**	-0.641	0	0.505	1.17E-25	1.82E-21
*CD7**	-0.663	0.537	0.892	2.62E-24	4.07E-20
*CTSW**	-0.670	0.854	0.987	3.59E-30	5.55E-26
TYROBP	-0.779	0.012	0.334	6.00E-13	9.30E-09
*NKG7**	-0.782	0.878	0.999	1.25E-50	1.94E-46
*CCL5**	-1.017	0.768	0.966	3.28E-50	5.08E-46
*CD8A**	-1.470	0	1	1.07E-193	1.65E-189

[#] Positive FC indicates increased expression in *CD4*$^+$ MAIT cells.

***** Differentially expressed in blood and liver.

Table 5.6. Differentially expressed genes between *CD4*$^+$ and *CD8A*$^+$*CD8B*$^+$/*CD8A*$^+$ MAIT cells in the liver.

Gene symbol	Average log_eFC[#]	% *CD4*$^+$	% *CD8*$^+$	p value	Adjusted p value
*CD4**	0.845	1	0	1.04E-195	1.61E-191
*LTB**	0.591	0.835	0.604	4.85E-07	7.51E-03
*ANXA1**	0.579	0.911	0.696	1.77E-08	2.75E-04
*LDHB**	0.492	0.861	0.674	3.06E-14	4.74E-10
*EMP3**	0.405	0.823	0.553	3.12E-10	4.83E-06
GSTK1	0.384	0.848	0.701	1.67E-09	2.59E-05
FTH1	0.379	0.987	0.996	5.00E-07	7.74E-03
*CTSB**	0.370	0.43	0.111	9.06E-16	1.40E-11
TMEM173	0.347	0.57	0.257	3.90E-10	6.04E-06
TIMP1	0.343	0.38	0.144	6.05E-09	9.38E-05
*FXYD5**	0.311	0.975	0.897	6.23E-09	9.65E-05
*CD5**	0.304	0.481	0.275	4.79E-07	7.43E-03
ARID5B	0.294	0.291	0.07	1.17E-09	1.82E-05
TMSB10	0.284	1	1	2.86E-09	4.44E-05
HINT1	0.275	0.962	0.877	2.22E-07	3.44E-03
GAPDH	0.275	1	0.988	7.11E-08	1.10E-03
*ITGB1**	0.258	0.266	0.094	1.41E-07	2.18E-03
FRMD4B	0.255	0.266	0.062	3.17E-09	4.91E-05
THEM4	0.252	0.38	0.169	3.27E-07	5.07E-03
*GZMM**	-0.303	0.481	0.733	5.86E-07	9.08E-03
KLRG1	-0.307	0.633	0.779	6.01E-07	9.32E-03
*HCST**	-0.314	0.911	0.979	2.64E-08	4.09E-04
*CST7**	-0.333	0.785	0.942	8.59E-09	1.33E-04
GZMK	-0.337	0.684	0.888	4.70E-09	7.29E-05
FYN	-0.345	0.532	0.764	1.79E-06	2.78E-02
KLRC1	-0.353	0	0.164	2.34E-07	3.63E-03
SLC4A10	-0.357	0.253	0.551	7.87E-08	1.22E-03
*KLRB1**	-0.396	0.886	0.998	8.59E-23	1.33E-18
*CD7**	-0.466	0.582	0.904	1.55E-14	2.41E-10
*TMIGD2**	-0.469	0.127	0.507	1.93E-12	3.00E-08
CCL3L1	-0.552	0.025	0.209	1.24E-08	1.91E-04
NCR3	-0.570	0.392	0.836	5.35E-18	8.29E-14
*CCL5**	-0.596	0.747	0.96	1.26E-22	1.96E-18
*CD8B**	-0.613	0	0.479	2.88E-24	4.46E-20
CD160	-0.639	0.063	0.575	2.13E-18	3.29E-14
*CTSW**	-0.695	0.684	0.951	2.32E-29	3.59E-25
*NKG7**	-0.762	0.557	0.986	9.69E-54	1.50E-49
*PRF1**	-0.784	0.481	0.905	1.95E-24	3.03E-20
*KLRD1**	-0.922	0.051	0.549	1.16E-21	1.79E-17
CCL4	-1.581	0.228	0.795	1.54E-27	2.38E-23
*CD8A**	-1.772	0	1	1.04E-195	1.61E-191

[#] Positive FC indicates increased expression in *CD4*$^+$ MAIT cells.

***** Differentially expressed in blood and liver.

Table 5.7. Table showing the cluster markers from all samples and whether they were upregulated or downregulated in MAIT cells under different stimulation conditions, as measured by bulk RNA-seq.

Cluster markers	TCR 6h	TCR 24h	*E. coli*	Cytokine
GNLY				
TYROBP				
ACTB				
IL7R	↓			↓
ACTG1				
MTRNR2L12				
PFN1				
TMSB10	↓			↑
IL32	↓			↑
MALAT1		↓		↓
PSME2				
TPM4	↓			↑
CALR	↑	↑		↑
HSP90AB1	↑	↑	↑	↑
SELL				
IL2RB			↑	↑
MTRNR2L8				
HIST1H1E	↓			
RNF213	↓		↑	
MYH9	↓			
FLNA	↓			
CX3CR1				
FGFBP2	↓			
GZMB	↑		↑	↑
GZMH			↑	↑
CD79A	↑			
ZEB2	↑			
HLA-DRA	↑	↓		
HLA-DQA1	↑			
ADGRG1	↓			
HLA-DRB1				
EEF1A1				
CD52				↓
HLA-DPB1	↓			↓
RPS12				
LTB				
GZMK	↓			
FCGR3A	↓		↓	
RPL32				
HLA-DRB5				
MT-CYB				
MT-ND1				

FGR	↓			
AQP3				↓
NKG7				
MT-ND2				
CCL4	↑	↑	↑	↑
HLA-DPA1				
RPS25				
COTL1				
CD8B	↓			
RPS13				
CTSA				↓
LITAF				
MT-ND3				
VIM				
HVCN1				
EFHD2				
C12orf75	↓			
SH3BGRL3				
CCL5				↓
C1orf21				
USP28	↓			
MT-ND4				
UCP2	↓			
LSP1	↓			↓
SRPK2	↓			
DUSP1	↓			↓
RASGRP2	↓	↓		↓
KLRD1				
FLT3LG	↓			
CD70				↑
FDXR	↓			↓
PVT1				
RPS19				
PHPT1				
RPS27L				
TNFSF8	↑			
DDB2	↓			↓
BAX				
SRGN	↑			
CD160	↑		↑	
CD69	↑	↑	↑	↑
CCL3	↑	↑	↑	↑
CCL3L1	↑	↑		↑
AC007952.4				
CD8A				
DUSP2	↑		↑	
XCL1	↑			

CCL4L2	↑	↑		
TNF	↑	↑	↑	↑
IFNG	↑		↑	↑
NFKBIZ	↑	↑	↑	↑
EGR1	↑	↑	↑	
JUN				↑
AC245014.3				
HSPA1A	↑			
Z93241.1				
IER2				
FOSB	↑		↑	
BTG2	↑	↑	↑	↑
JUNB	?	↑	↑	↑
PPP1R15A	↓			
FOS	↓			
WDR74	↑			
RSRP1				
ILF3-DT				
ZFP36	↓		↑	
MCL1			↑	
TMEM107	↓			↓
SAT1			↑	
EMP3	↓			
KLF6	↓		↑	
C12orf57				
SNHG9				
SRSF7				
TSPYL2	↑			
ID2	↑	↑	↑	↑
SLC2A3			↑	↑
NFKBIA	↑	↑	↑	↑
TNFAIP3				
XCL2	↑		↑	
REL	↑	↑	↑	
ICAM1		↑	↑	↑
FEZ1	↑	↑	↑	↑
PTGER4				↑
MT2A			↑	↑
HSPA1B	↑			
DNAJB1	↓			
UBC				
EIF1				↑
CXCR6	↓			
DDIT4	↓		↑	↑
GADD45B				↑

H3F3B			↑	↑
PMAIP1	↑	↑	↑	↑
FTH1				↑
NR4A2	↑	↑	↑	
GIMAP4	↓			↓
CCNH				
TSC22D3	↓			↓
GIMAP7	↓			
CSRNP1	↑		↑	
SERTAD1	↑		↑	
LCP1				
S100A11				
TUBA1A	↓			
CISH	↓			↑
TXNIP	↓		↓	
CCNL1				
CYCS	↑		↑	↑
DDX3X				
NEAT1				
IFRD1				
ATF3	↑		↑	↑
PHLDA1	↑	↑	↑	↑
SESN1				↓
POLR2A				
PRF1			↑	↑
DNAJA1	↑	↑	↑	↑
HSPA6				
HSPH1	↑	↑		↑
HSPE1	↑			
HSP90AA1				↑
HSPA8	↑			
ZFP36L2	↓			↓
ZFAND2A				
DNAJB4	↑			
BTG1				
CACYBP				
RGS2				
CHORDC1	↑			
PPP2R5C				
MT-CO3				
UBB				
MT-ND5				
CXCR4	↓			↓
PRDM1	↓		↑	↑
PNRC1	↓			
BRD2				

HSPB1				
HSPD1	↑			↑
AC004687.1				
LMNA				↑
MYADM	↓			
ANXA1	↓			↓
TAGLN2				
RGCC	↑		↑	
ATP2B1				
ARL4A				
CRIP1	↓			
CD44	↑		↑	
CA5B	↓			
AHNAK	↓			↓
ADGRE5				
C16orf54	↓		↓	
PBXIP1	↓		↓	
RNF19A	↑		↑	
DENND4A				
EZR				
ANXA2	↓			↑
YWHAQ				
TUBB4B	↑			
RAC2				
YWHAZ				
RGS1				↑
THEMIS	↓			
BATF		↑	↑	↑
COLQ	↓			
CNN2	↓			↓
HOPX				
LINC02446				
SLA	↑		↑	
KLRC1	↑		↑	
CASP1	↓			
CD40LG	↑		↑	↑
NR4A1	↑	↑	↑	
CARD16	↓			
FABP5	↑	↑	↑	↑
EIF4A3	↑			
YPEL5	↓			
TOB1	↓			
FTL				
SPRY1	↑		↑	
CDKN1A	↑	↑	↑	↑
CMC1	↓			

AREG	↑		↑	
CCL20	↑	↑	↑	↑
RGS18				
KLRF1	↓			
KLRG1		↑		↓
GPR55				
PRICKLE1				
LGALS3				↑
NCAM1	↓			
S1PR5	↓			↓
GLUL				
PLEK	↑	↑	↑	↑
TXK	↓			
GABARAPL1				
CST7				
TSPO	↓			
SRSF5	↓			
S100A4	↓		↓	
PRKACB	↓		↓	
SLC4A10	↓			↑
GPR171	↑		↑	
CD84	↑			↓
IRF1				
IL23R	↑	↑	↑	↑
GBP5			↑	↑
SLC38A1				
MAP3K8			↑	↑
GLIPR1				
GZMM	↓			
ITM2A				
BIN1	↓			
MFSD10				
FXYD5				
MYL12B				
KLF2	↓	↓		↓
CRTAM	↑	↑	↑	↓
AC016831.7				
ANKRD28				
FNBP1				
GZMA	↓			↑
TPT1				
RPL18A				
RPS8				
RPLP0				
AOAH				
FYB1				
AKIRIN2				

PTPRC				
TOX				
NOSIP				
SYTL3	↓			↑
IKZF3	↑		↑	
KLRC3	↑			
ZNF683				
SPRY2	↑	↑	↑	
FCRL6	↓			
CLIC3	↓		↓	
CXCR3				↓
MS4A1				
CD7			↑	
DUSP4	↑	↑	↑	↑
PABPC1				
PXN	↓			
NCR3	↓			
KLRB1				↓
GSTP1				
DOK2	↓			
CCR6	↓			
SMIM3	↓			
GNPTAB				
RIN3	↑		↑	
SLA2				↓
ITM2C				
RUNX2	↓			
GAPDH	↑			
S100A6	↓			
CD55				↑
IL6ST			↑	↑
CD247	↓			
DPP4	↓			
IFNGR1				
SLAMF1	↑		↑	↑
RORC	↑	↑	↑	↓
EVL				↓
CLEC2D	↑			
LST1	↓			
RUNX3	↓			
TRG-AS1	↓			
PDCD1	↑	↑	↑	↑
LY6E	↓			
RORA				
CD63				

SPOCK2				↓
ZNF331				
CYTOR				
MT-ND4L				
ZBTB38	↓		↓	
MAF	↑			
TBXAS1				↑
BLK	↓			↓
CBLB	↑			
PIK3R1	↓			
GSTK1	↓			
SCART1	↓			
TTC39C				
GPR65				
PRNP	↑		↑	
IFI44			↑	
RHOB				↑
ADRB2	↓			↓
AGTRAP	↓			
ST3GAL1				
NPRL3				
GLRX	↓			
LIMD2				
OSTF1	↓			
PLAC8	↓			↓
ZBTB16				
CDC42EP3				
PIM1	↓			↑
MGAT4A	↓		↓	
MYBL1	↓			↑
RBL2	↓			↓
DNAJB6				
S1PR1	↓		↓	↓
RPL6				
TMSB4X				
SARAF				
RIPOR2				
FYN				
CD2				
LYAR				
PIK3IP1	↓			↓
S100A10				
LINC01871				
PARP8	↓			↓
ERN1	↓			
ZFP36L1	↑		↑	

ZC3H12A			↑	↑
KDM6B	↑		↑	
BIRC3	↑		↑	
PFKFB3	↑		↑	↑
AL138963.3				
LGALS1				↑
PDE4D				
CD27				↓
CAPG	↓			
MYL12A	↓			
RBMS1				
S100B	↓			
IDS				↓
RILPL2	↑			
JUND	↑		↑	
GNG2	↓		↑	↑
CREM	↑	↑		↑
ARHGAP9				↓
IL16				↓
TRAF3IP3	↓		↓	
ARL4C	↓			
FAM177A1				
LCP2				
DENND2D				
IFITM1	↓		↑	
TIMP1	↓			
IFITM2	↓			
MT-ND6				
B2M				
HLA-B				
HLA-A				
RPS2				
RPL17				
HSPA5	↑			↑
CD74				
SPON2	↓			
SNHG8				
NEU1	↑			
IER3	↑	↑	↑	↑
TRAF1	↑	↑		↑
NINJ1		↑	↑	
IRF8	↑	↑	↑	↑
NFKB2	↑	↑	↑	↑
NFKBID	↑	↑	↑	
NFKB1	↑	↑	↑	↑
IL2RA	↑	↑	↑	↑

BCL2A1	↑	↑	↑	↑
SEC14L1	↑		↑	
FAS				
FURIN	↑	↑	↑	↑
GABPB1	↑		↑	
KMT2E				
HIF1A	↑		↑	↑
MIR155HG	↑	↑	↑	↑
SLC7A5	↑	↑	↑	↑
IL21R	↑	↑	↑	↑
LTA	↑	↑	↑	↑
IL4I1	↑		↑	
BAZ1A				
SQSTM1				
TANK		↑	↑	↑
PIM2	↓	↑	↑	↑
DDX21	↑	↑	↑	↑
RAB8B	↑		↑	
BHLHE40		↑	↑	↑
GTF3C1				
TAP1			↑	↑
MTFP1				
SERPINB9	↑		↑	
ISG20			↑	↓
EIF5A	↑	↑	↑	↑
NAMPT	↑	↑	↑	↑
SDCBP	↑			
SRSF2				
CD83	↑	↑	↑	↑
TNIP1				
BCL2L1	↑	↑	↑	↑
HNRNPC				
HSP90B1				
PER1				
BZW1	↑	↑	↑	↑
DDX3Y				
POLR2K	↑			↑
GPBP1				
ANP32A				
TUBB	↑			
RBM39				
GPX4				
RAB9A				
SYAP1				
STAT4				↑
NDUFV2				↑

APOL3				
PKM	↑		↑	
ITK			↑	↑
AC025164.1				

↑ = expression increased in stimulated vs. unstimulated, ↓ = expression decreased in stimulated vs. unstimulated, ? = contradicting results in different datasets.

Table 5.8. List of unique clonotype markers, with clonotype markers that overlap with cluster markers indicated by an asterisk.

ABHD14B	*DPP4**	*LGALS3BP*	*RPS14*
AC007384.1	*DSEL*	*LIMS1*	*RPS18*
*AC007952.4**	*DUSP1**	*LINC01644*	*RPS3*
*AC016831.7**	*DUSP2**	*LINC02446**	*RPS4X*
AC020659.1	*DUSP4**	*LITAF**	*RPS6*
AC108134.3	*ECI2*	*LPAR6*	*RUNX2**
AC243829.4	*EGR1**	*LSP1**	*RUNX3**
*AC245014.3**	*EMP3**	*LST1**	*S100A11**
ADTRP	*EPAS1*	*LTB**	*S100A4**
*AGTRAP**	*EPHB6*	*LY6E**	*S100A6**
*AKIRIN2**	*EVL**	*LYAR**	*S100A9*
ALDOC	*EZR**	*M6PR*	*S100B**
ALOX5AP	*FAM173A*	*MAF**	*S1PR1**
*ANKRD28**	*FAM50B*	*MAL*	*S1PR5**
*ANXA1**	*FCMR*	*MALAT1**	*SAMHD1*
*ANXA2**	*FCRL6**	*MAP3K8**	*SARAF**
*AOAH**	*FEZ1**	*MCL1**	*SCART1**
*AQP3**	*FGFBP2**	*MFSD10**	*SECISBP2*
*AREG**	*FGFR1OP2*	*MS4A1**	*SELENOM*
ARHGAP15	*FLT3LG**	*MT-CO1*	*SELL**
ARID5B	*FOS**	*MT-ND4**	*SEPT9*
*ATF3**	*FTCD*	*MT-ND4L**	*SERTAD1**
*BATF**	*FTL**	*MVB12A*	*SIT1*
*BIN1**	*FXYD5**	*MXD3*	*SLA**
*BLK**	*FYB1**	*MYADM**	*SLA2**
*BTG2**	*GABARAPL1**	*NAP1L4*	*SLAMF1**
*C12orf75**	*GAPDH**	*NCAM1**	*SLC2A3**
C1orf162	*GBP5**	*NCF4*	*SLC38A1**
CACNA2D4	*GCHFR*	*NCR3**	*SLC38A2*
CALM1	*GLIPR1**	*NELL2*	*SLC4A10**
*CAPG**	*GLRX**	*NFKBIA**	*SMIM3**
*CASP1**	*GLUL**	*NFKBIZ**	*SMIM37*
CASS4	*GNLY**	*NINJ1**	*SMYD3*
*CCL3**	*GNPTAB**	*NKG7**	*SPART*
*CCL3L1**	*GPI*	*NOSIP**	*SPINT2*
*CCL4**	*GPR171**	*NPDC1*	*SPOCK2**
*CCL4L2**	*GPR55**	*NR4A2**	*SPRY1**
*CCL5**	*GRAP2*	*NRN1*	*SPRY2**
*CCR6**	*GSTM2*	*NSG1*	*SPTSSB*
CCSER2	*GSTP1**	*OASL*	*SRGN**
*CD160**	*GTDC1*	*OAZ1*	*SRRT*
*CD2**	*GTSF1*	*OSBPL10*	*SRSF5**
*CD247**	*GZMA**	*PABPC1**	*SRSF7**
*CD27**	*GZMB**	*PAGE5*	*ST3GAL1**
CD300A	*GZMH**	*PCNT*	*STIP1*
CD4	*GZMK**	*PDCD1**	*STK38*
*CD40LG**	*GZMM**	*PDE4D**	*STMN3*
CD5	*H3F3B**	*PECAM1*	*TARBP2*

*CD52**	*HCST*	*PIK3R1**	*TBXAS1**
*CD55**	*HLA-A**	*PIM2**	*THEMIS**
*CD63**	*HLA-DPA1**	*PITRM1*	*TIAM1*
*CD69**	*HLA-DPB1**	*PLCB1*	*TIGIT*
*CD7**	*HLA-DQB1*	*PLEK**	*TIMP1**
*CD74**	*HLA-DRB1**	*PMAIP1**	*TMEM107**
*CD84**	*HLA-E*	*PPIA*	*TMEM171*
*CD8A**	*HOPX**	*PPP1R15A**	*TMEM273*
*CD8B**	*HSPA1A**	*PPP1R2C*	*TMIGD2*
CD96	*HSPA1B**	*PRF1**	*TMSB10**
CD99	*IFNG**	*PRICKLE1**	*TNF**
CDKN2B	*IFNG-AS1*	*PRKACB**	*TNFAIP3**
CEBPD	*IFNGR1**	*PRSS21*	*TNFRSF18*
CFH	*IL10RA*	*PTGER4**	*TNFSF13B*
CH25H	*IL12RB2*	*PXN**	*TOB1**
CHRM3-AS2	*IL17A*	*RAB3IP*	*TOX**
CIB1	*IL23R**	*RBPJ*	*TPRKB*
*CISH**	*IL2RB**	*RCAN3*	*TRG-AS1**
*CLEC2D**	*IL32**	*RCBTB2*	*TSPO**
*CLIC3**	*IL6ST**	*REL**	*TSTD1*
*CMC1**	*IL7R**	*RGS1**	*TTC39C**
CMTM6	*IRF1**	*RGS18**	*TUBA1A**
*CNN2**	*ISG15*	*RHOB**	*TXK**
COIL	*ITGB1*	*RIN3**	*TXNIP**
COL6A2	*ITM2A**	*RNF185*	*TYROBP**
*COLQ**	*ITM2C**	*RORA**	*UBB**
*COTL1**	*ITPRIP*	*RORC**	*UNG*
CPNE2	*KDSR*	*RPL10*	*VILL*
*CRTAM**	*KIR3DL2*	*RPL11*	*VIM**
*CST7**	*KLF2**	*RPL12*	*XCL1**
*CTSA**	*KLF6**	*RPL13*	*XCL2**
CTSB	*KLRB1**	*RPL18A**	*YPEL5**
CTSW	*KLRC1**	*RPL19*	*Z93241.1**
*CXCR3**	*KLRC3**	*RPL23A*	*ZBTB38**
*CXCR6**	*KLRD1**	*RPL29*	*ZFP36**
*CYCS**	*KLRF1**	*RPL32**	*ZFP36L2**
*CYTOR**	*KLRG1**	*RPL6**	*ZNF331**
DKK3	*LDHB*	*RPLP0**	*ZNF652*
*DNAJB1**	*LGALS1**	*RPLP1*	*ZNF683**
DNAJC2	*LGALS3**	*RPS13**	*RPS14*

6. Delineation of the transcriptional response of human blood MAIT cells to TCR and cytokine stimulation at a single-cell level

6.1. Introduction

In Chapter 5, I identified an association between the single-cell resting transcriptional profile of MAIT cells from the blood and liver, and their clonal origin. The genes differentially expressed between TCR clonotypes were enriched for genes associated with MAIT cell activation, both TCR- and cytokine-mediated. Additionally, MAIT cell clonotypes varied in their expression of NK cell receptors. Expression of certain NK cell receptors, such as CD94 (*KLRD1*), is associated with an enhanced capacity of MAIT cells to respond to cytokine stimulation (Dias et al., 2017). Based on these findings, I hypothesise that the varying activation states of MAIT cell clonotypes reflect differences in their capacity for activation in response to TCR and/or cytokine stimuli, and thus in their history of *in vivo* activation. In this chapter, I have investigated this hypothesis through scRNA-seq and scTCR-seq of unstimulated, TCR-stimulated and cytokine-stimulated human blood MAIT cells.

While cytokine responsiveness has not previously been associated with MAIT cell TCR usage, there is some evidence for differences in ligand affinity between MAIT cell TCRs, as well as for altered MAIT cell activation in response to bacteria or riboflavin metabolites dependent on TCR usage. *TRBV6-1* MAIT cell TCRs with different CDR3β sequences show varying affinity for MR1/5-OP-RU and MR1/6-FP

(Eckle et al., 2014). While the differences in affinity for MR1/6-FP are pronounced, they are minor for the potent MR1/5-OP-RU antigen. Nevertheless, when transduced into SKW3 cells, TCRs with varying affinity for MR1/5-OP-RU induce differential CD69 upregulation in response to non-saturating concentrations of rRL-6-CH_2OH or 5-A-RU. This suggests a functional impact of small changes in MAIT cell ligand affinity. In another study, four distinct MAIT cell clones were shown to exhibit differential responsiveness to stimulation with the weak MAIT cell ligand RL-6,7-diMe, with only two clones producing IFNγ in response to this ligand (Gold et al., 2014).

As well as differences in their responsiveness to purified ligands, MAIT cells with distinct TCRs show differential activation by bacteria both *in vitro* and *in vivo*. MAIT cells with distinct Vβ segments show differences in IFNγ and TNFα production following *in vitro* stimulation with *E. coli* and *C. albicans* (Dias et al., 2017). The proliferative capacity of MAIT cells in response to bacterial stimulation is also associated with their Vβ usage (Dias et al., 2017). *In vitro* stimulation of $CD8^+$ T cells with different bacteria (*M. smegmatis*, *S. typhimurium*, *C. albicans*) activates populations of MAIT cells with distinct TCR repertoires, suggesting MAIT cell TCRs have some ligand or pathogen specificity (Gold et al., 2014). During human *in vivo* infection with *S.* Paratyphi A, the MAIT cell clonal landscape changes, with some clones expanding and others contracting (Howson et al., 2018). Jurkat MAIT cell lines expressing recombinant TCRs containing the CDR3β sequences of the expanded or contracted clonotypes show comparatively increased or decreased activation, respectively, in response to bacterial supernatant or riboflavin metabolites. This differential activation is observed irrespective of the bacteria used for stimulation, namely *S.* Paratyphi A or *E. coli*. Therefore, the enhanced responsiveness of the expanded clonotypes relative to the contracted clonotypes is

not pathogen specific. This contrasts with *in vitro* data suggesting pathogen specificity in the MAIT cell TCR repertoire (Gold et al., 2014), as discussed above.

Another question that remains to be addressed within the MAIT cell field is whether MAIT cells comprise functionally distinct subsets in humans. In mice, MAIT cells comprise an abundant RORγt^{+} IL-17-secreting type 17 subset and a minor subset of T-bet^{+} IFNγ-secreting type 1 cells (Rahimpour et al., 2015). These populations bifurcate within the thymus from a shared precursor population, as demonstrated through pseudotime analysis of scRNA-seq data from MR1/5-OP-RU^{+} mouse thymic T cells (Koay et al., 2019a; Legoux et al., 2019b). In contrast, human MAIT cells coexpress RORγt and T-bet (Leeansyah et al., 2015) and distinct subsets have not been identified under homeostatic conditions. This is supported by my scRNA-seq data in Chapter 5, suggesting that transcriptional clusters of MAIT cells reflect different states of activation rather than distinct differentiated subsets. Nevertheless, it remains possible that subsets of MAIT cells show distinct functional capacity following activation. For example, it is unclear why only a fraction of MAIT cells produce IL-17 upon PMA and ionomycin stimulation (Dusseaux et al., 2011), despite all MAIT cells expressing RORγt (Leeansyah et al., 2015). Moreover, using mass cytometry, Lepore and colleagues identified populations of MAIT cells producing distinct combinations of cytokines following PMA and ionomycin stimulation (Lepore et al., 2017), although whether this indicates the presence of bona fide MAIT cell subsets with fundamentally distinct functional capacity is unclear.

The genes altered in expression upon TCR or cytokine stimulation of MAIT cells have been characterised in detail, particularly in three recent publications (Hinks et al., 2019; Lamichhane et al., 2019b; Leng et al., 2019). However, these examined

gene expression by bulk RNA-seq at a single time point. Consequently, they were not able to assess the kinetics of gene expression following MAIT cell activation. I hypothesised that by performing a similar single time point experiment using scRNA-seq, I would capture individual MAIT cells at different stages of activation and thus would be able to investigate time-dependent changes in gene expression upon TCR- and cytokine-driven activation.

6.1.1. Summary and unanswered questions

In summary, my data from Chapter 5 indicates an association between the activation state of MAIT cells and their clonal origin. There is some evidence in the literature for differential activation of MAIT cells depending on their TCR usage. However, previous studies have examined small numbers of TCRs, often in the context of transduced cell lines, and/or have lacked paired TCRαβ sequencing data, which is essential for defining clonotypes. Therefore, the relationship between MAIT cell clonal origin and TCR ligand responsiveness remains unclear. The cytokine responsiveness of different MAIT cell clonotypes has yet to be examined. While human MAIT cells do not appear to comprise multiple subsets under homeostatic conditions, it remains possible that there are subsets of MAIT cells with distinct functional capacity. Finally, the genes altered following TCR- and/or cytokine-dependent MAIT cell activation are well-described, but the kinetics of gene expression following activation remain poorly characterised.

6.2. Aim

To characterise the pathway(s) of MAIT cell activation following TCR or cytokine stimulation, and to establish the association between the clonal origin of MAIT cells and their capacity for activation.

6.3. Key questions

1. Do different MAIT cell clonotypes show altered capacity for activation following TCR or cytokine stimulation?
2. Do MAIT cells comprise functionally distinct subsets?
3. How does gene expression vary along the trajectory/trajectories from non-activated to activated MAIT cells?

6.4. Experimental design (Experiment 2)

Enriched peripheral blood $CD8^+$ T cells from three donors were incubated with media (hereafter unstimulated), plate-bound MR1/5-OP-RU + αCD28 (hereafter TCR-stimulated) or IL-12 and IL-18 (hereafter cytokine-stimulated) for 20 hours. MAIT cells were sorted and analysed by scRNA-seq and scTCR-seq using the droplet-based Chromium Single Cell Immune Profiling Solution from 10x Genomics (protocol outlined in Fig. 6.1A).

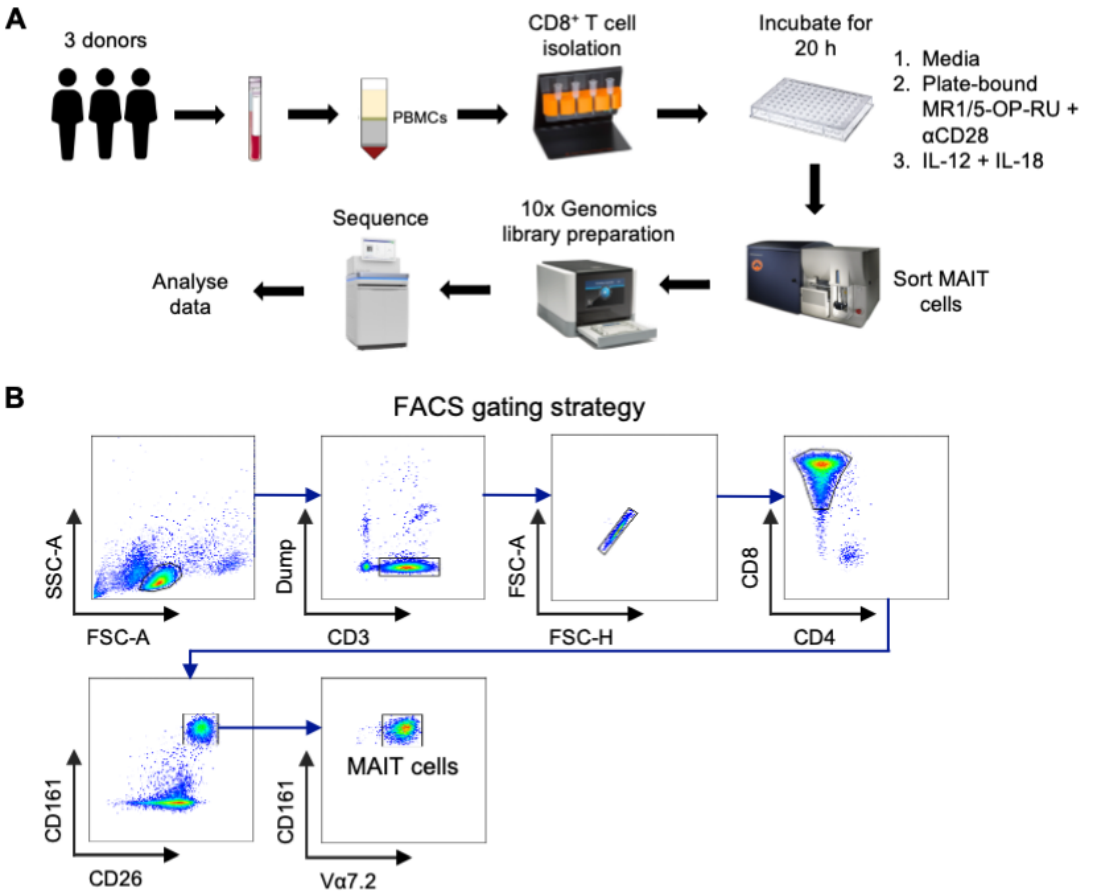


Figure 6.1. Experimental workflow. (A) Schematic diagram illustrating the experimental protocol. Blood was collected from three donors, PBMCs isolated and $CD8^+$ T cells enriched by magnetic-activated cell sorting (MACS). Isolated $CD8^+$ T cells were incubated for 20 h with RPMI media, plate-bound MR1/5-OP-RU + αCD28, or IL-12 + IL-18. MAIT cells were isolated by FACS according to the gating strategy shown in (B). Finally, scRNA-seq and scTCR-seq libraries were generated (10x Genomics technology) and sequenced, ready for analysis. **(B)** FACS gating strategy. Dump = anti-CD19, anti-CD14, anti-TCR γδ, anti-TCR Vδ2, anti-TCR Vα24-Jα18, SYTOX Green Nucleic Acid Stain.

6.5. Results

6.5.1. Quality control

Following stimulation, MAIT cells were sorted as Vα7.2^{+}CD161^{hi}CD26^{+}CD8^{+} T cells for combined scRNA-seq and scTCR-seq (Fig. 6.1B). The resulting single-cell data was filtered to remove cells with low UMI counts, low gene counts and/or a high fraction of mitochondrial reads (Fig. 6.2A-H). Following MAIT cell activation, the average number of UMIs and genes per cell increased, along with the range of UMI and gene counts (Fig. 6.2A-C,E-G). Consequently, cell multiplets could not be removed from the data based on their expected high UMI counts, as this approach would have biased towards the removal of activated MAIT cells. A fraction of cell multiplets (predominantly doublets) will therefore be present within the data. The proportion of cell multiplets increases linearly with loading cell concentration. Given that we loaded ~6000-8000 MAIT cells per channel of the Chromium Controller, the multiplet rate should be in the range 2.3-3.9% per sample (10x Genomics technical information). As only a single cell type was loaded per lane, these doublets will be homotypic doublets (i.e. doublets containing two cells of the same type). This precludes the use of computational doublet removal algorithms that identify heterotypic doublets (i.e. doublets containing two cells of a different type) by their hybrid transcriptional profile (DePasquale et al., 2019; McGinnis et al., 2019a; Wolock et al., 2019). Despite the inclusion of some homotypic doublets, their impact on data interpretation should be minor (DePasquale et al., 2019; McGinnis et al., 2019a; Wolock et al., 2019). While heterotypic doublets can comprise a distinct cluster, suggesting the presence of an intermediate cell state, homotypic doublets

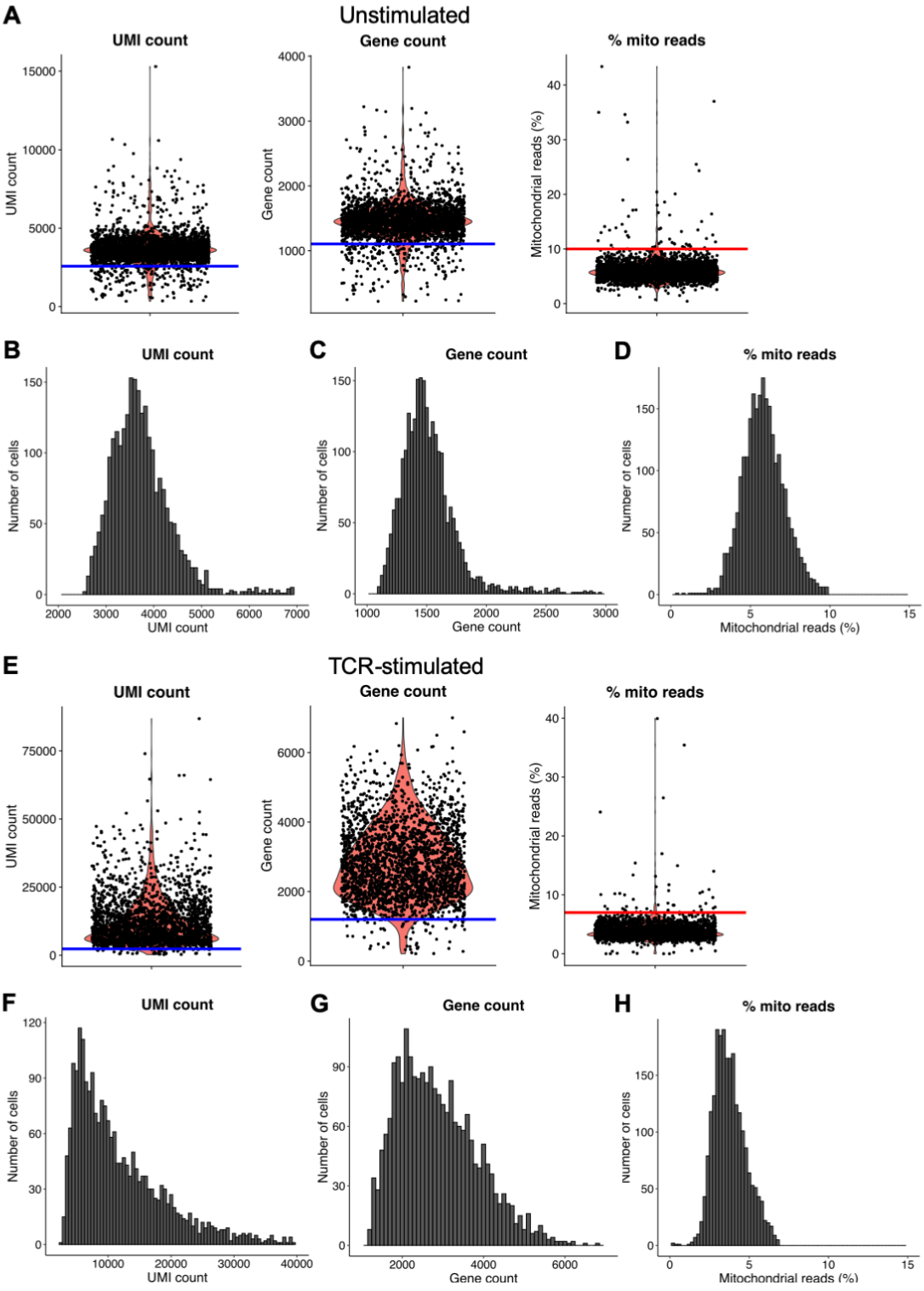


Figure 6.2. Removal of poor quality cells with a low UMI count, low gene count and/or high percentage of mitochondrial reads from single-cell data. (A,E) Violin plots showing the pre-filtering distributions of UMI counts (left), gene counts (middle) and percent mitochondrial reads (right) for a representative unstimulated (A) and TCR-stimulated (E) sample. Lower and upper filtering thresholds are indicated by blue and red lines, respectively. **(B-D,F-H)** Histograms showing the post-filtering distributions of UMI counts (B,F), gene counts (C,G) and percent mitochondrial reads (D,H) for a representative unstimulated (B-D) and TCR-stimulated (F-H) sample.

are transcriptionally similar to their composite singlet cells and typically result in only small quantitative effects on gene expression and cluster abundance.

6.5.2. Comparison of the gene expression and TCR repertoire of unstimulated MAIT cells with blood MAIT cells from Experiment 1

The genes differentially expressed between unstimulated MAIT cell clusters are enriched for cytokine signalling pathways

Unstimulated MAIT cells comprised multiple clusters (5-7) in each donor (Fig. 6.3A-C), consistent with blood MAIT cells analysed directly *ex vivo* in Experiment 1 (Fig. 5.5B). Despite examining the same number of donors and the samples comprising a similar number of clusters in the two experiments, the total number of cluster markers (genes differentially expressed between each cluster and all other cells) identified in Experiment 2 was considerably larger than in Experiment 1 (Fig. 6.3D). Cluster markers combined from the three unstimulated samples showed relatively low (11%) overlap with the cluster markers from blood MAIT cells in Experiment 1, although 44% of the cluster markers identified in Experiment 1 were again identified in Experiment 2 (Fig. 6.3D). Variation in the cluster markers from the two experiments could be explained by the different experimental setups. While MAIT cells in Experiment 1 were sequenced immediately post-thawing, unstimulated MAIT cells in Experiment 2 were rested for 20 hours prior to sequencing. Regardless of the differences in the cluster marker lists from the two experiments, Reactome pathway enrichment analysis again identified cytokine and TCR signalling pathways as a major component of resting MAIT cell transcriptional heterogeneity, as in Experiment 1 (Fig. 6.3E and 5.8A). However, while TCR signalling pathways were

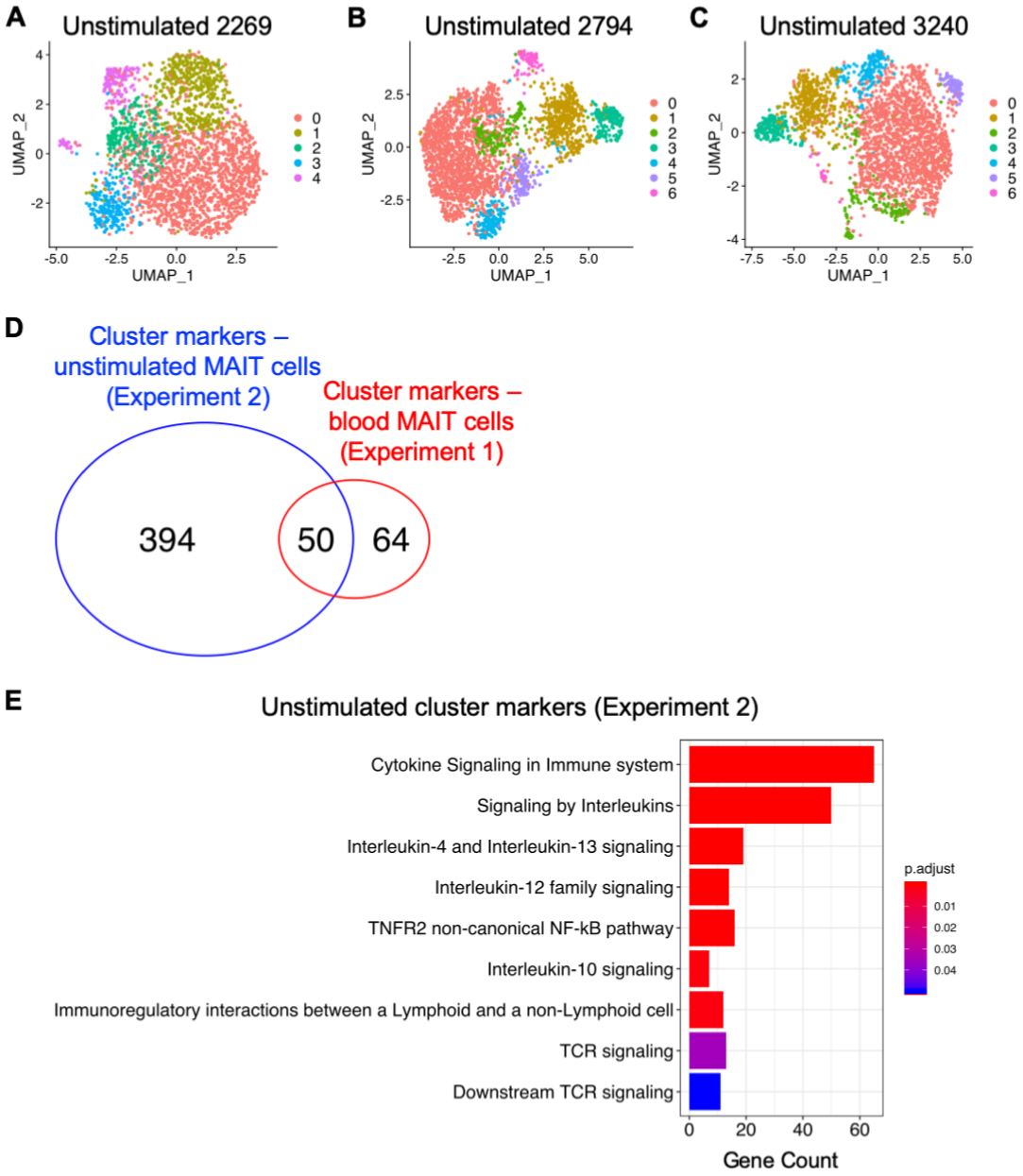


Figure 6.3. Unstimulated MAIT cells comprise multiple clusters that vary in the expression of cytokine and TCR signalling pathways. (A-C) Clusters in the three samples of unstimulated MAIT cells as visualised by UMAP. Plots show donors 2269 (A), 2794 (B) and 3240 (C). **(D)** Venn diagram illustrating the overlap of cluster markers from unstimulated MAIT cells (combined from the three donors) with the cluster markers from blood MAIT cells in Experiment 1. **(E)** Reactome pathway enrichment analysis of the cluster markers combined from the three samples of unstimulated MAIT cells. Selected enriched pathways are shown.

most enriched amongst blood MAIT cell cluster markers in Experiment 1 (Fig. 5.8A), cytokine signalling pathways were dominant in Experiment 2 (Fig. 6.3E).

Consistent with the enrichment of cytokine and TCR signalling pathways amongst cluster markers, suggesting differences in MAIT cell activation state between the clusters, 61% of unstimulated cluster markers overlapped with genes differentially expressed in TCR- or cytokine-stimulated MAIT cells (Fig. 6.4A; lists of differentially expressed genes were obtained from three recent bulk RNA-seq studies (Hinks et al., 2019; Lamichhane et al., 2019b; Leng et al., 2019)). Of the activation-associated cluster markers, 57% were altered in expression upon cytokine stimulation, 89% upon TCR stimulation (6 h or 24 h) and 43% upon *E. coli* stimulation (Fig. 6.4B). Therefore, similar to Experiment 1, a large proportion of the genes with variable expression amongst unstimulated MAIT cells were associated with MAIT cell activation, both TCR- and cytokine-mediated. This may reflect differential *in vivo* activation histories of the cells, as hypothesised in Chapter 5.

A *GNLY*$^{+}$ MAIT cell cluster is present in all donors

When analysing the samples of unstimulated MAIT cells, I noticed that *GNLY* showed a distinct expression pattern. In all donors, *GNLY*$^{+}$ MAIT cells were almost entirely restricted to a single cluster (Fig. 6.5A-C) and *GNLY* was a positive marker for this cluster (i.e. upregulated in this cluster compared with all other cells; Fig. 6.5D-F). In one donor (3240; Fig. 6.5F), *GNLY* was the only positive marker for this cluster.

Given this finding, I decided to examine the blood samples from Experiment 1 to see whether a *GNLY*$^{+}$ MAIT cell cluster was also present in these donors. The results were similar to Experiment 2, with *GNLY* marking a single cluster within each donor

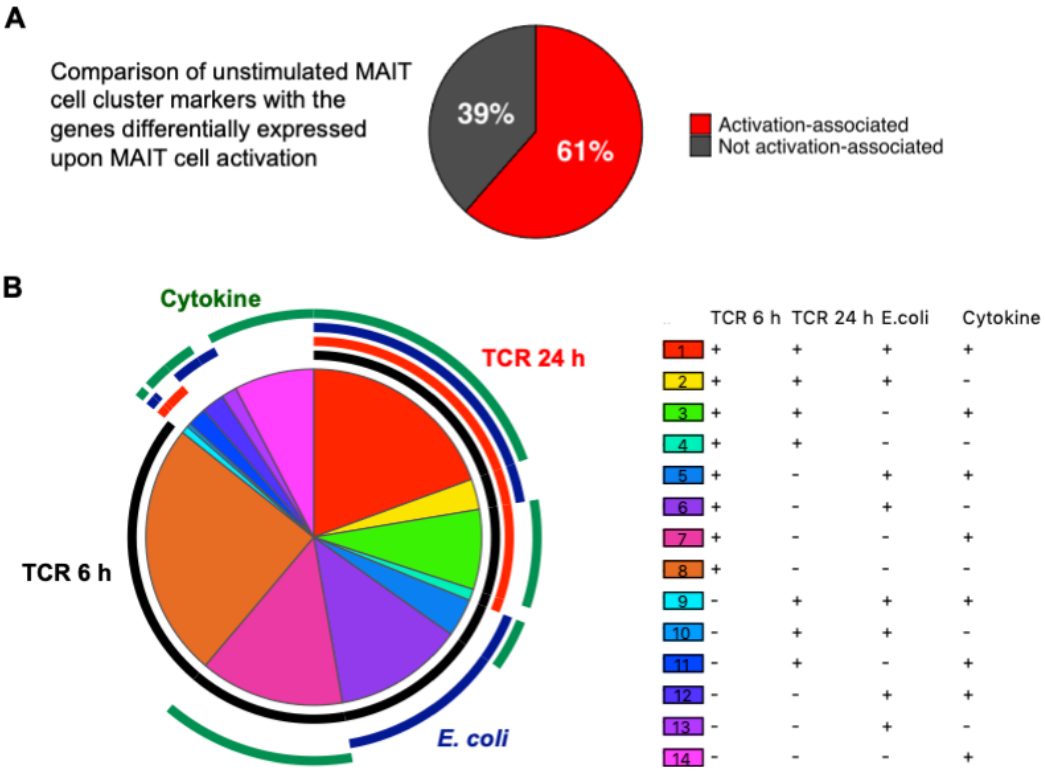


Figure 6.4. Clusters of unstimulated MAIT cells differentially express genes associated with MAIT cell activation. Comparison of unstimulated MAIT cell cluster markers with the genes differentially expressed following MAIT cell activation in three published RNA-seq studies (Hinks et al., 2019; Lamichhane et al., 2019b; Leng et al., 2019). Figure 5.9 explains how this comparison was performed for cluster markers in Experiment 1. **(A)** Percentage of cluster markers associated with MAIT cell activation. **(B)** Pie chart showing the fraction of activation-associated cluster markers that were differentially expressed in response to TCR (6 h or 24 h), *E. coli*, and/or cytokine stimulation.

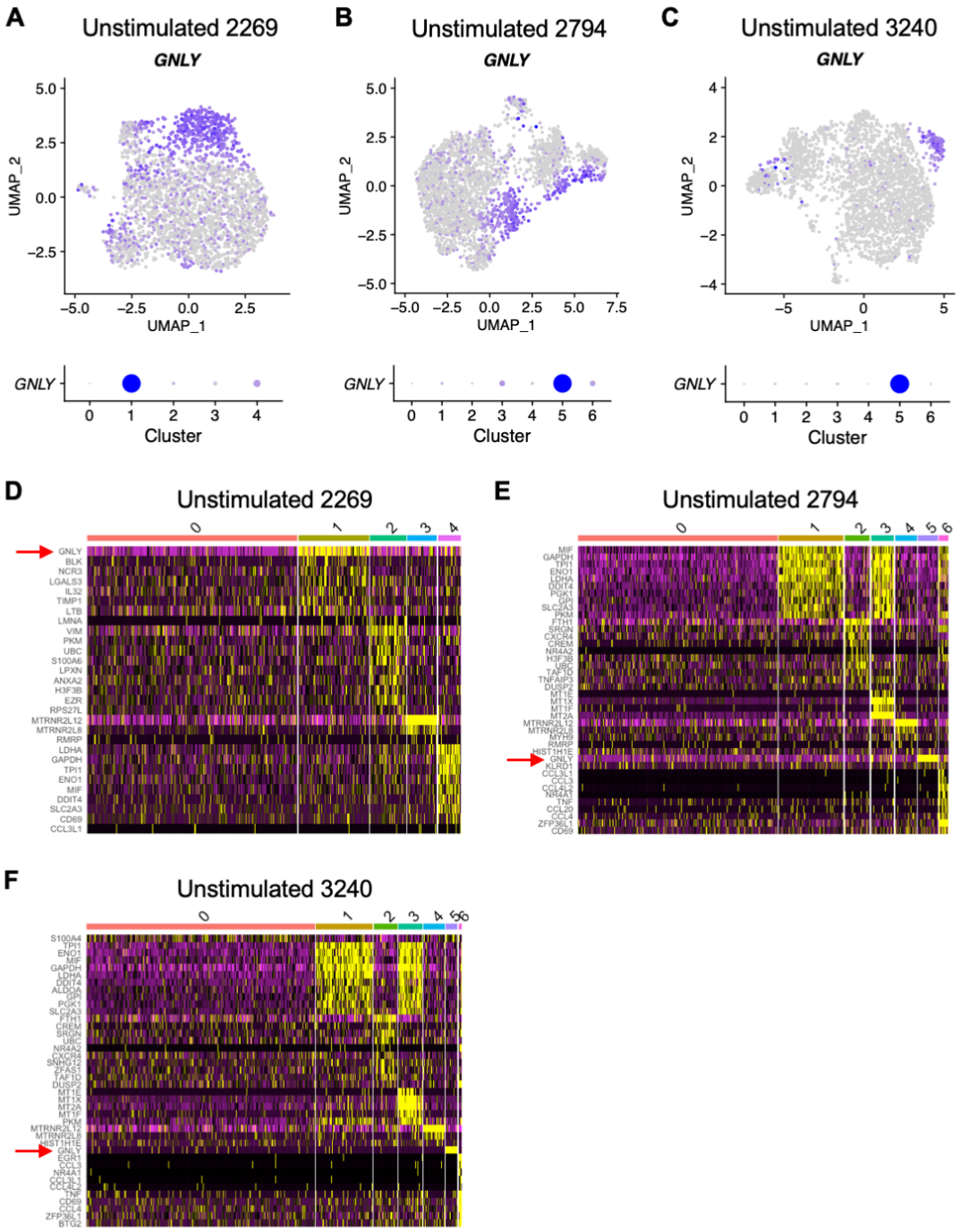


Figure 6.5. A *GNLY*$^{+}$ MAIT cell cluster is present in all donors. (A-C) Expression of *GNLY* by unstimulated MAIT cells from donors 2269 (A), 2794 (B) and 3240 (C). Top: UMAP, bottom: dot plot. **(D-F)** Heatmaps showing the top 10 positive markers for each cluster of unstimulated MAIT cells from donors 2269 (D), 2794 (E) and 3240 (F).

(Fig. 6.6A-C). Again, *GNLY* was a positive marker for these clusters (Fig. 6.6D-F) and in one donor (GI3008 Blood) was the sole positive marker for the cluster (Fig. 6.6D).

Localisation of TCR clonotypes amongst MAIT cell clusters

In Experiment 1, I identified a number of instances where MAIT cell TCR clonotypes were biased towards localisation within particular transcriptional clusters. I therefore examined whether a similar relationship between clonotype and cluster could be identified in unstimulated MAIT cells from Experiment 2. In contrast with Experiment 1, the majority of TCR clonotypes appeared randomly distributed across clusters (Fig. 6.7A-C). However, individual clonotypes with highly specific cluster localisation could be identified. For example, clonotype 6 from donor 2794 was almost exclusively restricted to cluster 6 (Fig. 6.7D). Furthermore, while 27% of cells within cluster 6 belonged to clonotype 6, 80% of cells within this cluster shared an identical *TRAV1-2*/*TRAJ12* TCRα chain paired with various TCRβ chains. Consistent with the differential expression of activation-associated genes between clonotypes in Experiment 1, cells within cluster 6 were more activated than cells within the other clusters (Fig. 6.7E), displaying increased expression of chemokines (e.g. *CCL3*, *CCL4*, *CCL20*), cytokines (e.g. *TNF*, *LTB*) and activation markers (e.g. *CD69*). Visualising the localisation of the largest 20 MAIT cell clonotypes across the clusters within each sample revealed a couple of other clonotypes with seemingly biased localisation to particular clusters, such as clonotype 7 in donors 2269 and 2794, and clonotype 8 in donor 3240 (Fig. 6.8A-C). When interpreting these plots, it is important to bear in mind that some of the clonotypes contain very few cells (e.g. clonotypes 13-20 from unstimulated 2269 each contain 7-8 cells), and thus highly biased cluster localisation can result from the presence of just a few cells in a particular cluster.

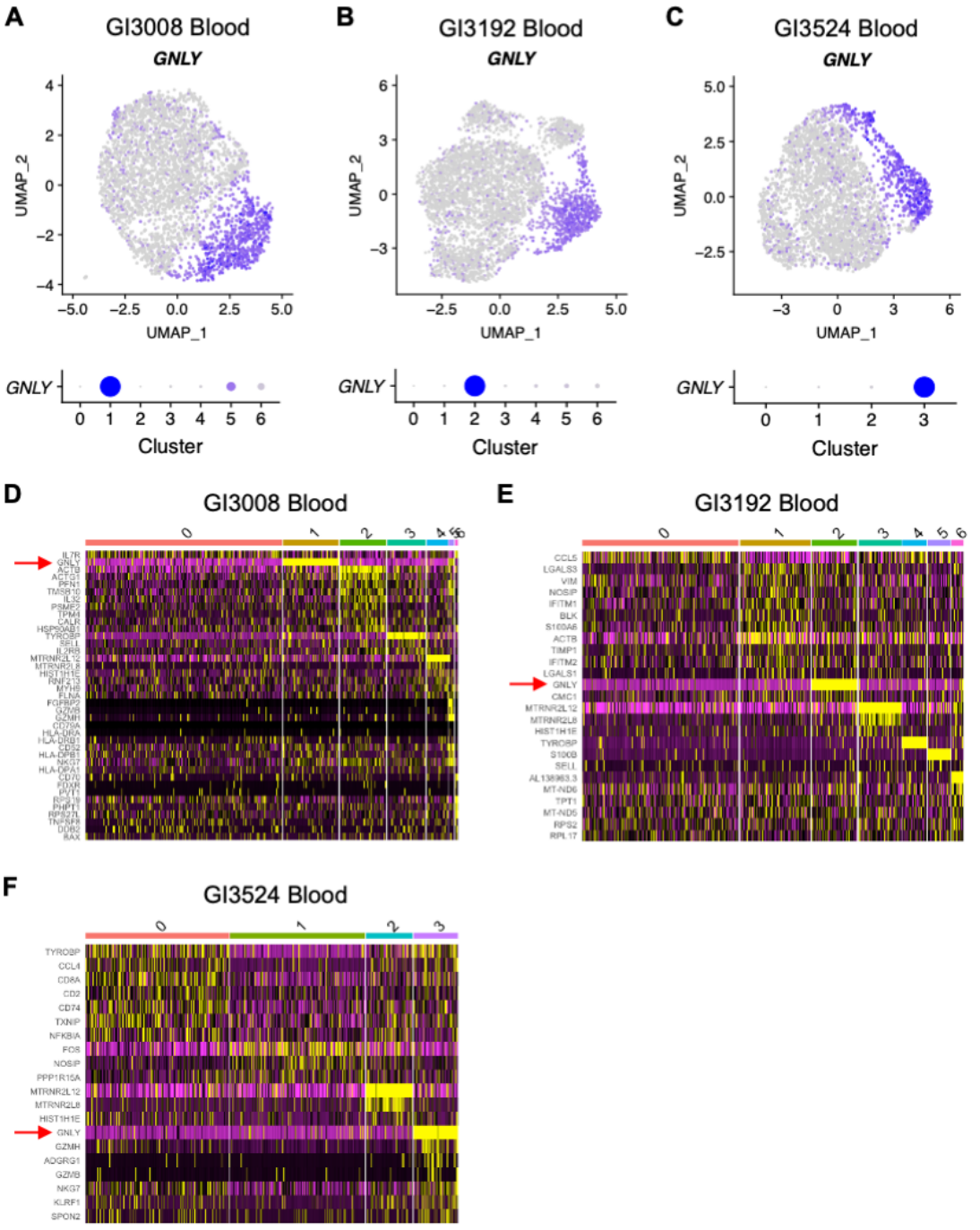


Figure 6.6. A *GNLY*$^{+}$ MAIT cell cluster was present in all blood samples from Experiment 1. (A-C) Expression of *GNLY* by blood MAIT cells from donors GI3008 (A), GI3192 (B) and GI3524 (C). Top: UMAP, bottom: dot plot. **(D-F)** Heatmaps showing the top 10 positive markers for each cluster of blood MAIT cells from donors GI3008 (D), GI3192 (E) and GI3524 (F).

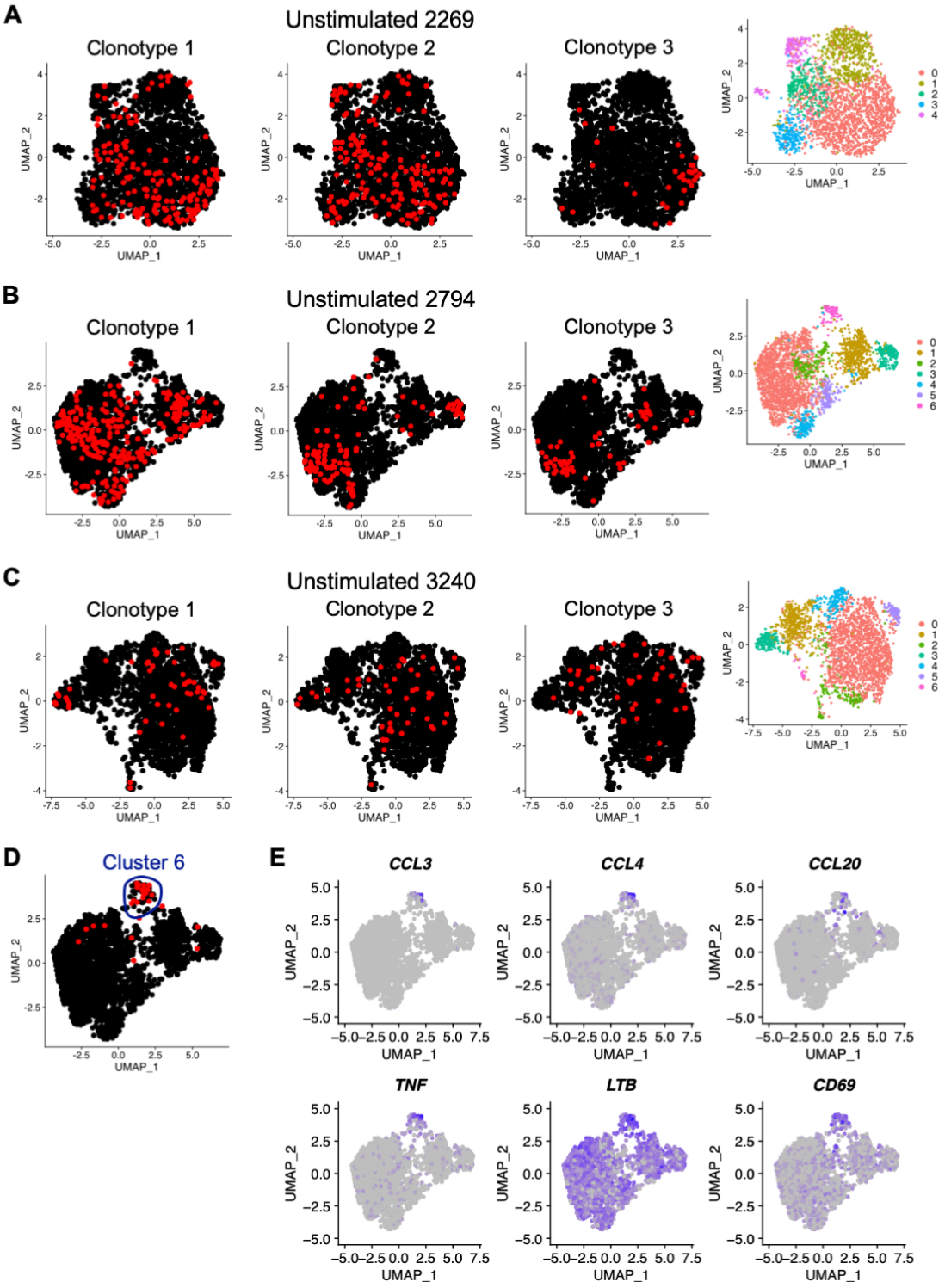


Figure 6.7. Cluster localisation of TCR clonotypes from the three samples of unstimulated MAIT cells. (A-C) Cluster localisation of the three largest nucleotide clonotypes from each sample of unstimulated MAIT cells. Plots show donors 2269 (A), 2794 (B) and 3240 (C). The three UMAP plots on the left of (A-C) each show the localisation of a single TCR clonotype in red, while the right UMAP is a reminder of the transcriptional clusters identified in the corresponding sample. **(D)** The sixth largest clonotype from donor 2794 was mostly restricted to a single cluster (cluster 6). **(E)** UMAP plots showing the expression of example genes upregulated in cluster 6 of donor 2794.

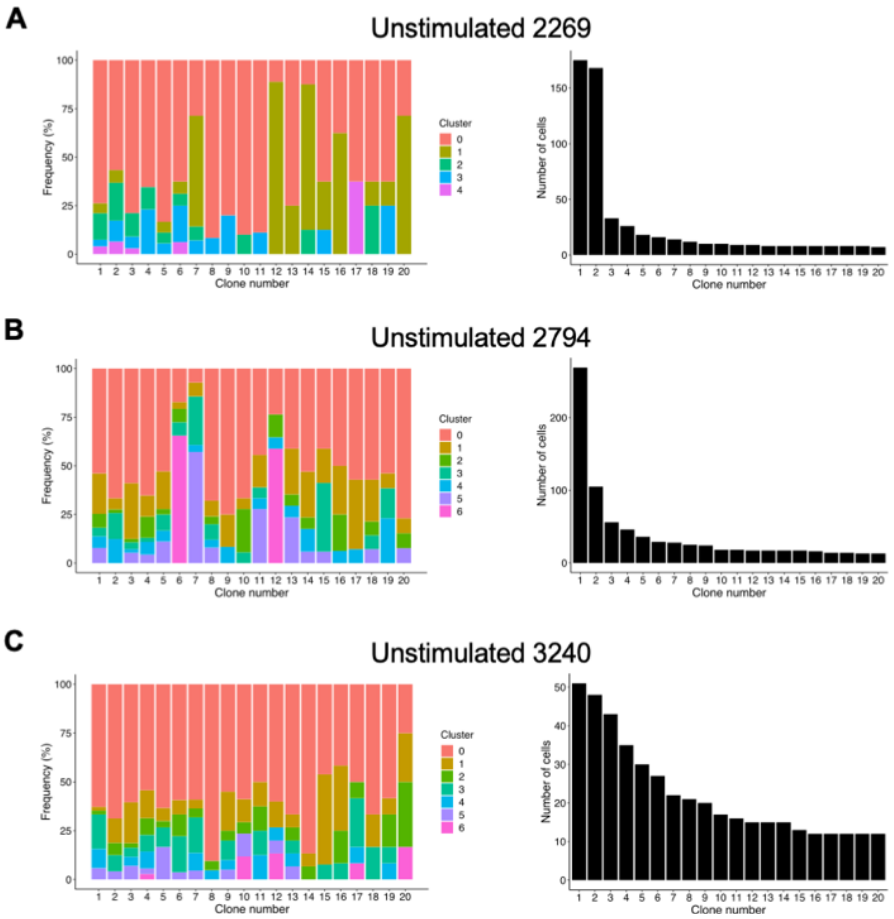


Figure 6.8. Cluster localisation of the top 20 largest TCR clonotypes from each sample of unstimulated MAIT cells. (A-C) Left – bar chart showing the percentage of cells from each of the top 20 largest clonotypes within each MAIT cell cluster. Right – number of cells in each of the top 20 largest clonotypes. Plots show donors 2269 (A), 2794 (B) and 3240 (C).

Clonotype markers are enriched for NK cell receptors and genes associated with cell activation

Despite the majority of clonotypes appearing unbiased in their cluster localisation, pathway enrichment analysis on a pooled list of unstimulated MAIT cell clonotype markers (i.e. genes differentially expressed between each clonotype and all other cells) revealed an enrichment for cytokine signalling pathways and genes associated with lymphoid to non-lymphoid cell interactions (predominantly NK cell receptors; Fig. 6.9A; Table 6.1), similar to blood MAIT cell clonotype markers from Experiment 1 (Fig. 5.22A). However, in contrast with Experiment 1, TCR signalling pathways were not enriched. In total, 32 clonotype markers were identified in both experiments (Fig. 6.9B), including cytokines (e.g. *IL32*, *LTB*), cytotoxic molecules (e.g. *GNLY*, *GZMK*) and NK cell receptors (e.g. *KLRD1*, *KLRF1*). Clonotype markers exhibited a 64% overlap with genes differentially expressed upon MAIT cell activation (Figure 6.9C; gene lists obtained from three published bulk RNA-seq studies (Hinks et al., 2019; Lamichhane et al., 2019b; Leng et al., 2019)). Therefore, whilst there was generally no clear association between TCR clonotypes and clusters in unstimulated MAIT cells from Experiment 2 (Fig. 6.7A-C and 6.8), clonotypes did vary in the expression of activation-associated genes and NK cell receptors, similar to Experiment 1. Therefore, it remained possible that MAIT cell clonotypes may show a differential capacity for activation in response to TCR or cytokine stimulation, as hypothesised based on my findings in Chapter 5.

The general features of the TCR repertoire are comparable with Experiment 1

I next characterised the TCR repertoire of unstimulated MAIT cells from the three donors. As the donors in this experiment (Experiment 2) were younger than those

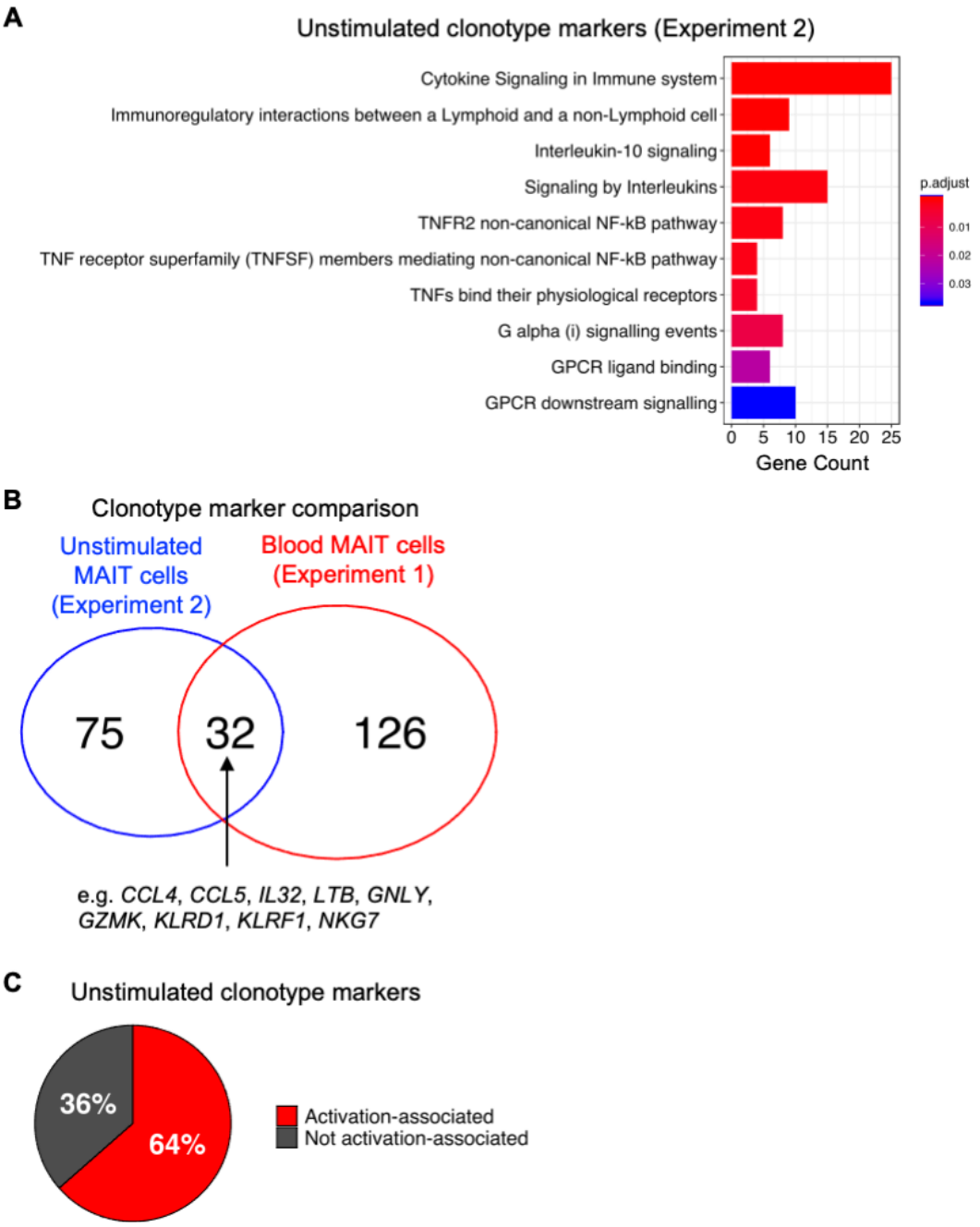


Figure 6.9. Clonotype markers are enriched for cytokine signalling pathways and NK cell receptors. (A) Reactome pathway enrichment analysis of clonotype markers combined from the three samples of unstimulated MAIT cells. Selected enriched pathways are shown. **(B)** Venn diagram showing the overlap of clonotype markers between unstimulated MAIT cells (blue) and blood MAIT cells from Experiment 1 (red). **(C)** Pie chart showing the overlap of unstimulated MAIT cell clonotype markers with the genes differentially expressed between stimulated (6 h TCR, 24 h TCR, *E. coli* or cytokine) and unstimulated MAIT cells in three published bulk RNA-seq datasets (Hinks et al., 2019; Lamichhane et al., 2019b; Leng et al., 2019). Figure 5.9A,B explains how this comparison was performed for cluster markers in Experiment 1.

in Experiment 1 (aged 26-31 vs. 50-66) and did not have any underlying health conditions, this allowed me to confirm that my previous findings were not an artefact of donor age or health status. Overall, the results of Experiment 2 were highly similar to Experiment 1, although a few minor differences were noted.

The TCR repertoire was oligoclonal (Fig. 6.10A) and consistent with the semi-invariant nature of the MAIT cell TCRα chain, TCRα only clonotypes were more oligoclonal than TCRβ only clonotypes (Fig. 6.10B-D). However, the degree of oligoclonality in this experiment was significantly reduced compared with Experiment 1 (Fig. 6.10E).

The majority of MAIT cells used a *TRAV1-2*/*TRAJ33* TCRα chain with a smaller fraction using a *TRAV1-2*/*TRAJ12* (8.8-11.1%) or *TRAV1-2*/*TRAJ20* TCRα (3.4-6.7%; Fig. 6.11A). The CDR3α region was 12 amino acids in length in essentially all MAIT cells (Fig. 6.11B) and had a highly conserved sequence (Fig. 6.11C), with variation at specific positions dependent on *TRAJ* gene usage (Fig. 6.11D).

MAIT cells showed biased use of *TRBV6-1*, *TRBV6-4* and *TRBV20-1* in all donors (Fig. 6.12A-D), consistent with published studies (Lepore et al., 2014; Reantragoon et al., 2013; Tilloy et al., 1999) and my data from Experiment 1 (Fig. 5.13). Some additional *TRBV* segments were enriched to a variable degree in all donors, particularly *TRBV4-2*, *TRBV6-3*, and *TRBV28* (Fig. 6.12A-D). CDR3β length showed greater variability than CDR3α length, with an overall similar distribution of lengths across donors (Fig. 6.12E). In contrast with the CDR3α sequence, the CDR3β sequence was highly diverse (Fig. 6.12F-H).

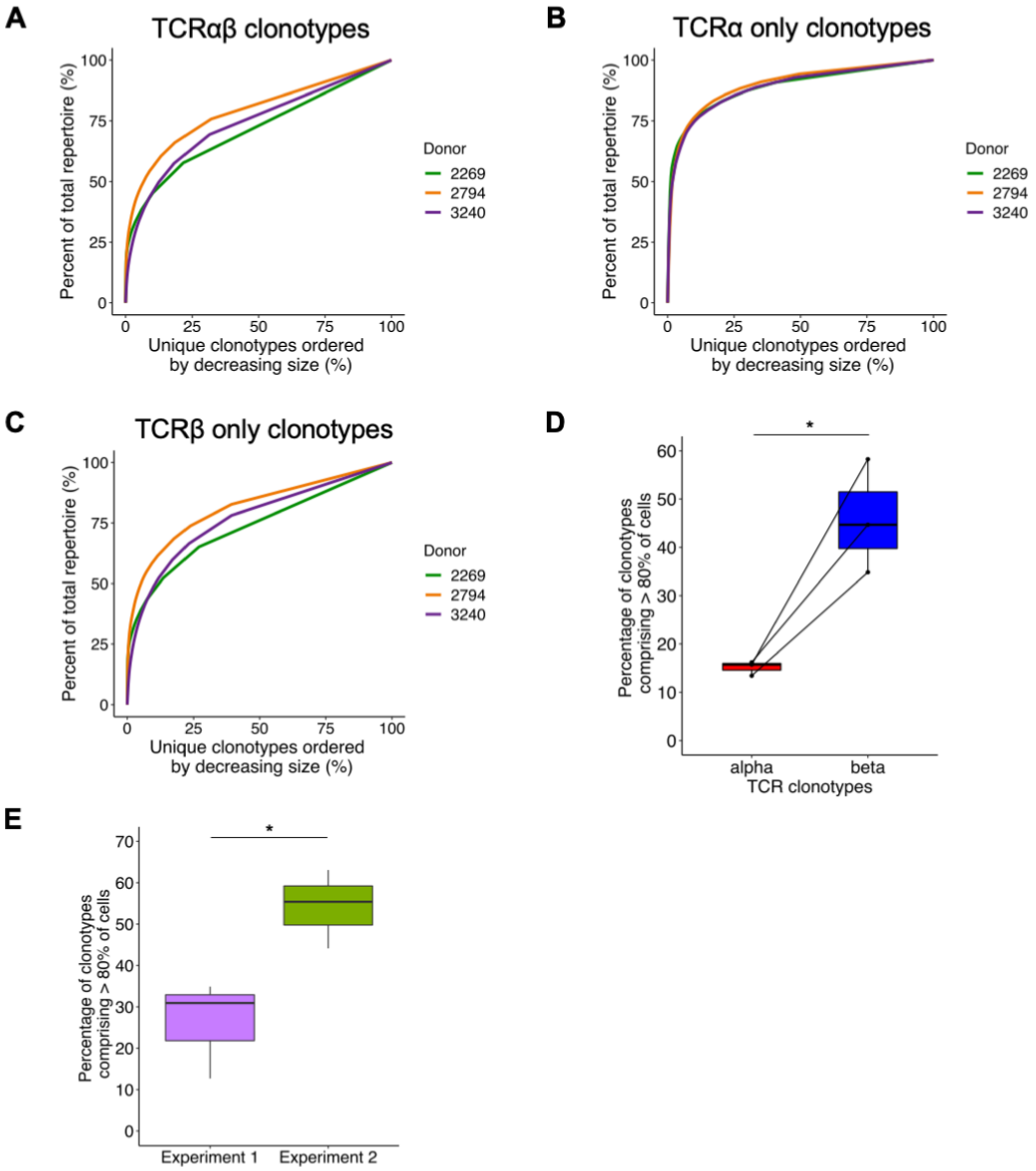


Figure 6.10. The MAIT cell TCR repertoire is oligoclonal, but shows reduced oligoclonality compared with Experiment 1. (A-C) Line graphs illustrating the oligoclonality of unstimulated MAIT cell TCRαβ (A), TCRα only (B) and TCRβ only (C) nucleotide clonotypes. **(D)** Box plot showing the percentage of unique TCRα only (red) or TCRβ only (blue) clonotypes comprising > 80% of cells. **(E)** Box plot showing the percentage of unique TCRαβ clonotypes comprising > 80% of cells in blood samples from Experiment 1 (purple) and in unstimulated MAIT cells from Experiment 2 (green). Two-tailed paired sample t-test (D) and two-tailed independent samples t-test (E). * $p < 0.05$.

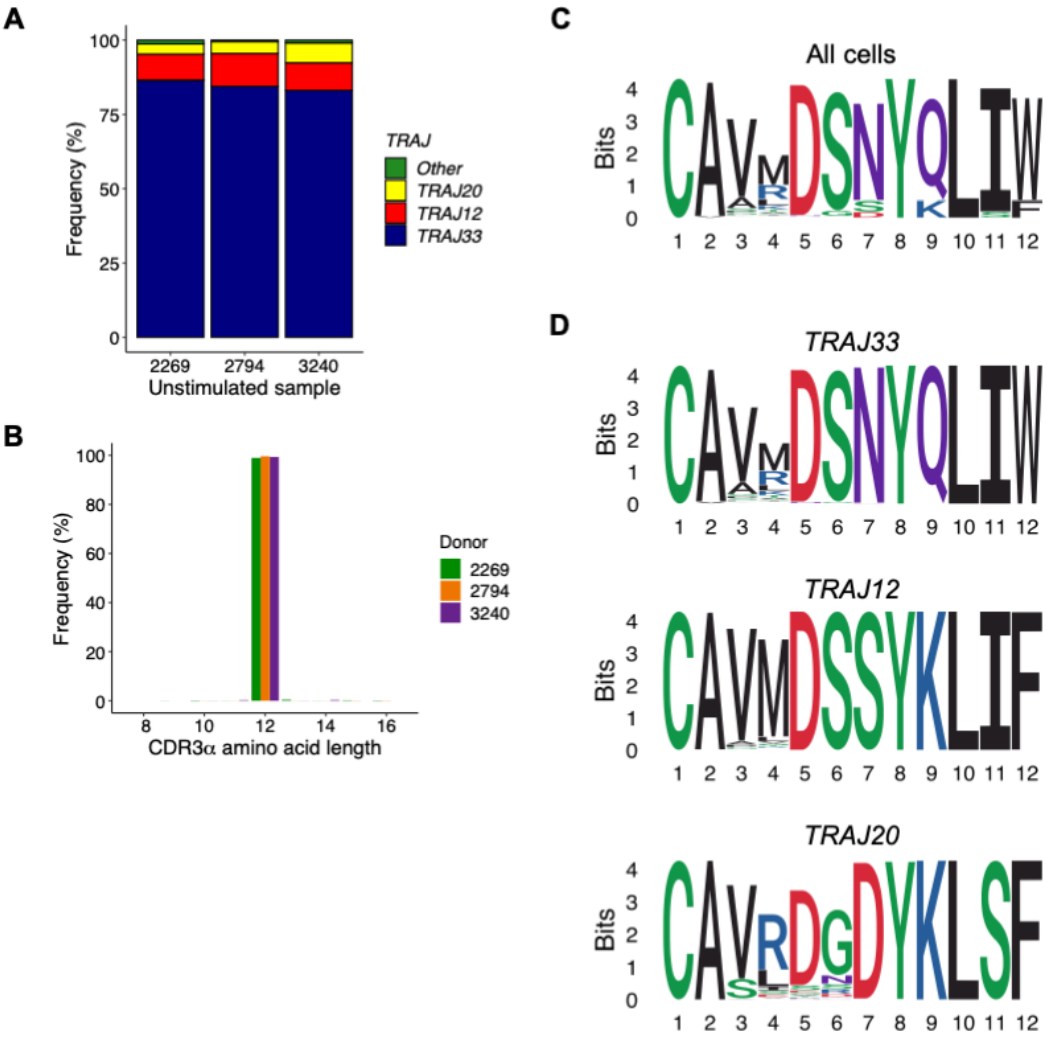


Figure 6.11. Features of the MAIT cell TCRα chain repertoire. (A) Percentage use of *TRAJ33*, *TRAJ12*, *TRAJ20* and other *TRAJ* segments amongst *TRAV1-2* TCRs in the three unstimulated samples. **(B)** CDR3α amino acid length in each unstimulated sample. **(C,D)** Sequence logos generated from *TRAV1-2* CDR3α amino acid sequences of length 12, combined from all unstimulated samples. Plots show all CDR3α sequences (C) or CDR3α sequences split by *TRAJ* gene usage (D). The overall height of the stacked letters at each position indicates the sequence conservation, while the relative abundance of each amino acid is indicated by the height of individual letters within the stack. Acidic bases are shown in red, basic residues in blue, hydrophobic in black and polar in green.

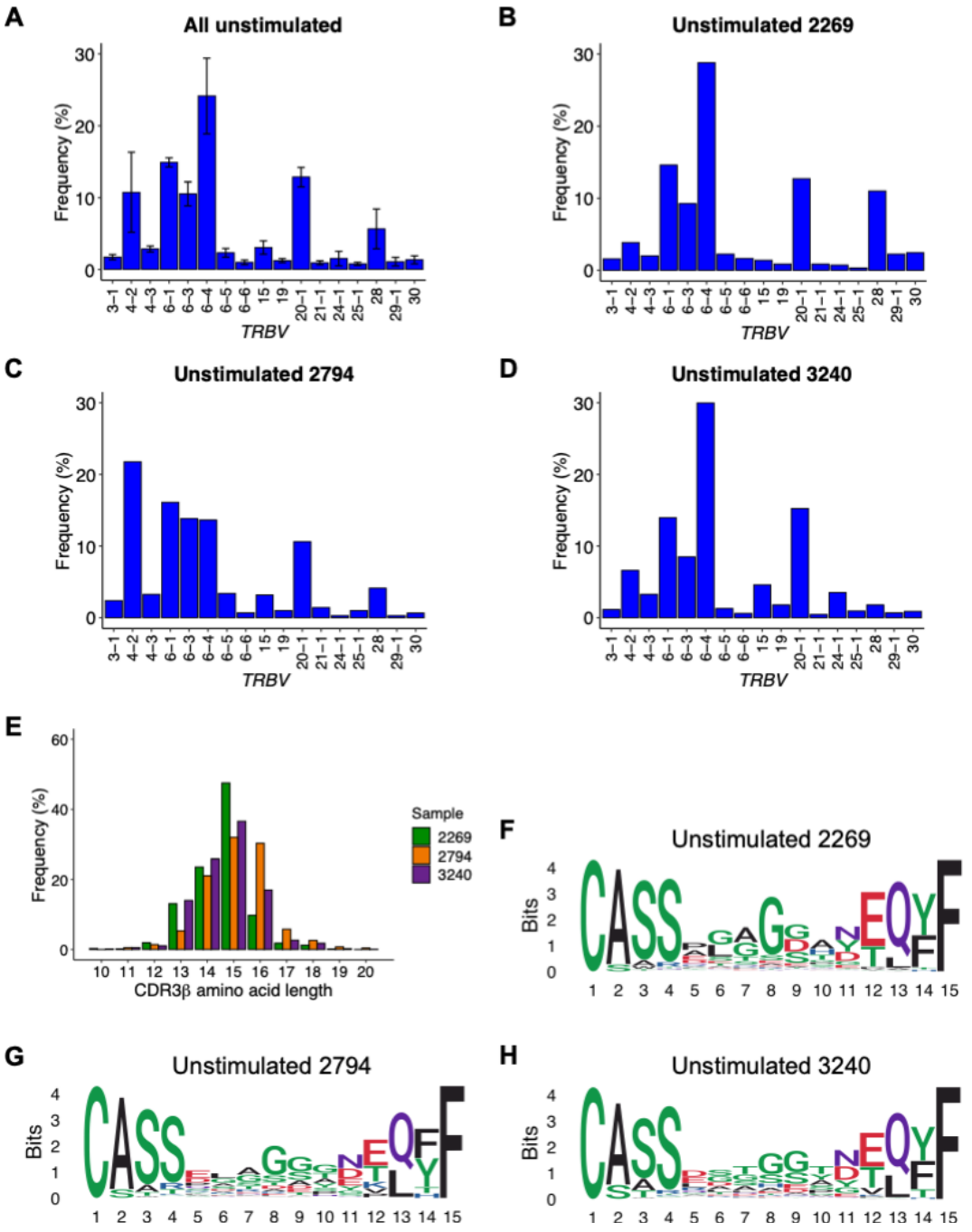


Figure 6.12. Features of the MAIT cell TCRβ chain repertoire. (A) Bar chart showing the mean ± SEM percentage of MAIT cells using the indicated *TRBV* gene segments in the three unstimulated samples. **(B-D)** Bar charts showing the percentage of unstimulated MAIT cells using the indicated *TRBV* gene segments in donors 2269 (B), 2794 (C) and 3240 (D). Bar charts in (A-D) include *TRBV* segments with a frequency > 1% in any one sample. **(E)** Frequency of CDR3β sequences of different lengths in each sample. **(F-H)** Sequence logos generated from all unstimulated MAIT cell *TRAV1-2* CDR3β amino acid sequences of length 15 from donors 2296 (F), 2794 (G) and 3240 (H). The overall height of the stacked letters at each position indicates the sequence conservation, while the relative abundance of each amino acid is indicated by the height of individual letters within the stack. Acidic bases are shown in red, basic residues in blue, hydrophobic in black and polar in green.

In Experiment 1, I identified some bias in the pairing of *TRAJ* and *TRBV* segments (Fig. 5.14). Namely, compared with *TRAJ33* TCRs, *TRAJ12* and *TRAJ20* TCRα chains showed a preference for pairing with *TRBV6-4* TCRβ chains, while *TRAJ33* TCRα chains showed increased pairing with *TRBV20-1* TCRβ chains. Analogous to Experiment 1, *TRBV6-4* pairings were relatively enriched amongst *TRAJ12* and *TRAJ20* TCRs compared with *TRAJ33* TCRs in Experiment 2, although the differences did not reach statistical significance (Fig. 6.13A,B). While *TRBV20-1* showed enriched pairing with *TRAJ33* TCRs compared with *TRAJ12*, there was no significant difference in *TRBV20-1* pairing between *TRAJ33* and *TRAJ20* TCRs in Experiment 2 (Fig. 6.13A,C). Consistent with Experiment 1, use of *TRBV6-1* was comparable across *TRAJ* segments (Fig. 6.13A,D).

Functional TCR clonotypes show little overlap between donors, consistent with Experiment 1

I next examined the overlap of MAIT cell functional TCR clonotypes between donors. As in Experiment 1, there was very little overlap of functional TCRαβ clonotypes between individuals, with ≥ 93% of MAIT cells belonging to a clonotype that was unique to a particular donor (Fig. 6.14A). In contrast, functional TCRα clonotypes were highly shared, with 91% of all MAIT cells belonging to one of 41 clonotypes present in all three donors (Fig. 6.14B). This large amount of sharing was not observed for functional TCRβ clonotypes (Fig. 6.14C), indicating that diversity in the MAIT cell TCR repertoire across individuals is governed by the TCRβ chain, confirming my results from Chapter 5.

In summary, my findings from unstimulated MAIT cells in Experiment 2 and blood MAIT cells in Experiment 1 were largely analogous. Namely, MAIT cells comprised

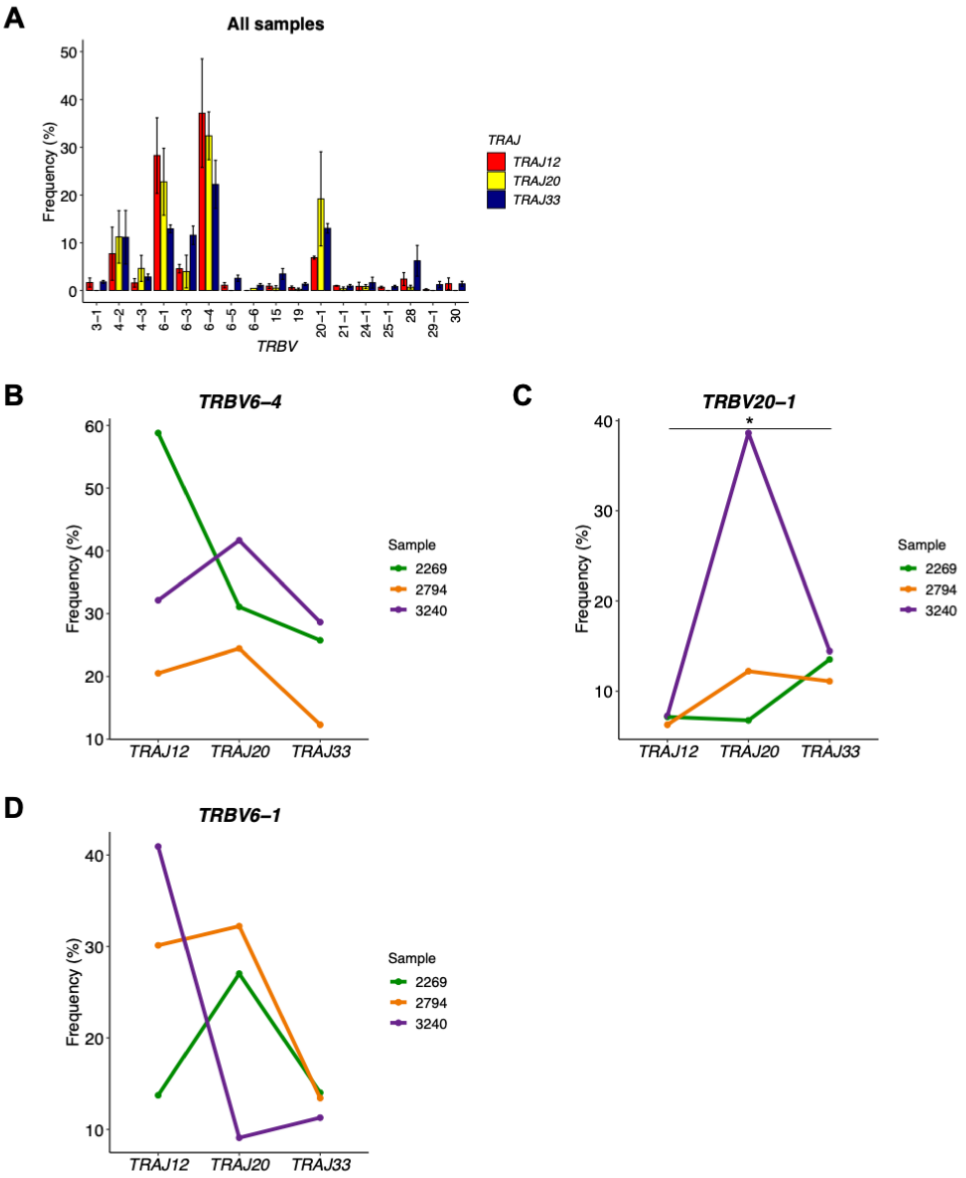


Figure 6.13. Association between *TRAJ* and *TRBV* segment usage in MAIT cell TCRs from the three unstimulated samples. (A) Bar chart showing the mean ± SEM percentage of *TRAJ12* (red), *TRAJ20* (yellow) and *TRAJ33* (blue) *TRAV1-2* MAIT cell TCRs using each *TRBV* gene segment in the three unstimulated samples. Bar chart includes *TRBV* segments with a frequency > 1% in any one sample. **(B-D)** Line plots showing the frequency of *TRAJ12*, *TRAJ20* and *TRAJ33 TRAV1-2* TCRs using *TRBV6-4* (B), *TRBV20-1* (C) and *TRBV6-1* (D) in each sample. Repeated measures one-way ANOVA with the Greenhouse-Geisser correction followed by Tukey's multiple comparisons test with individual variances computed for each comparison. * $p < 0.05$.

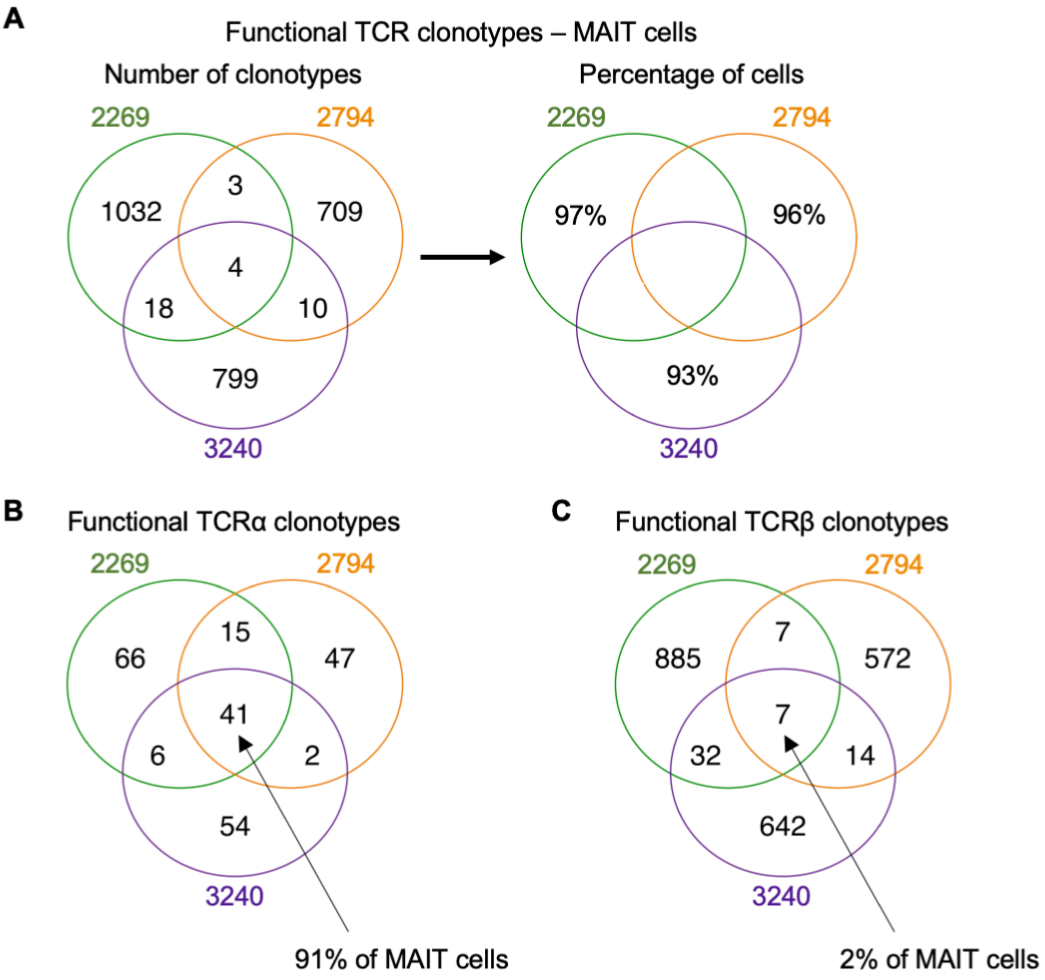


Figure 6.14. Overlap of MAIT cell functional TCR clonotypes between donors.
(A) Venn diagrams illustrating the overlap of MAIT cell functional (amino acid) TCRαβ clonotypes between donors (left) and the corresponding percentage of cells from each donor that have a donor-specific clonotype (right). **(B,C)** Overlap of MAIT cell functional TCRα (B) and TCRβ (C) clonotypes between donors. The numbers in the Venn diagrams indicate the number of clonotypes that are donor-specific or shared between donors.

a similar number of clusters, and cluster markers were enriched for genes differentially expressed following MAIT cell activation. Additionally, *GNLY* expression was essentially limited to a single cluster in each donor. While the localisation of clonotypes across clusters appeared more random in Experiment 2, clonotype markers were still enriched for cell activation pathways and NK cell receptors, suggesting that clonotypes may vary in activation state. The general properties of the TCR repertoire were consistent across the two experiments, including a highly restricted CDR3α length and sequence, a diverse TCRβ chain, and little overlap of TCRαβ clonotypes between donors. However, the degree of MAIT cell oligoclonality in Experiment 2 was significantly decreased compared with Experiment 1.

6.5.3. Pathways of MAIT cell activation following TCR and cytokine stimulation

MAIT cells exhibit divergent transcriptional profiles following TCR or cytokine stimulation

Having established that the unstimulated MAIT cell data from this experiment (Experiment 2) was largely comparable with that of Experiment 1, I moved on to analysis of the stimulation data. For each donor, MAIT cells from the three stimulation conditions were combined for analysis. Clustering at low resolution identified 4-7 clusters per donor (Fig. 6.15A-C), with cells from the different stimulation conditions largely located within distinct clusters (Fig. 6.15A-F). The lack of overlap between unstimulated and stimulated (either TCR or cytokine) MAIT cells indicates that essentially all cells within the TCR and cytokine stimulation conditions had undergone transcriptional changes.

MAIT cells activated via the TCR or with cytokines were present in distinct clusters, suggesting that different stimuli activate divergent transcriptional programmes. This is consistent with published bulk RNA-seq data from TCR- and cytokine-activated MAIT cells (Hinks et al., 2019; Lamichhane et al., 2019b; Leng et al., 2019).

Activation-associated genes show a graded pattern of expression

To gain further insight into the data, I visualised the expression of genes that were upregulated (*IFNG*, *TNF*, *GZMB*, *IL26*) or downregulated (*IL7R*, *KLRB1*, *CD27*, *TCF7*) upon TCR- or cytokine-dependent MAIT cell activation in a published study (Leng et al., 2019) (Fig. 6.16A-C). Both upregulated and downregulated genes were expressed in a gradient, with upregulated genes most highly expressed in the region of the UMAP furthest from unstimulated MAIT cells. Downregulated genes showed the opposite pattern of expression. This suggested that rather than activated MAIT cells comprising subsets with distinct functional capacity, they instead fall along an activation trajectory going from non-activated MAIT cells to the most highly activated MAIT cells. The kinetics of gene expression along the activation trajectory appear specific to the stimulation condition, although many genes were differentially expressed in response to both TCR and cytokine stimulation.

Visualising the expression of these activation-associated genes also provided some insights into the differences in gene expression between TCR- and cytokine-stimulated MAIT cells. Namely, *IFNG* and *IL26* were more highly expressed upon cytokine stimulation, *TNF* showed slightly higher expression upon TCR stimulation, and *GZMB* showed similar expression across the two activation conditions. *IL7R* and *KLRB1* were similarly downregulated by the two stimuli, while *CD27* was more strongly downregulated following cytokine stimulation. These distinctions may

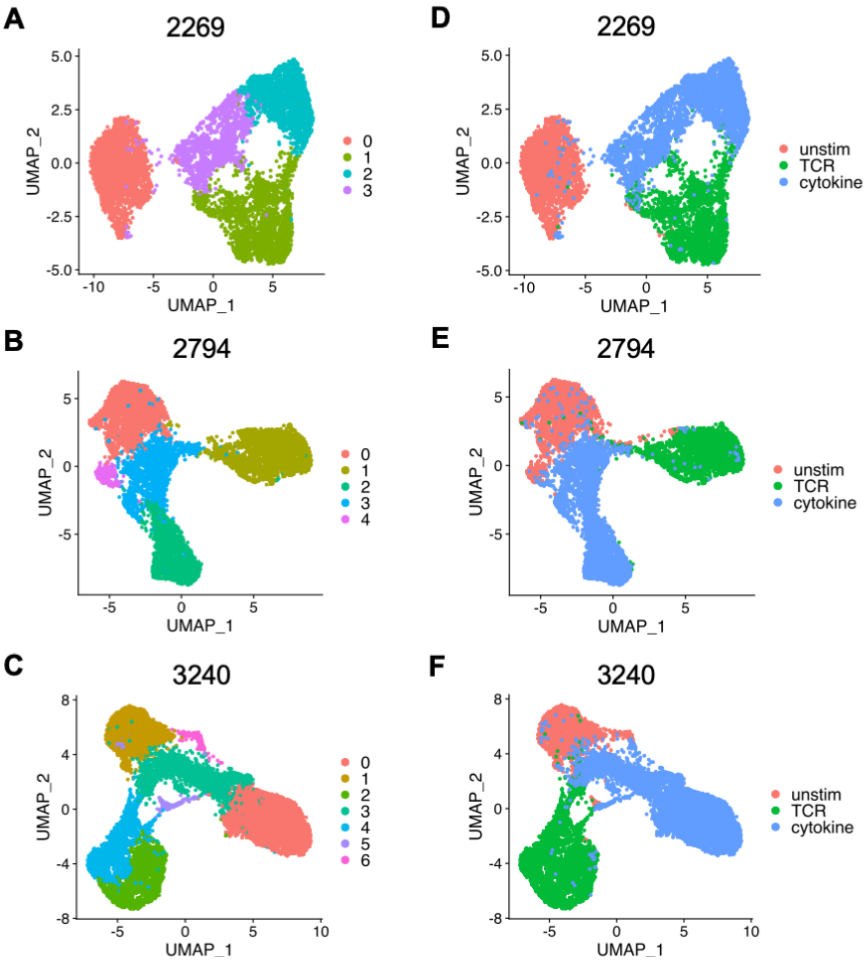


Figure 6.15. Clustering analysis of merged unstimulated, TCR-stimulated and cytokine-stimulated MAIT cells. (A-C) UMAP visualisation of MAIT cell clusters from the combined analysis of unstimulated, TCR-stimulated and cytokine-stimulated samples from donors 2269 (A), 2794 (B) and 3240 (C). **(D-F)** Stimulation status overlaid onto UMAP plots in (A-C).

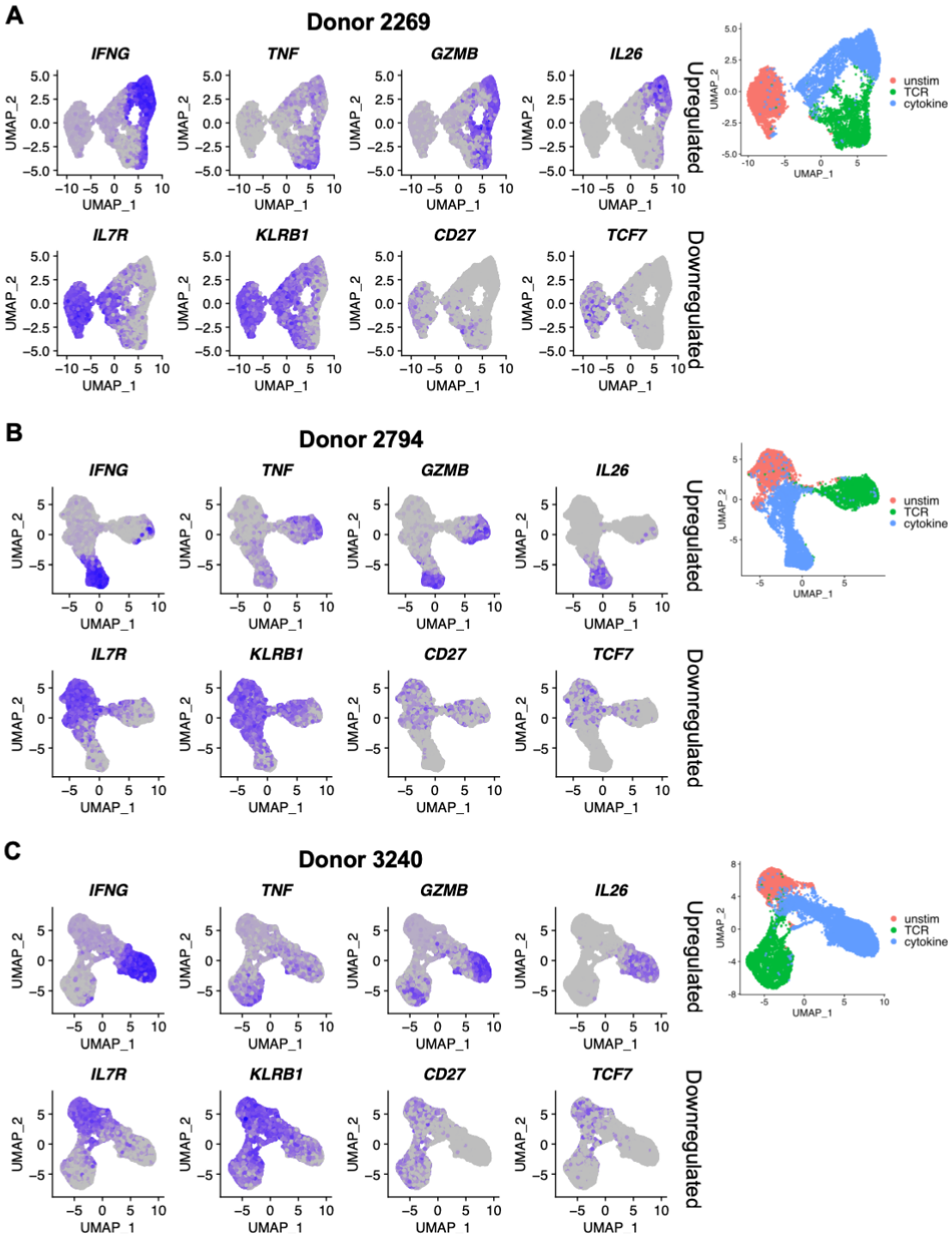


Figure 6.16. Activation-associated genes exhibit a graded expression pattern.
(A-C) Expression of example genes that were upregulated (top) or downregulated (bottom) upon MAIT cell activation in a published study (Leng et al., 2019), as visualised by UMAP. Grey indicates low expression level and dark blue indicates high expression level.

reflect differences in the transcriptional changes induced by the two stimuli or in the kinetics of gene expression following TCR- and cytokine-dependent activation. Prior reports suggest that MAIT cells show a peak activation level 4-6 hours after TCR stimulation, compared with 20-24 hours for cytokine stimulation (Lamichhane et al., 2019b; Slichter et al., 2016). However, this conclusion is based on the expression kinetics of a limited number of genes and proteins, and therefore these timepoints may not reflect the overall peak of TCR- and cytokine-dependent activation.

Aggregated analysis of all samples

As the UMAP plots and gene expression distributions looked similar for the three donors, I decided to perform an aggregated analysis of all samples. Low resolution clustering identified nine clusters (Fig. 6.17A). As with the individual donor analyses, these clusters largely comprised cells from a single stimulation condition (Fig. 6.17B,C), with the exception of cluster 6, which contained ~43% unstimulated, ~18% TCR-stimulated and ~40% cytokine-stimulated MAIT cells. Cluster 6 was highly enriched for the expression of *GNLY* (Fig. 6.17D). This suggests that the *GNLY*$^{+}$ MAIT cells found in unstimulated samples (Fig. 6.5) do not undergo major global transcriptional changes following activation. *GNLY* expression was upregulated following TCR- and cytokine-dependent activation (TCR: FC 1.83, $p = 2.24 \times 10^{-57}$; cytokine: FC 1.96, $p = 6.56 \times 10^{-89}$), but the fraction of cells expressing *GNLY* was relatively similar in all conditions (26.6% in unstimulated, 28.6% in TCR-stimulated, 30.8% in cytokine-stimulated).

The majority of clusters contained a mixture of cells from the three donors (Fig. 6.17E), indicating that changes in gene expression following TCR and cytokine stimulation were conserved across donors. However, cells from donor 2269 were almost absent from cluster 8, which contained mostly unstimulated cells (~75%; Fig.

6.17A-C,E). Compared with other unstimulated MAIT cells (cluster 0), cells from cluster 8 showed enriched expression of metallothioneins, such as *MT1X* and *MT2A*, which were also upregulated in cytokine-stimulated, but not TCR-stimulated, MAIT cells (Fig. 6.17F). In the individual analyses of unstimulated MAIT cells from donors 2794 and 3240, metallothionein-expressing cells also comprised a distinct cluster (Fig. 6.17G and data not shown). However, in the blood samples from Experiment 1, metallothionein expression was distributed throughout MAIT cell clusters (Fig. 6.17H and data not shown). Metallothioneins function to regulate the metabolism of essential metals (Juárez-Rebollar et al., 2017), suggesting metabolic differences between the samples in Experiment 1 and 2. However, the specific function of this metallothionein-expressing cluster is unknown.

I next visualised the expression of the activation-associated genes previously examined in the individual donors. Similar patterns of gene expression were identified (Fig. 6.17I), with genes showing a gradient of expression moving from unstimulated cells on the right of the UMAP to the most highly activated cells on the left of the UMAP. Expression of *IFNG* and *IL26* was strongly upregulated following cytokine but not TCR stimulation, while *TNF* showed greater upregulation upon TCR stimulation. *GZMB* was similarly upregulated, and *IL7R* and *KLRB1* were similarly downregulated, in response to the two stimuli. *CD27* showed greater downregulation upon cytokine stimulation compared with TCR stimulation.

Pseudotime analysis confirms a branching trajectory of activation, with distinct branches for TCR- and cytokine-activated MAIT cells

Given that activation-associated genes exhibited a graded expression pattern, I decided to perform pseudotime analysis to characterise the gene expression

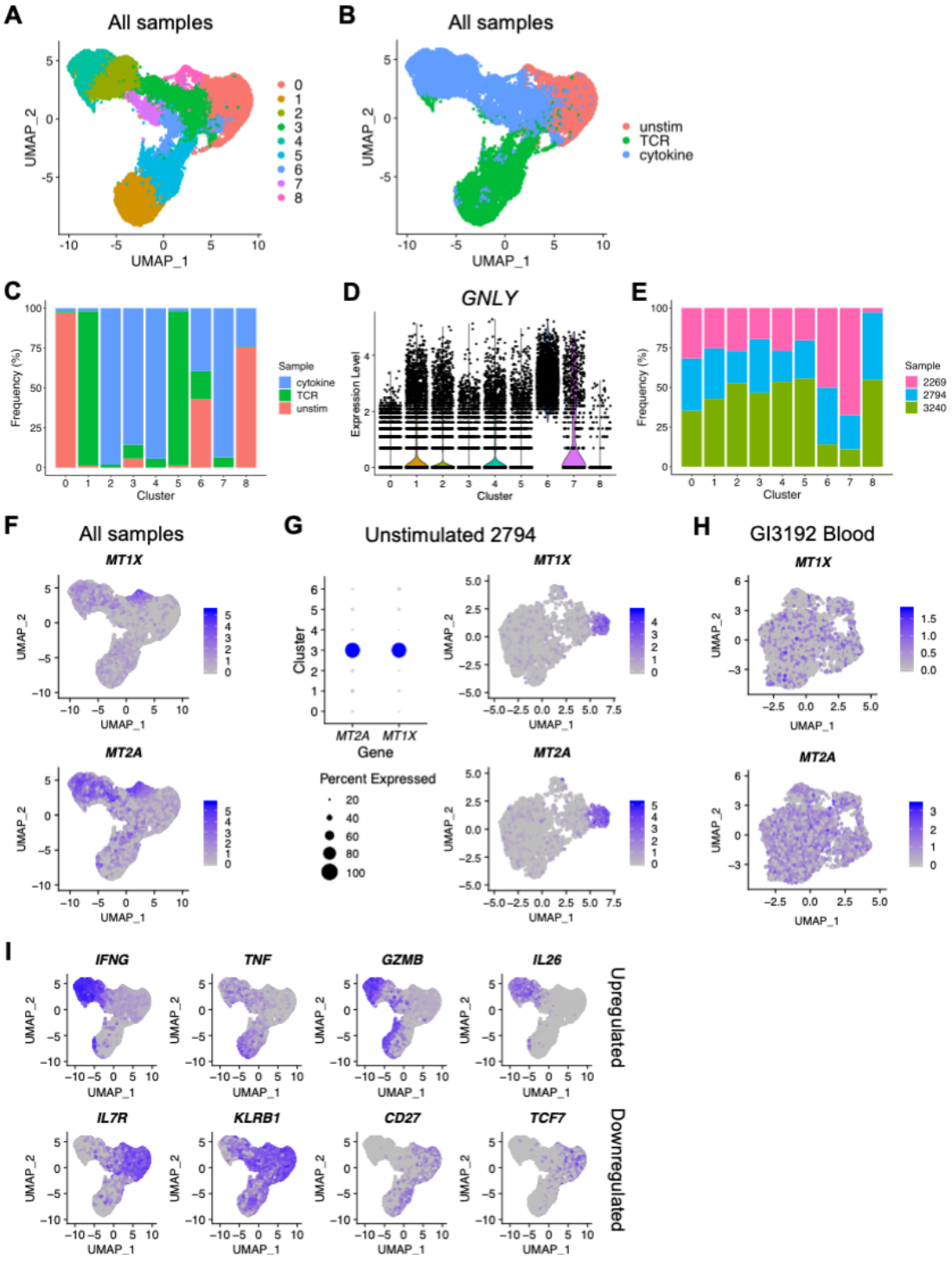


Figure 6.17. Transcriptional changes upon TCR and cytokine activation are conserved across donors. (A) Clusters of MAIT cells from the combined analysis of all samples, as visualised by UMAP. **(B)** Location of unstimulated, TCR-stimulated and cytokine-stimulated MAIT cells on the UMAP shown in (A). **(C)** Bar graph showing the proportion of MAIT cells from the three stimulation conditions found in each cluster. **(D)** Violin plot showing the expression of *GNLY* within each cluster. **(E)** Bar graph showing the proportion of MAIT cells from the three donors found in each cluster. **(F-H)** Expression of the metallothioneins *MT1X* and *MT2A* by MAIT cells from all samples (F), unstimulated cells from donor 2794 (G) or blood MAIT cells from donor GI3192 in Experiment 1 (H). **(I)** Expression of activation-associated genes by MAIT cells from all samples.

trajectories going from a non-activated to a TCR- or cytokine-activated MAIT cell. Cluster 0, which contains unstimulated MAIT cells, was specified as the starting cluster for Slingshot trajectory analysis (Street et al., 2018).

Slingshot revealed a branching trajectory, with one branch containing TCR-stimulated MAIT cells and a second branch containing cytokine-stimulated MAIT cells (Fig. 6.18A). Expression of activation-associated genes, for example *IFNG* (upregulated), *GZMB* (upregulated), *KLRB1* (downregulated) and *IL7R* (downregulated), showed the expected pattern along the TCR and cytokine activation trajectories (Fig. 6.18B-E).

SCORPIUS trajectories for TCR- and cytokine-activated MAIT cells look similar to the trajectories identified using Slingshot

Having identified that the TCR and cytokine activation pathways of MAIT cells are distinct linear pathways, I wanted to identify the key genes that predict progression along the trajectories and to characterise modules of genes that show similar patterns of expression. These analyses are easy to perform using the linear pseudotime algorithm SCORPIUS (Cannoodt et al., 2016).

First, I checked that the trajectories identified by SCORPIUS were similar to those identified by Slingshot. Analysis of unstimulated MAIT cells together with TCR-stimulated MAIT cells identified a linear trajectory (Fig. 6.19A), as did combined analysis of unstimulated and cytokine-stimulated MAIT cells (Fig. 6.19B). The expression patterns of activation-associated genes along the two pseudotime trajectories were similar to those observed with Slingshot (Fig. 6.19C-D). For example, *GZMB* was upregulated relatively early in pseudotime and was expressed

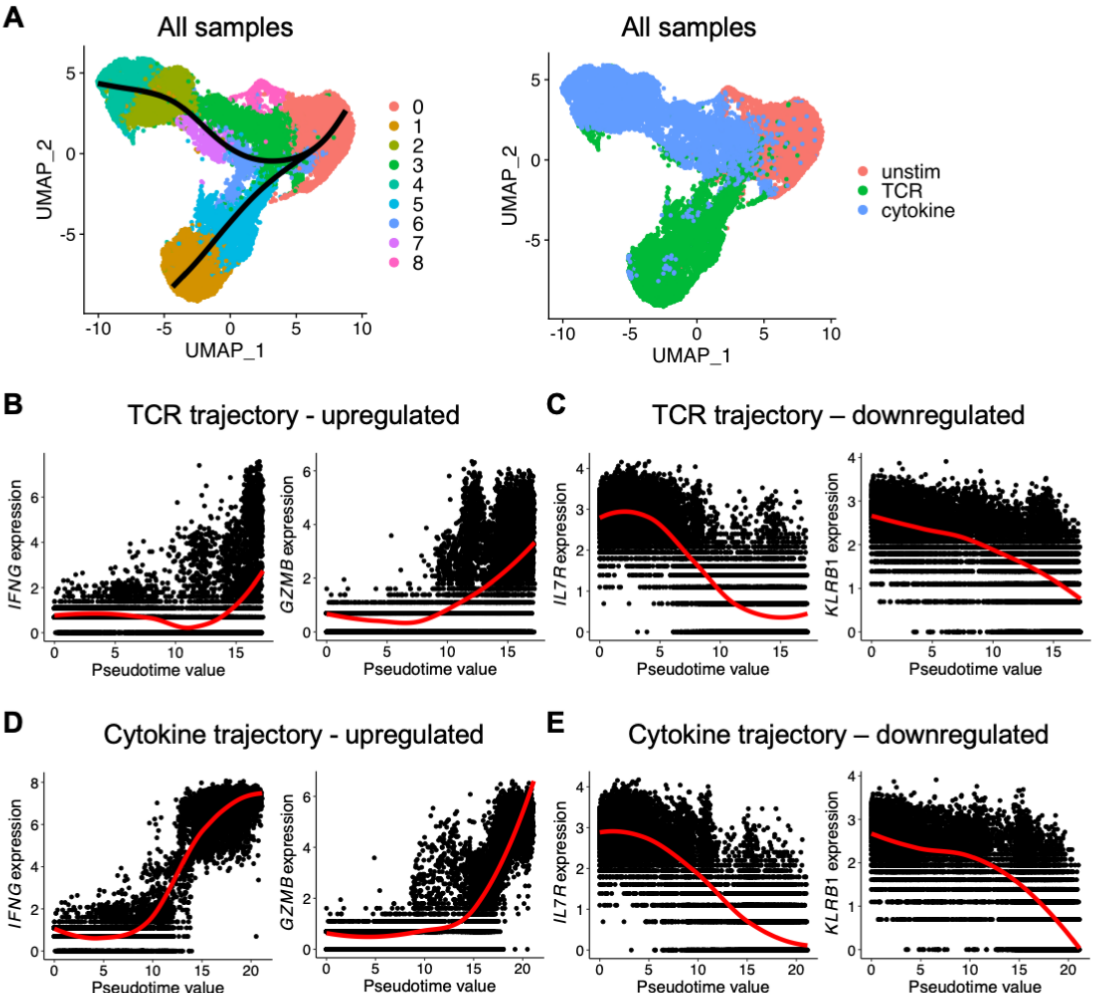


Figure 6.18. Pseudotime analysis reveals distinct trajectories for cytokine- and TCR-activated MAIT cells. (A) Left: UMAP of all samples showing the branching trajectory (indicated by the black line) identified with Slingshot. Right: location of cells from the different stimulation conditions on the UMAP. **(B-E)** Scatter plots showing the expression of genes that are upregulated (B,D) or downregulated (C,E) over pseudotime on the TCR (B,C) or cytokine (D,E) activation trajectory. Loess curves are shown in red.

by a large fraction of TCR- and cytokine-activated MAIT cells. *IFNG* was highly expressed by MAIT cells following cytokine stimulation, but only expressed by the most highly activated TCR-stimulated MAIT cells. Lastly, TCR and cytokine stimulation induced comparable downregulation of *IL7R* and *KLRB1*, with *IL7R* downregulated prior to *KLRB1* on both trajectories. *CD69* showed an unexpected but analogous expression pattern following TCR and cytokine stimulation (Fig. 6.19E). Its expression increased in early activated MAIT cells, reached a peak around halfway through the activation trajectory, and then decreased gradually to baseline levels. This is contrasting with what is observed at the protein level, where expression of CD69 is highest on the most activated, IFNγ-expressing MAIT cells (Emgård et al., 2019).

A mix of shared and pathway-specific genes predict progression along TCR and cytokine activation trajectories

SCORPIUS uses the Random Forest (Breiman, 2001) algorithm to identify the genes whose expression level is best able to predict progression along a defined pseudotime trajectory. The top 15 most predictive genes for the TCR and cytokine activation trajectories included *HSP90AB1*, *PKM*, *TUBA1B* and *RPS27* (Fig. 6.20A). *HSP90AB1* encodes a member of the heat shock protein 90 family that regulates a variety of cellular processes (including transcription and the cell cycle); *PKM* encodes the glycolytic enzyme pyruvate kinase; *TUBA1B* encodes tubulin, the major component of microtubules; and *RPS27* encodes a component of the 40S ribosomal subunit. The nature of these genes suggests changes in a number of key cellular processes following MAIT cell activation regardless of the stimulus used. *IL7R*, which is downregulated following MAIT cell activation (Fig. 6.17I, 6.18C,E and

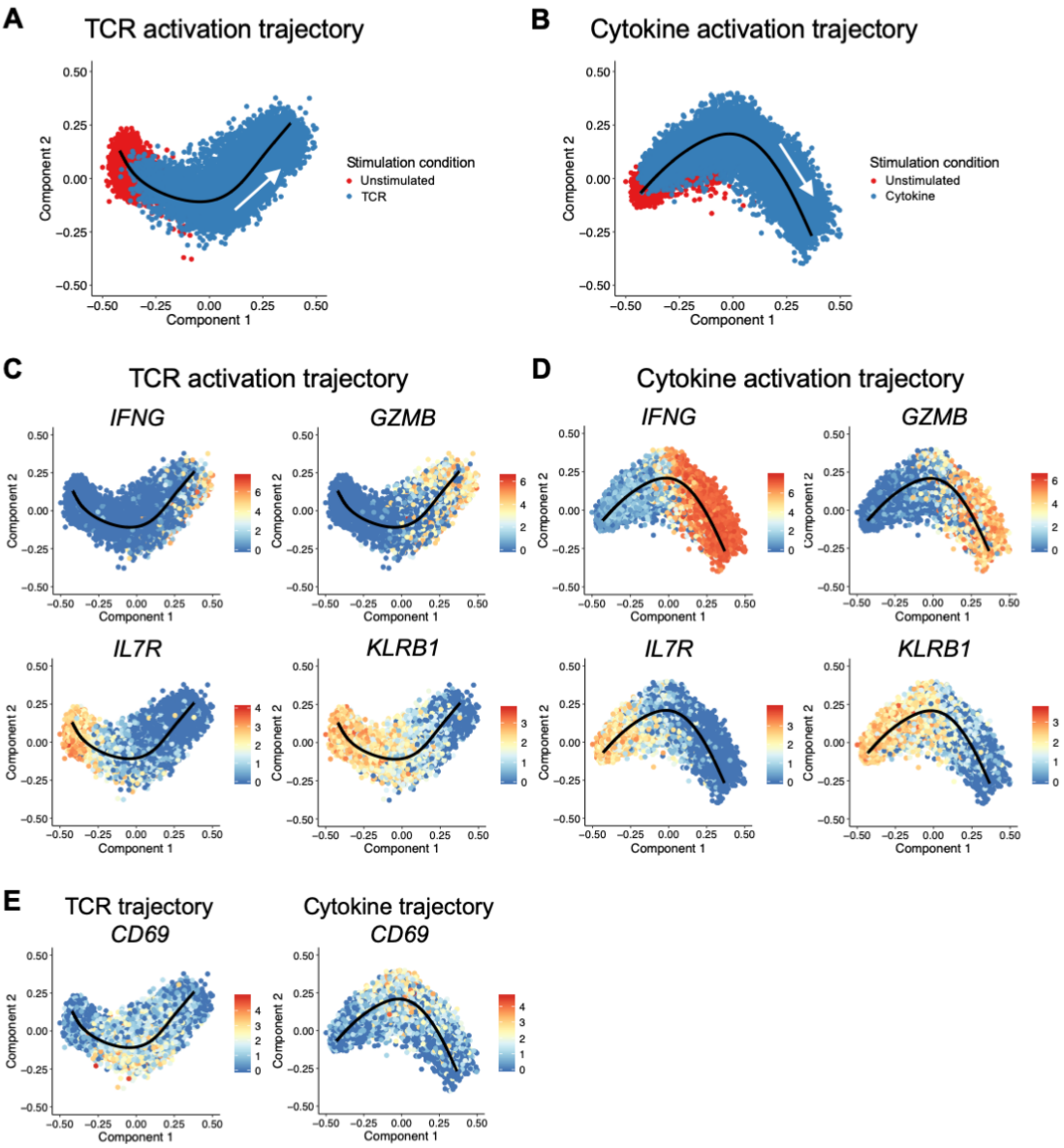


Figure 6.19. The linear TCR and cytokine activation trajectories identified using SCORPIUS are comparable with the Slingshot trajectories. (A,B) SCORPIUS TCR (A) or cytokine (B) activation trajectory (indicated by the black line) generated from combined unstimulated and TCR-stimulated MAIT cells, or unstimulated and cytokine-stimulated MAIT cells, respectively. The white arrow indicates the direction going from the least activated to the most activated MAIT cells. **(C,D)** Expression of activation-associated genes along the TCR (C) and cytokine (D) activation trajectories. **(E)** *CD69* expression on the TCR (left) and cytokine (right) activation trajectories.

A **SCORPIUS**

TCR trajectory	Cytokine trajectory
1. *HSP90AB1*	1. *IFNG*
2. *RAN*	2. *HSP90AB1*
3. *IL7R*	3. *IL7R*
4. *TUBA1B*	4. *IL26*
5. *IL2RA*	5. *LAPTM5*
6. *FABP5*	6. *TUBA1B*
7. *NPM1*	7. *PKM*
8. *PKM*	8. *CD52*
9. *PRDX1*	9. *CYTIP*
10. *EIF5A*	10. *SEC61B*
11. *HSPE1*	11. *ITIMB*
12. *MIR155HG*	12. *EIF1*
13. *HSPD1*	13. *GZMB*
14. *HSPA8*	14. *CHCHD2*
15. *RPS27*	15. *RPS27*

B **Slingshot**

TCR trajectory	Cytokine trajectory
1. *HSP90AB1*	1. *IFNG*
2. *IL7R*	2. *PKM*
3. *CCL4*	3. *IL7R*
4. *CCL3*	4. *GNLY*
5. *GNLY*	5. *GZMB*
6. *FTH1*	6. *IFITM1*
7. *CCL4L2*	7. *CD52*
8. *PKM*	8. *IL26*
9. *RPS27*	9. *MX1*
10. *IL4I1*	10. *FTH1*
11. *MTRNR2L12*	11. *RPS27*
12. *TNFRSF9*	12. *NPPC*
13. *TUBA1B*	13. *NKKBIA*
14. *VIM*	14. *ENO1*
15. *CD7*	15. *CYTIP*

Figure 6.20. Identification of key genes that predict progression along MAIT cell TCR and cytokine activation trajectories. (A,B) Table showing the top 15 genes whose expression is most predictive of progression along the SCORPIUS (A) or Slingshot (B) TCR and cytokine activation trajectories. Genes in orange and green are within the top 15 most predictive genes for the TCR and cytokine activation pathways, respectively, identified with SCORPIUS and Slingshot.

6.19C,D), was the only immune gene amongst the top 15 most predictive genes for both pathways.

To confirm that the genes identified were not specific to the SCORPIUS trajectories, I also used the Random Forest (Breiman, 2001) algorithm to identify the genes best able to predict progression along the Slingshot TCR and cytokine trajectories (Fig. 6.20B). Of the top 15 most predictive genes for the TCR and cytokine activation trajectories, five and eight, respectively, were identified by both SCORPIUS and Slingshot (Fig. 6.20A,B). This provides validation of the trajectories and genes identified using SCORPIUS, and provides a list of key genes whose expression can predict MAIT cell activation state following TCR- or cytokine-dependent activation.

Having validated the SCORPIUS trajectories, I further characterised the similarities and differences between the predictor genes for the TCR and cytokine activation pathways (Fig. 6.21A). In total, 60 of the top 100 predictor genes were shared across the two pathways (Table 6.2 and 6.3). In addition to genes associated with key cellular processes (e.g. *HSP90AB1*; Fig. 6.21), as discussed above, these encompassed a variety of immune-related genes, including chemokines (e.g. *CCL5*, *CCL20*), cytokine and chemokine receptors (e.g. *IL7R*, *IL2RA*, *CXCR4*), cytotoxic molecules (e.g. *GZMK*, *GZMB*), costimulatory molecules (e.g. *TNFRSF18* [encodes GITR]), and *KLRB1* (encodes CD161), a key marker of MAIT cells. Nevertheless, differences could be identified between the predictor genes for the TCR and cytokine trajectories. *IFNG* was the best predictor of progression along the cytokine activation trajectory (Fig. 6.20), whilst it was much less important for the TCR trajectory (220th most important gene). This is consistent with the observed upregulation of *IFNG* in a large proportion of cytokine-stimulated MAIT cells, but only a small fraction of TCR-stimulated MAIT cells (Fig. 6.17I, 6.18B,D and

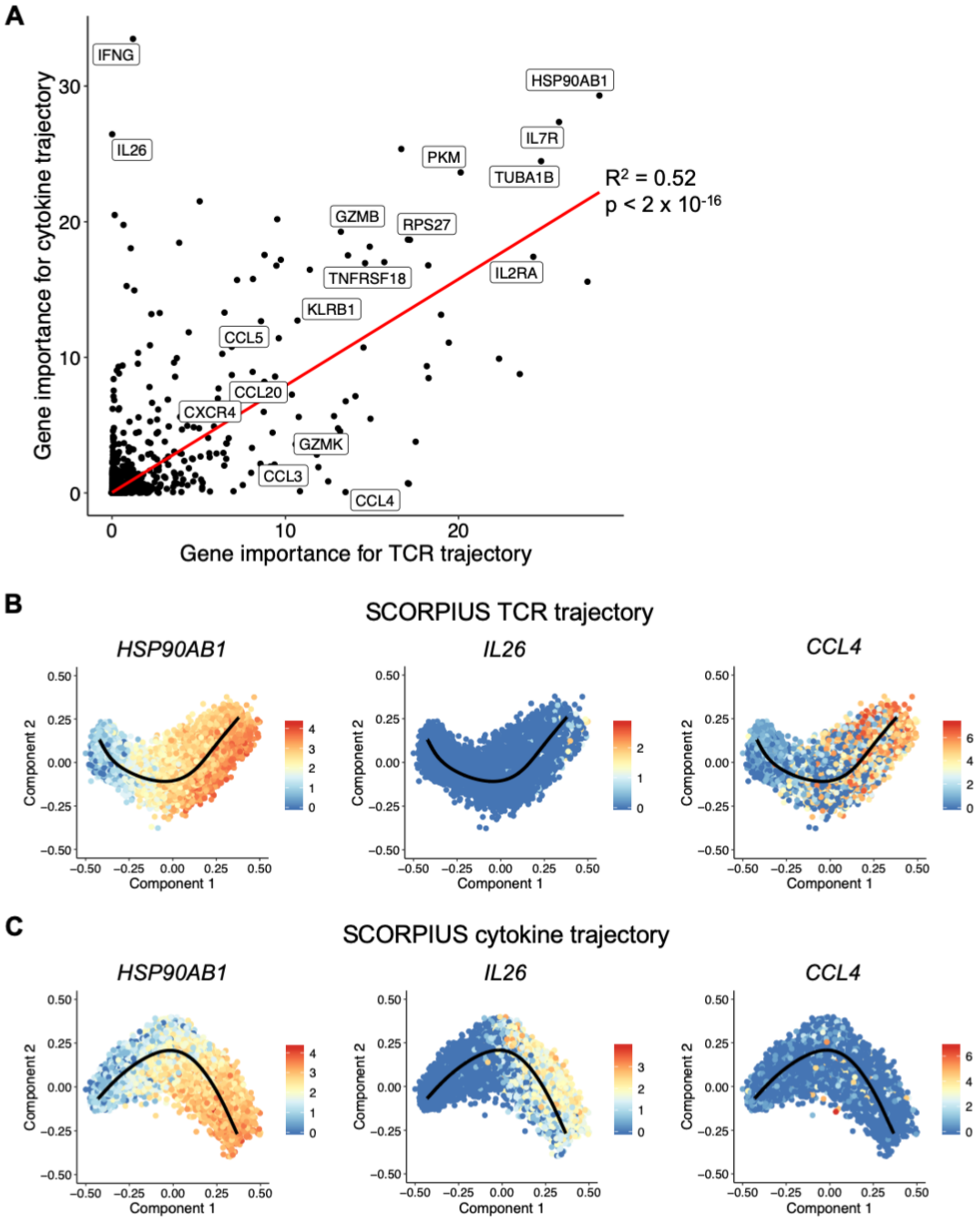


Figure 6.21. Similarities and differences in the importance of specific genes for predicting progression along the TCR and cytokine activation trajectories identified using SCORPIUS. (A) Scatter plot of all genes showing their importance for predicting progression along the TCR and cytokine activation trajectories identified using SCORPIUS. The linear regression line is shown in red, and the associated R^2 and p value are indicated. **(B,C)** Expression of example genes along the SCORPIUS TCR (B) or cytokine (C) activation trajectory. *HSP90AB1* is highly predictive of progression along both pathways, while *IL26* is more predictive for the cytokine trajectory and *CCL4* for the TCR trajectory.

6.19C,D). Additionally, *IL26* was a key predictor of progression along the cytokine but not the TCR activation trajectory (Fig. 6.20 and 6.21). In contrast, *CCL3* and *CCL4* were better able to predict progression along the TCR activation trajectory compared with the cytokine trajectory (Fig. 6.21; TCR trajectory: 39th and 30th best predictors, respectively; cytokine pathway: 671st and 933rd best predictors, respectively).

Identification of four gene expression modules with distinct patterns of expression following TCR and cytokine stimulation

The top 100 genes from the SCORPIUS TCR and cytokine activation trajectories were clustered into modules of genes with similar patterns of expression over pseudotime. My analysis identified four distinct modules of genes for both the TCR and cytokine activation pathways (Fig. 6.22), two modules containing genes that were upregulated upon activation and two modules containing genes that were downregulated upon activation. These were further divided into modules of early (module A) and late (module B) upregulated genes, and early (module C) and late (module D) downregulated genes.

Rapidly upregulated genes (module A) included *PKM* (glycolytic enzyme), *HSP90AA1* and *HSP90AB1* (heat shock proteins), *EIF5A* (translation initiation factor), *PRDX1* (antioxidant enzyme) and *SEC61B* (component of the endoplasmic reticulum protein translocation apparatus), among others. In general, rapidly upregulated genes were associated with basic cellular functionality, although *IFNG* was an early upregulated gene within the cytokine activation trajectory (Fig. 6.22B). Late upregulated genes (module B) comprised *GZMB*, *IL2RA* and *CCL20* in both pathways. Late upregulated genes specific to the TCR activation trajectory included

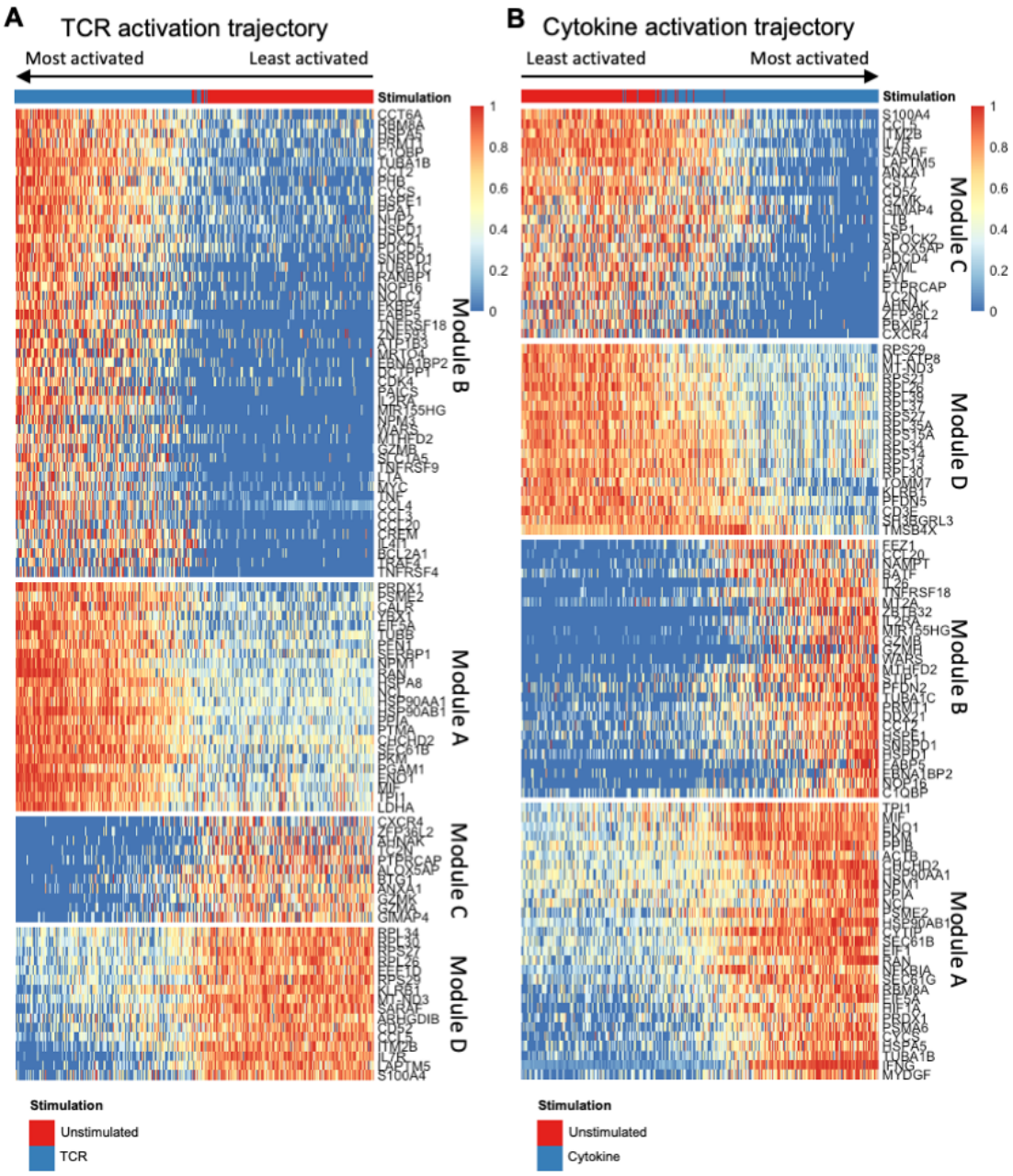


Figure 6.22. Modules of genes showing similar expression patterns along the MAIT cell TCR and cytokine activation pathways. (A,B) Heatmaps showing the top 100 genes most predictive of progression along the SCORPIUS TCR (A) and cytokine (B) activation pathways. The black arrow indicates the direction going from the least activated to the most activated MAIT cells. Modules of genes with shared expression patterns are indicated. Module A = early upregulated genes, module B = late upregulated genes, module C = early downregulated genes, module D = late downregulated genes.

CCL3, *CCL4* and *LTA*, while late upregulated genes specific to the cytokine activation trajectory included *BATF* (TF with diverse immune functions e.g. regulates Th17 cell differentiation (Schraml et al., 2009)) and *IL26*. Therefore, upregulated genes with immune functionality were predominantly found within module B, indicating relatively late upregulation following activation.

GZMK and *CXCR4* showed rapid downregulation (module C) following TCR and cytokine stimulation, while more slowly downregulated genes (module D) included *KLRB1* and a number of genes encoding ribosomal proteins. *IL7R* was found within module C for the cytokine trajectory, but module D for the TCR trajectory.

Clonotypes show some variation in their relative position along the TCR and cytokine activation trajectories

Finally, I addressed the question of whether MAIT cell clonotypes differ in their capacity for activation in response to either TCR or cytokine stimuli. To achieve this, I visualised the location of the top 10 largest clonotypes from each donor along the TCR (Fig. 6.23) and cytokine (Fig. 6.24) activation trajectories identified using SCORPIUS. By eye, there did not appear to be any major differences between clonotypes in their localisation along the TCR pseudotime axis (Fig. 6.23). The majority of clonotypes contained a mixture of cells at different stages of activation i.e. at different positions along the pseudotime trajectory. In general, this was also the case for MAIT cells activated with cytokines (Fig. 6.24).

To more systematically examine the relationship between MAIT cell clonal origin and localisation along the TCR and cytokine pseudotime axes, I generated box plots showing the localisation of the top 10 largest clonotypes from each donor along the

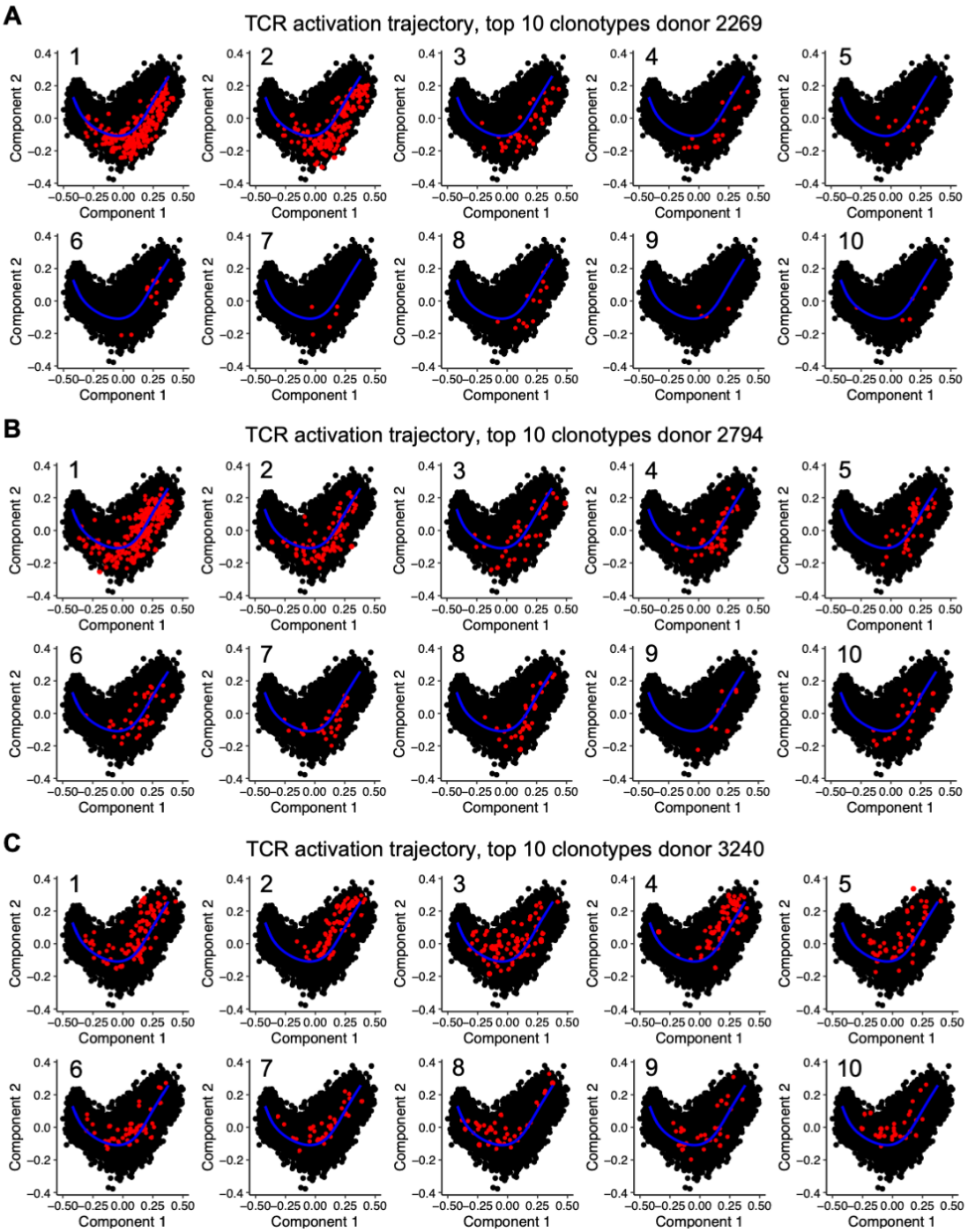


Figure 6.23. Localisation of the top 10 largest clonotypes from each donor along the SCORPIUS TCR activation trajectory. (A-C) Location of the top 10 largest clonotypes from donors 2269 (A), 2794 (B) and 3240 (C) on the SCORPIUS TCR activation trajectory. Cells not within the clonotype are shown in black, cells within the clonotype are shown in red, and the TCR trajectory is indicated by the blue line.

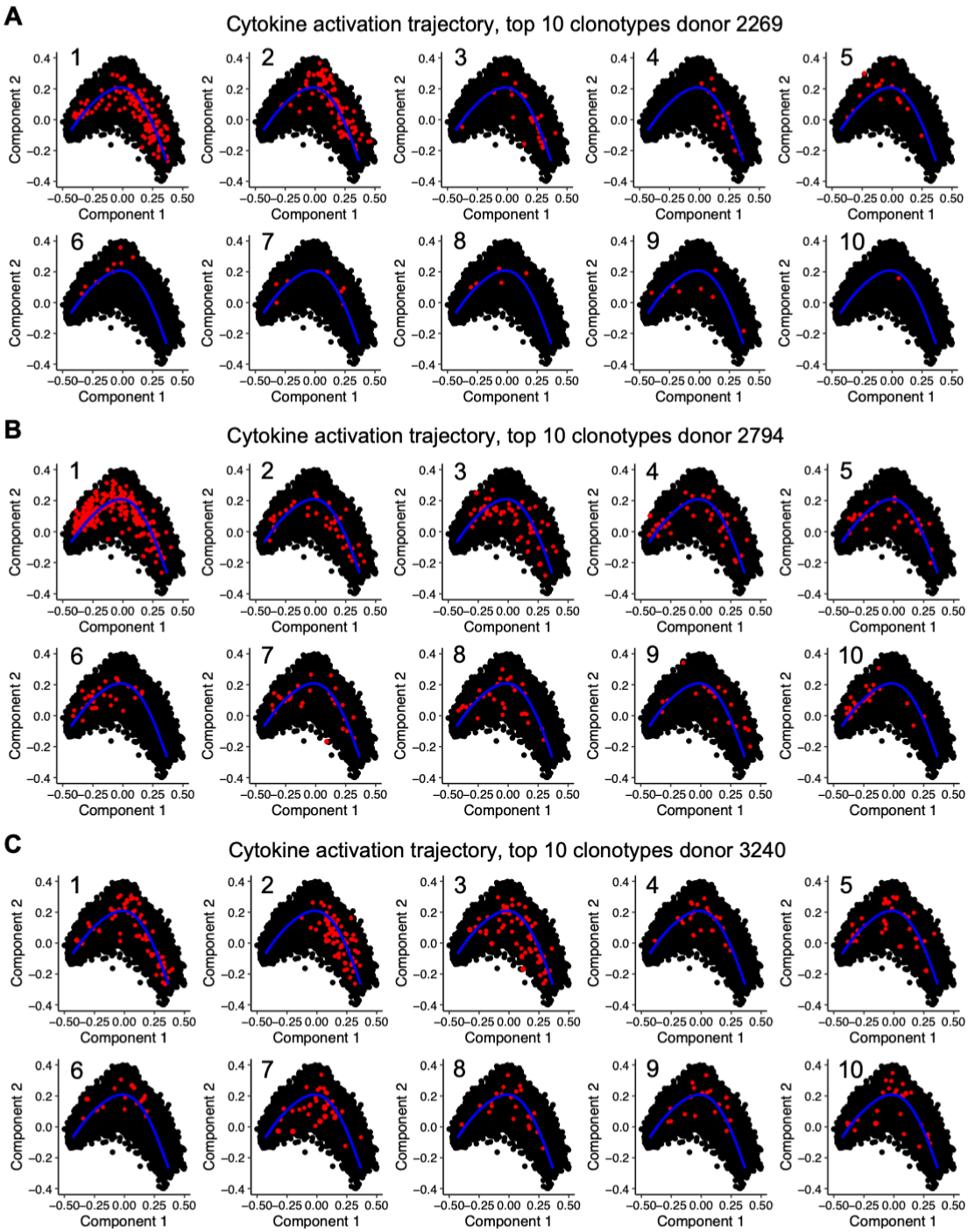


Figure 6.24. Localisation of the top 10 largest clonotypes from each donor along the SCORPIUS cytokine activation trajectory. (A-C) Location of the top 10 largest clonotypes from donors 2269 (A), 2794 (B) and 3240 (C) on the SCORPIUS cytokine activation trajectory. Cells not within the clonotype are shown in black, cells within the clonotype are shown in red, and the cytokine trajectory is indicated by the blue line.

SCORPIUS TCR and cytokine trajectories (Fig. 6.25A-F). I then performed a one-way ANOVA for each donor-trajectory combination (e.g. donor 2269, TCR trajectory) to test the null hypothesis that there was no significant difference in the mean pseudotime value between clonotypes. The ANOVA result was significant (i.e. $p < 0.05$) in each case except for clonotypes from donor 2794 on the TCR trajectory ($p = 0.08$; Fig. 6.25C). This suggests that there are some differences between clonotypes in their capacity for activation in response to TCR or cytokine stimulation, although these differences were small in the setting examined.

6.6. Discussion

In summary, I have demonstrated that unstimulated MAIT cells comprise multiple clusters that largely reflect different states of activation rather than distinct differentiation states. While the majority of TCR clonotypes were randomly distributed across transcriptional clusters of unstimulated MAIT cells, some clonotypes were localised to specific clusters. This indicates that the transcriptome of individual resting MAIT cells can be related to their clonal origin. MAIT cells were shown to follow a linear activation pathway that differed depending on the stimulus, with activation-associated genes displaying a graded expression pattern along the trajectories. Thus, MAIT cells did not comprise functionally distinct subsets, but instead their functionality at any given time reflected their position on an activation trajectory. While clonal origin had a minor effect on activation capacity, the potential of an individual MAIT cell to be activated by TCR and cytokine stimuli was largely independent of clonotype.

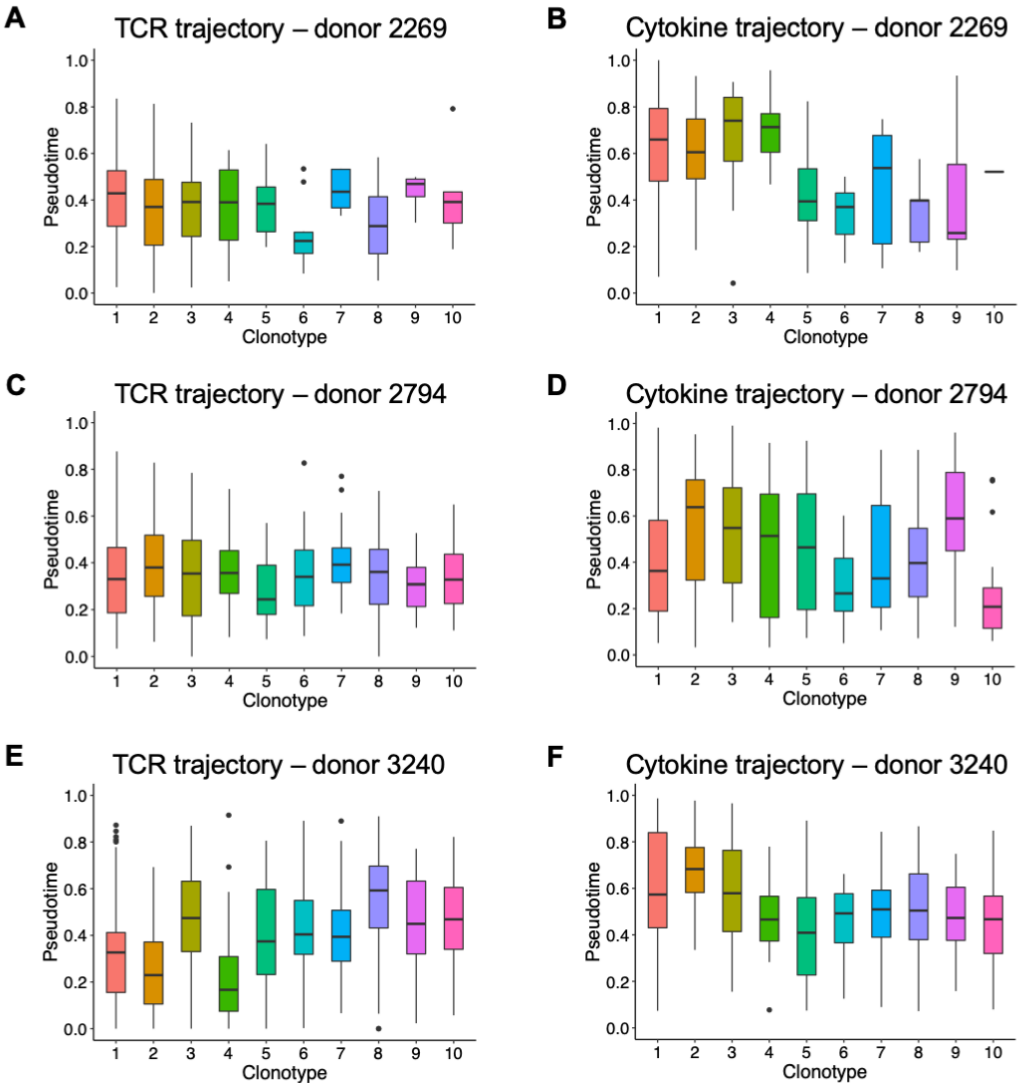


Figure 6.25. TCR clonotypes show minor variation in their position along MAIT cell TCR and cytokine activation pathways. (A-F) Box plots showing the localisation of MAIT cells from the top 10 largest clonotypes from each donor along the TCR (A,C,E) and cytokine (B,D,F) activation trajectories identified using SCORPIUS. The position of TCR- and cytokine-stimulated MAIT cells was examined on the TCR and cytokine activation trajectories, respectively. Plots show cells from donors 2269 (A,B), 2794 (C,D) and 3240 (E,F).

Cluster markers from unstimulated MAIT cells in Experiment 2 indicated variation in MAIT cell activation (Fig. 6.3E and 6.4), as in Experiment 1 (Fig. 5.7B,C, 5.8A and 5.9). However, the number of unique cluster markers identified in Experiment 2 was over threefold more than in Experiment 1 (Fig. 6.3D). This was despite the same number of donors being analysed and unstimulated blood MAIT cells comprising a similar number of clusters per donor. The reasons for this are unclear, however there are a number of plausible explanations. Firstly, the age and disease status of the donors in the two experiments was considerably different. The donors in Experiment 1 were aged 50-66 and were undergoing a liver resection to remove a tumour (benign or malignant), whereas the donors in Experiment 2 were younger (aged 26-31) and disease-free. Secondly, in Experiment 1, blood MAIT cells were analysed directly *ex vivo*, whereas the unstimulated MAIT cells in Experiment 2 were cultured *in vitro* for 20 hours prior to analysis. The cells will have undergone transcriptional changes during incubation that could have included, for example, alterations in the expression of activation-associated genes or changes in metabolism.

A major goal of this work was to determine if MAIT cells form transcriptionally distinct subsets. However, across both single-cell experiments, the data from resting MAIT cells failed to support such a model. The possible exception was the identification of a cluster of MAIT cells within each donor that showed upregulated expression of *GNLY* and encompassed the majority of *GNLY*-expressing cells (Fig. 6.5 and 6.6). This subset has not previously been described, but a small population of granulysin-expressing MAIT cells can be detected by flow cytometry (Leeansyah et al., 2015). Granulysin is a cytotoxic pore-forming protein that mediates the killing of intracellular bacteria, is chemotactic for a variety of cell types, and can induce monocytes to produce proinflammatory cytokines such as RANTES and MIP-1α

(CCL3) (Krensky and Clayberger, 2009; Walch et al., 2014). In every sample, *GNLY* was either the sole marker or one of a few markers for the cluster. None of the other cluster markers were consistent across all donors and most were donor-specific. Thus it was not possible to define a gene signature for these *GNLY*$^{+}$ MAIT cells, which may have helped with predicting their specific function. Upon TCR or cytokine stimulation, *GNLY* showed only a small change in expression, consistent with previous studies that showed no change in granulysin protein expression following MAIT cell activation (Kurioka et al., 2015; Le Bourhis et al., 2013; Leeansyah et al., 2015), and *GNLY*$^{+}$ MAIT cells failed to express canonical effector genes such as *GZMB*. The absence of shared marker genes across donors and the lack of effector molecule expression following stimulation means the exact nature of these cells remains unclear. While *GNLY*-expressing MAIT cells might be a unique subset, further studies are required to clarify the nature and function of these cells.

My findings regarding the MAIT cell TCR repertoire were largely consistent with Chapter 5, and together these experiments provide the most comprehensive analysis of the MAIT cell TCR repertoire to date. However, the degree of MAIT cell clonality significantly differed in Experiment 1 and Experiment 2 (Fig. 6.10E). This is probably a consequence of the different age and/or disease status of the donors, however which of these factors is predominantly responsible for the observed difference is unknown. Nevertheless, the difference in clonality suggests preferential expansion or retention of particular MAIT cell clonotypes, likely in response to microbial infection. In accordance with this hypothesis, published data provides evidence for the expansion of certain MAIT cell clonotypes, and the contraction of others, during human *in vivo S*. Paratyphi A infection (Howson et al., 2018), as well as the activation of populations of MAIT cells with distinct TCR repertoires in response to different microbes *in vitro* (Gold et al., 2014).

The localisation of TCR clonotypes appeared largely random across clusters of unstimulated MAIT cells (Fig. 6.7A-C and 6.8). However, there were instances where TCR clonotypes were localised to particular clusters (Fig. 6.7D and 6.8), as observed to a greater degree in Chapter 5. In the example of clonotype 6 from donor 2794, the cluster in which the cells were localised displayed an activated transcriptional profile (Fig. 6.7D,E). The reason for the less striking association between clonotype and cluster in Experiment 2 is unclear, but factors including reduced oligoclonality of the population and resting of the samples prior to sequencing may have contributed. Regardless, consistent with clonotypes differing in their activation state, as implied by my findings in Chapter 5, pathway enrichment analysis of the clonotype markers combined from all samples of unstimulated MAIT cells showed enrichment for pathways associated with cytokine signalling (Fig. 6.9A), and clonotype markers highly overlapped with genes differentially expressed following MAIT cell activation (Fig. 6.9C). However, in contrast with Experiment 1, TCR signalling pathways were not enriched. Thus, Experiment 2 broadly supports the finding from Experiment 1 of clonotypes differing in their activation state, despite the more limited association between clonotypes and clusters.

Following TCR- and cytokine-dependent activation, MAIT cells followed distinct activation pathways. Activation-associated genes, for example *IFNG* and *IL7R*, showed a graded expression pattern along these trajectories. The Random Forest algorithm was used to identify the genes most predictive of progression along the two pathways. Genes highly predictive of progression along both pathways included *IL7R* (downregulated upon activation) and a number of genes associated with basic cellular functionality. With regards to immune-related genes, *CCL3* and *CCL4* were more predictive of progression along the TCR trajectory, whilst *IFNG* and *IL26* were more predictive of progression along the cytokine trajectory. The top

100 predictive genes for progression along the TCR and cytokine activation pathways could be divided into two modules of upregulated genes (early and late) and two modules of downregulated genes (early and late). In general, immune genes were upregulated at a fairly late stage following activation, with the exception of the proinflammatory cytokine *MIF* following TCR and cytokine stimulation, and *IFNG* following cytokine stimulation. Immune genes were found in both modules of downregulated genes. Therefore, this study has provided novel insights into the kinetics of gene expression following MAIT cell activation and has defined key genes associated with their activation state.

While statistical analysis identified a significant association between the clonal origin of MAIT cells and their position in pseudotime along the TCR and cytokine trajectories, the differences identified between clonotypes were small. Whether these differences are biologically meaningful will require further investigation. In contrast, cells from individual clonotypes showed very different levels of activation. Thus, my results suggest greater variability in MAIT cell activation capacity between cells from an individual clonotype rather than between different clonotypes. This implies that activation capacity varies stochastically amongst MAIT cells, or that another unidentified factor is responsible for variation in activation capacity. However, cell activation in this experiment was achieved using strong stimuli. Therefore, it is possible that clonal origin has a more important role in governing the activation potential of MAIT cells in response to weaker stimuli. This could explain the distinct clustering of TCR clonotypes in Experiment 1, if the cells had previously been activated *in vivo* by low concentrations of cytokines or by microbes expressing a variety of riboflavin metabolites. In Experiment 2, the differences in activation state between clonotypes may have been reduced due to *in vitro* culture.

6.7. Limitations

Given the cost of scRNA-seq and scTCR-seq, I was only able to study three donors and three different stimulation conditions. Therefore, while my data provide evidence for distinct TCR and cytokine activation trajectories, it is unclear if and how the pathways would differ following MAIT cell activation with alternative riboflavin metabolites or different combinations of cytokines. The pathway(s) of activation following combined TCR and cytokine stimulation, remains unknown.

As pseudotime trajectories are derived from multiple cells, rather than by following an individual cell over time, it is not possible to say for certain that the activation trajectories identified accurately reflect the true pathways of MAIT cell activation. However, the expression of known activation-associated genes, such as *GZMB*, *IFNG* and *IL7R*, progressively increased or decreased along the trajectory, suggesting the trajectories identified by Slingshot and SCORPIUS are good approximations of the pathways followed by TCR- and cytokine-activated MAIT cells. Differences were identified in the genes most predictive of progression along the two pathways. Previous studies have shown differences in the kinetics of TCR- and cytokine-dependent MAIT cell activation (Lamichhane et al., 2019b; Slichter et al., 2016). As MAIT cells were stimulated for 20 hours with both stimuli, differences between the pathways could reflect differences in their kinetics, or genuine differences in the transcriptional response to TCR and cytokine stimulation.

While MAIT cells differed in their level of activation, subsets with distinct functional capacity were not identified. It is possible that 20 hours of stimulation is not sufficient for the differentiation of any potential MAIT cell subsets. For example, IL-

13 production is detected only after the stimulation of MAIT cells for several days with PHA, IL-2, IL-17 and αCD3/αCD28 (Kelly et al., 2019). After 8 days of culture, ~10% of MAIT cells have the capacity to produce IL-13. These IL-13-producing MAIT cells may represent a distinct MAIT cell subset or alternatively may reflect the most activated MAIT cells at the end of a linear activation trajectory. Alternatively, detection of functionally distinct MAIT cell subsets may require activation with particular stimuli. For example, very few IL-17-producing MAIT cells were detected following TCR or IL-12 + IL-18 stimulation (data not shown). IL-7 can enhance the capacity of MAIT cells to produce IL-17 upon αCD3/αCD28 stimulation, with maximal production reached after several days (Tang et al., 2013). Whether IL-17-producing MAIT cells comprise a distinct functional subset remains unknown. Moreover, specialised subsets of MAIT cells could be restricted to certain tissues, for example IL-10-producing MAIT cells in adipose tissue (Carolan et al., 2015). Therefore, my data do not provide any evidence for functional MAIT cell subsets, but their existence in other settings remains possible.

Although I did not identify a major association between MAIT cell clonal origin and the capacity for activation, strong stimuli were used. Thus, it may be that TCR usage has a greater effect on activation capacity in response to weaker riboflavin metabolites or suboptimal concentrations of cytokines. In previous studies, differences in the ligand affinity of MAIT cell TCRs and TCR-dependent changes in responsiveness to stimulation were largely identified in the presence of microbes or weak riboflavin ligands, for example RL-6,7-diMe (Dias et al., 2017; Eckle et al., 2014; Gold et al., 2014; Howson et al., 2018). Therefore, it may be that the stimuli chosen for this study were not optimal for identifying potential differences in activation capacity between clonotypes. Moreover, use of a single stimulus may not reflect the setting of natural infection *in vivo*, where multiple riboflavin ligands may

be present simultaneously in the presence or absence of cytokines. Thus, how my findings may differ from the *in vivo* situation, will require further study.

As there was no apparent difference in the global transcriptional profile of *CD8*$^+$, DN and *CD4*$^+$ MAIT cells in Experiment 1, I decided to focus on CD8$^+$ MAIT cells for Experiment 2. Consequently, whether DN and CD4$^+$ MAIT cells follow the same activation trajectories as CD8$^+$ MAIT cells is unknown.

6.8. Conclusion

Consistent with Chapter 5, my findings demonstrate that MAIT cells comprise multiple clusters in the blood that largely reflect different states of activation rather than distinct differentiated populations. By performing analysis of the MAIT cell TCR repertoire on a further three donors, I have validated my findings from Chapter 5, and together my experiments provide a detailed characterisation of the resting MAIT cell TCR repertoire. Of particular interest, my results show for the first time that the paired TCRαβ repertoire of MAIT cells is largely donor-specific, despite large amounts of sharing in TCRα chain usage. My data suggests that MAIT cells do not comprise functionally distinct subsets. Instead, upon activation they progress along a linear trajectory from non-activated to highly activated cells. While the trajectory is distinct depending on the stimulus, the genes differentially expressed along the two pathways are highly overlapping. MAIT cell activation in the presence of strong stimuli is largely independent of clonal origin, although TCR usage may have a minor role. Further studies should investigate the role of MAIT cell clonal origin *in vivo* and in settings of suboptimal stimulation.

6.9. Summary

- Described the MAIT cell TCR repertoire in unprecedented detail.
- Characterised distinct linear trajectories for TCR- and cytokine-activated MAIT cells.
- Demonstrated that the activation capacity of MAIT cells in response to TCR and cytokine stimuli is largely clonotype independent.
- Critically, failed to support the model that human MAIT cells are comprised of functionally distinct subsets.

Table 6.1. List of unique clonotype markers.

ANKRD28	*GFPT2*	*MT-CO3*	*RPL15*
ANXA1	*GNA15*	*MT-ND1*	*RPL32*
AOAH	*GNG5*	*MT-ND3*	*RRS1*
BACH2	*GNLY*	*MT1F*	*S100B*
BZW1	*GPR183*	*MYL6*	*SELENOH*
C1orf21	*GSTP1*	*NAB2*	*SGK1*
CACNA2D4	*GZMK*	*NCR3*	*SH3BGRL3*
CAMK4	*HIVEP3*	*NDC1*	*SIGLEC9*
CCL20	*HLA-DPA1*	*NFKBID*	*SRGN*
CCL3	*HOPX*	*NINJ1*	*STK17B*
CCL3L1	*HSP90AB1*	*NKG7*	*TKTL1*
CCL4	*HSPA8*	*NOSIP*	*TMEM273*
CCL4L2	*IL27RA*	*NR4A1*	*TMIGD2*
CCL5	*IL32*	*NR4A2*	*TNF*
CCT5	*IL4I1*	*NRN1*	*TNFAIP8*
CD27	*IL7R*	*PDCD1*	*TNFRSF12A*
CD3D	*IMP4*	*PFN1*	*TNFRSF18*
CD8A	*IRF4*	*PHLDA1*	*TNFRSF9*
CD8B	*KIR3DL2*	*PKM*	*TRAF1*
CISD3	*KLRD1*	*PLK3*	*TRAF2*
CMC1	*KLRF1*	*PLP2*	*TSPO*
CST7	*LPAR6*	*PTPN6*	*TSPOAP1-AS1*
EGR1	*LTA*	*PTPN7*	*TYROBP*
EGR2	*LTB*	*PUF60*	*WDR46*
EGR3	*LYST*	*RAB8B*	*YBX1*
ELP5	*MDM2*	*RARA*	*ZFP36L1*
GAPDH	*MIR155HG*	*REL*	

Table 6.2. Top 100 genes predictive of progression along the SCORPIUS TCR activation trajectory.

Gene	Importance
HSP90AB1	28.1424966
RAN	27.4537441
IL7R	25.8078061
TUBA1B	24.7728592
IL2RA	24.3178899
FABP5	23.5404454
NPM1	22.3410201
PKM	20.1247097
PRDX1	19.4337787
EIF5A	18.99374
HSPE1	18.2713191
MIR155HG	18.2496312
HSPD1	18.1681006
HSPA8	17.5266499
RPS27	17.191133
TNFRSF9	17.1248176
PTMA	17.0742711
CHCHD2	17.0717266
ITM2B	16.7263784
LAPTM5	16.6935214
TNFRSF18	15.7099883
NCL	14.9099737
HSP90AA1	14.8656121
RPL26	14.6001371
MT-ND3	14.5093783
NOP16	14.0425697
ENO1	13.6108617
C1QBP	13.4817494
CCL4	13.4713873
GZMB	13.2049593
CALR	13.160869
GZMK	13.0417973
PPIA	12.8089564
TUBB	12.4701922
LTA	11.904572
NHP2	11.8120276
SARAF	11.4093827
CCL3	10.8402172
TUBA1C	10.7685807
KLRB1	10.6955852
PPA1	10.5910057
EBNA1BP2	10.374637
S100A4	9.73205892

RPS29	9.61287996
SEC61B	9.52727869
CYCS	9.47523242
CCL20	9.40321853
PFN1	9.36543218
PHB	9.2546867
IL4I1	9.1179246
MTHFD2	8.7899285
TC2N	8.77945087
SNRPD1	8.74568833
MIF	8.69182399
CCL5	8.5904637
SERBP1	8.5567617
RPL30	8.13250009
DDX21	8.11053901
YBX1	8.07131363
DCTPP1	8.0265543
PTPRCAP	7.90956151
GZMA	7.54257466
PRMT1	7.53117256
ANXA1	7.20961618
BTG1	7.0044084
GIMAP4	6.90983775
ALOX5AP	6.9047793
SLC1A5	6.72567048
TNF	6.61242575
MRTO4	6.58946404
PAICS	6.58512151
RPL34	6.48478273
NPM3	6.46968228
ATP1B3	6.42363058
HSPA5	6.35666951
AHNAK	6.13793809
PSME2	6.09910047
ZFP36L2	6.07001194
EEF1D	6.0241081
CCT2	5.8804617
TNFRSF4	5.64937486
LDHA	5.62738795
ZNF593	5.60239846
BCL2A1	5.54183853
TPI1	5.37894178
CCT6A	5.22279916
CREM	5.1931357
MYC	5.18834297
CDK4	5.13154879
FKBP4	5.09987173

CD52	5.04817761
WARS	5.01656774
RANBP1	4.79547792
CXCR4	4.70112656
TRAF4	4.67288521
ARHGDIB	4.65547563
PGAM1	4.6095607
NOLC1	4.4244835
RBM8A	4.41670891
PDCD5	4.35425123

Table 6.3. Top 100 genes predictive of progression along the SCORPIUS cytokine activation trajectory.

Gene	Importance
IFNG	33.475409
HSP90AB1	29.3055912
IL7R	27.3543403
IL26	26.4482576
LAPTM5	25.3669961
TUBA1B	24.4730241
PKM	23.641624
CD52	21.5075026
CYTIP	20.4951347
SEC61B	20.1874942
ITM2B	19.967224
EIF1	19.7697335
GZMB	19.2673436
CHCHD2	18.6831639
RPS27	18.6638175
ZBTB32	18.4492943
HSP90AA1	18.1632584
SEC61G	18.0446833
MTHFD2	17.5518924
ENO1	17.5204672
IL2RA	17.4159789
S100A4	17.1916533
TNFRSF18	17.022453
RPL26	16.9452231
MIR155HG	16.7823958
CYCS	16.7619769
SARAF	16.4609815
RPL30	15.7759467
ANXA1	15.7023928
RAN	15.5780431
CD3E	15.2592441
NFKBIA	14.9382811
RPL34	13.296858
MYDGF	13.2668638
JAML	13.1903345
EIF5A	13.1414506
KLRB1	12.7136247
CCL5	12.6632511
RBM8A	11.8436625
PRMT1	11.7815524
RPS29	11.4052606
PRDX1	11.0832353
TMSB4X	10.8878988

ALOX5AP	10.7732502
MT-ND3	10.7202734
FEZ1	10.3301154
HSPA5	10.2558933
RPS15A	9.94118967
NPM1	9.89886941
PPIB	9.60544493
RPL37	9.52807115
LSP1	9.38876781
HSPD1	9.3492758
HIF1A	9.30853151
PSMA6	9.02681908
DDX21	8.92084022
LTB	8.80625034
FABP5	8.76368504
GIMAP4	8.69615874
CCL20	8.57606425
MT-ATP8	8.5627347
HSPE1	8.46473787
BATF	8.39500876
TC2N	8.19384821
PFDN5	7.80542344
MIF	7.73783702
AHNAK	7.69539056
SH3BGRL3	7.54382215
PTPRCAP	7.40657077
EBNA1BP2	7.24923415
NOP16	7.12965011
CST7	7.11275935
PSME2	6.95592392
SPOCK2	6.87753608
RPL13	6.83285117
C1QBP	6.75145284
TOMM7	6.64256482
TPI1	6.46246465
ZFP36L2	6.37970798
EVL	6.23514452
MT2A	6.19970527
GZMH	6.15157588
PFDN2	6.1457894
SNRPD1	5.97412471
PPIA	5.664899
PDCD4	5.60825581
TUBA1C	5.60040537
ACTB	5.56800094
NCL	5.46617506
RPS21	5.40915586

NAMPT	5.26815472
RPL35A	5.25162028
STIP1	4.94669585
CCT2	4.90260312
RPL39	4.89793311
CXCR4	4.84437673
RPS14	4.78649758
WARS	4.76179657
GZMK	4.7610239
PBXIP1	4.67086379

7. Discussion

The aim of this thesis was to characterise the regulation and heterogeneity of human MAIT cells. In summary, I have: (1) characterised the unique transcriptional and chromatin accessibility profile of MAIT cells relative to conventional T_{mem} cells (Fig. 7.1), (2) described the transcriptional and TCR repertoire heterogeneity of blood and liver MAIT cells (Fig. 7.2), and (3) delineated MAIT cell activation trajectories following TCR and cytokine stimulation (Fig. 7.2).

In addition to their distinctive transcriptional profile, MAIT cells displayed a unique chromatin accessibility landscape (Fig. 7.1). Based on the genes most upregulated in expression and accessibility, MAIT cells appeared primed for both innate-like and type 17 functionality compared with T_{mem} cells. While the type 17-biased transcriptome and chromatin accessibility landscape of MAIT cells is consistent with their expression of type 17-associated molecules under homeostatic conditions, including the IL-23R, RORγt and CCR6 (Billerbeck et al., 2010; Dusseaux et al., 2011), it is somewhat at odds with the relatively limited capacity of MAIT cells to exhibit type 17 functionality, in particular IL-17 secretion (Provine et al., 2018). In contrast, MAIT cells rapidly upregulate other effector molecules, such as IFNγ and granzyme B, following TCR- and cytokine-dependent activation (Kurioka et al., 2015; Leeansyah et al., 2015; Slichter et al., 2016; Ussher et al., 2014), suggesting particularly tight regulation of type 17 functionality. Further work should investigate the optimal conditions for driving type 17 functions, especially as MAIT cells show increased IL-17 production in several diseases, including inflammatory bowel disease (Haga et al., 2016; Serriari et al., 2014) and multiple sclerosis (Willing et al., 2018).

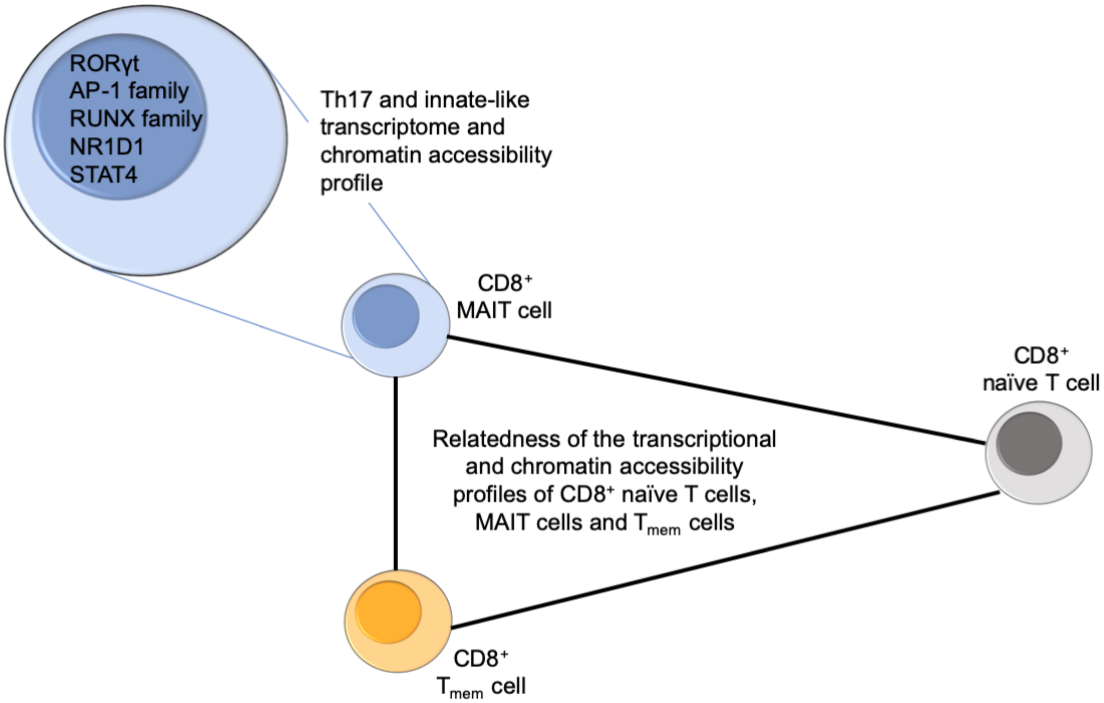


Figure 7.1. Summary of Chapter 4 results – transcriptional and chromatin accessibility landscape of human blood MAIT cells. $CD8^+$ human blood MAIT cells displayed a transcriptional profile and chromatin accessibility landscape that was distinct from both $CD8^+$ naïve T cells and $CD8^+$ T_{mem} cells. MAIT cells and T_{mem} cells were more similar to one another than either was to naïve T cells in relation to both gene expression and chromatin accessibility, consistent with their overlapping phenotype and function. Compared with T_{mem} cells, MAIT cells showed upregulated expression and accessibility of genes associated with innate-like and type 17 functionality. Motif enrichment and IMAGE (predicts causal transcriptional regulators) analysis suggested an important role for known (e.g. RORγt) and novel (e.g. AP-1 family, RUNX family, NR1D1, STAT4) TFs in regulating MAIT cell biology.

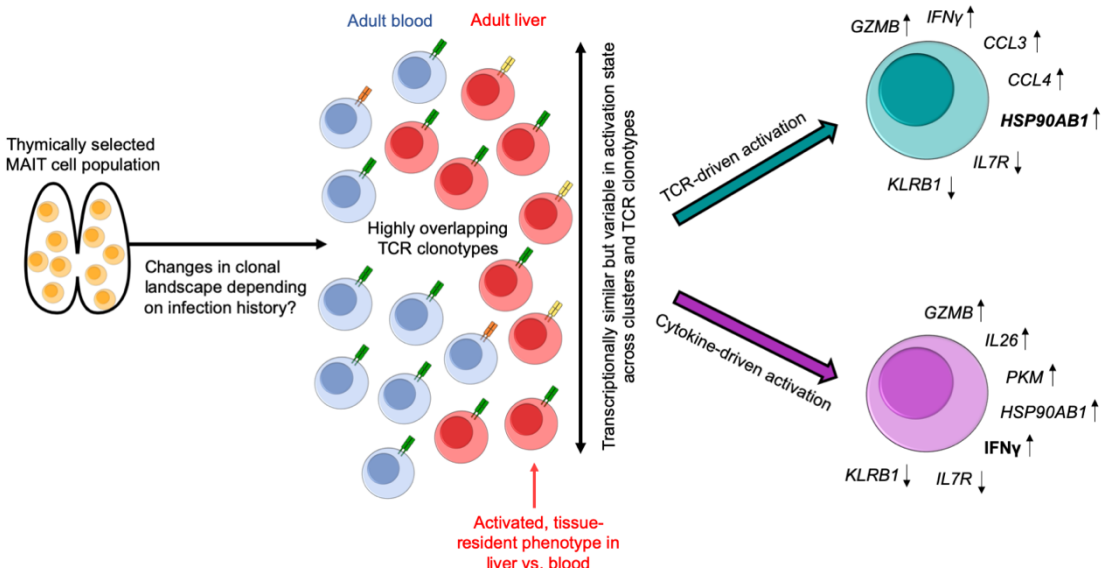


Figure 7.2. Summary of Chapter 5 and 6 results – single-cell transcriptional analysis of human blood and liver MAIT cells. MAIT cells in the liver were transcriptionally distinct from blood MAIT cells and displayed an activated, tissue-resident phenotype. Within the blood and liver, MAIT cells comprised multiple transcriptional clusters that reflected different states of activation, rather than fundamentally distinct subsets. MAIT cells displayed an oligoclonal TCR repertoire with high clonal sharing between blood and liver. In contrast, the TCR repertoire was essentially unique to each individual. Similar to clusters, clonotypes displayed variable expression of activation-associated molecules, perhaps reflecting differences between clonotypes in the time since *in vivo* activation. Following *in vitro* stimulation via the TCR or with cytokines, MAIT cells did not comprise functionally distinct subsets. Instead, individual MAIT cells fell along a linear trajectory from non-activated to highly activated cells, with a distinct activation pathway identified depending on the stimulus. While the expression of a number of genes was highly predictive of progression along both pathways (e.g. *GZMB*, *IL7R*), genes with greater importance for either the TCR (e.g. *CCL3*, *CCL4*) or cytokine (e.g. *IL26*) pathway, were also identified. The gene most highly predictive of progression along the TCR or cytokine activation pathway is indicated in bold.

The reduced accessibility of type 2 genes in MAIT cells is consistent with their general lack of type 2 functionality, and with the absence of a MAIT2 subset in mice (Koay et al., 2019a; 2016; Legoux et al., 2019b; Rahimpour et al., 2015). However, human MAIT cells have recently been shown to secrete type 2 cytokines following prolonged stimulation *in vitro* (Kelly et al., 2019), implying that MAIT cells are more functionally plastic than currently appreciated. The need for chromatin remodelling might provide one explanation for the time delay in acquisition of type 2 functionality. Given the lack of functional subsets identified in my scRNA-seq data from activated MAIT cells and the fact that the majority of IL-13-producing MAIT cells coexpress TNFα (Kelly et al., 2019), I hypothesise that type 2-secreting MAIT cells do not comprise a distinct functional subset, but instead localise towards the end of an activation trajectory, the specific nature of which will be stimulus-dependent.

As well as confirming a key role for RORγt in the regulation of MAIT cell phenotype and function, matched RNA-seq and ATAC-seq analysis of MAIT cells allowed me to identify several other TFs or families of TFs with potentially important roles in MAIT cells, including the AP-1 and RUNX families, NR1D1, and STAT4 (Fig. 7.1), the roles of which should now be investigated experimentally. MAIT cells can have protective and pathogenic roles in disease (Toubal et al., 2019) and in any given disease, could simultaneously exhibit behaviours of both a protective and pathogenic nature. For example, a recent study demonstrated a pathological role for MAIT cells in several mouse models of cancer, which was in part dependent on MAIT cell IL-17A production (Yan et al., 2019). However, MAIT cells can kill cancer cell lines *in vitro* (Gherardin et al., 2018a; Ling et al., 2016; Won et al., 2016; Yan et al., 2019), suggesting their cytotoxic activity might be beneficial in cancer. Understanding the regulation of MAIT cell functions, including the role of different

TFs, will therefore be important for the potential development of therapeutics that can modulate specific aspects of MAIT cell function.

My scRNA-seq data demonstrates that MAIT cells in the blood and liver are transcriptionally distinct (Fig. 7.2), validating and extending prior findings using bulk RNA-seq (Salou et al., 2018). In general, liver MAIT cells appeared to be an activated, tissue-resident counterpart of blood MAIT cells. While I did not identify any unique TFs expressed by liver MAIT cells, a number of TFs, including several AP-1 family members (e.g. *JUN*, *JUNB*, *FOSB*) and *RUNX2*, showed upregulated expression in the liver compared with the blood, while *MYC* was downregulated in liver MAIT cells. This suggests tissue-specific roles for these TFs. Differences in the expression of AP-1 family members and *RUNX2* between MAIT cells of varying phenotype further supports a potentially important role for these TFs in regulating MAIT cell phenotype and/or function, as suggested by my ATAC-seq data. Deletion of the AP-1 factor *Junb* in immune cells is protective in mouse models of hepatitis that involve NK/NKT cells, and is associated with reduced secretion of IFNγ by NK/NKT cells, which is hepatotoxic in this setting (Thomsen et al., 2013). Given the similar biology of iNKT cells and MAIT cells (Garner et al., 2018), jun-B may also regulate MAIT cell IFNγ production in the liver, indicating a possible avenue for initial investigations into the role of AP-1 factors in MAIT cells.

Within the blood and liver, MAIT cells comprised multiple clusters that varied in the expression of TCR and cytokine signalling pathways, activation-associated genes and NK cell receptors. NK cell receptors show diverse expression on resting MAIT cells and can exhibit altered expression upon MAIT cell activation (Dias et al., 2017). My results therefore imply that previously identified heterogeneity in the phenotype and function of MAIT cells within individual tissues reflects variation in

the activation state of MAIT cells, rather than the presence of functionally distinct MAIT cell subsets. This is further supported by my scRNA-seq data from TCR- and cytokine-activated MAIT cells, which demonstrated the presence of stimulus-specific MAIT cell activation trajectories (Fig. 7.2), rather than subsets of MAIT cells with distinct functional capacity. Altogether, my findings indicate that human MAIT cells comprise a single subset, as suggested by their mixed type 1/type 17 phenotype (Dias et al., 2017), in contrast with mouse MAIT cells that comprise developmentally and transcriptionally distinct MAIT1 and MAIT17 subsets (Koay et al., 2019a; 2016; Legoux et al., 2019b; Rahimpour et al., 2015). The reason for this disparity between human and mouse MAIT cell populations remains unclear.

The activation trajectories identified for TCR- and cytokine-stimulated MAIT cells showed considerable overlap in the genes predictive of progression from a non-activated to an activated MAIT cell (Fig. 7.2). Nevertheless, some clear differences could be identified. *IFNG* and *IL26* were uniquely predictive of progression along the cytokine activation pathway, while the chemokines *CCL3* and *CCL4* were better predictors of progression along the TCR activation pathway. As well as reflecting differences in TCR- and cytokine-induced functional responses, this may partly reflect differing kinetics of gene expression depending on the stimulus. For example, IFNγ protein expression peaks earlier upon TCR compared with cytokine stimulation (Slichter et al., 2016).

Antimicrobial activity through IFNγ production is often considered the main function of MAIT cells as evidenced by their ability to promote bacterial clearance (e.g. following *Legionella* infection (Wang et al., 2018)). However, reduced IFNγ production following TCR compared with cytokine stimulation (Slichter et al., 2016) suggests that, in the absence of costimulation (e.g. with cytokines), the main

response of MAIT cells to TCR triggering might not be antimicrobial. This hypothesis is supported by recent studies demonstrating that TCR but not cytokine stimulation induces a tissue repair program in MAIT cells, which can promote wound healing in an experimental model (Constantinides et al., 2019; Hinks et al., 2019; Lamichhane et al., 2019b; Leng et al., 2019). Thus, in the steady state, the primary function of MAIT cells appears to be maintaining homeostasis, for example through tissue repair, and it is only in the context of an "emergency" (e.g. infection) that MAIT cells play an antimicrobial role.

The MAIT cell TCR repertoire was highly shared across matched blood and liver (Fig. 7.2), but was essentially unique to each individual due to diverse TCRβ usage. Nevertheless, TCRα chains were highly shared across individuals. Tyrosine 95 of the conserved CDR3α region forms a hydrogen bond with the ribityl moiety of stimulatory riboflavin metabolites and is essential for MAIT cell activation (Corbett et al., 2014; Eckle et al., 2014; Patel et al., 2013; Reantragoon et al., 2012), explaining the essentially fixed antigen specificity of MAIT cells. It is possible that diverse TCRβ usage has no functional consequence. However, evidence for distinct TCR usage following stimulation with different pathogens *in vitro* (Gold et al., 2014), altered affinity of the MAIT cell TCR for riboflavin metabolites dependent on TCRβ sequence (Eckle et al., 2014), and contraction and expansion of different clonotypes during human *in vivo* infection with *S*. Paratyphi A infection (Howson et al., 2018), suggests otherwise. In this study, I identified differences in the activation state of MAIT cell clonotypes amongst resting MAIT cells. However, as both TCR and cytokine signalling pathways varied between clonotypes, and clonotypes showed altered expression of cell surface molecules known to influence their cytokine responsiveness, whether the identified differences between clonotypes directly result from differences in the antigen affinity or specificity of their TCRs, is unclear.

In contrast with the varying activation state of MAIT cell clonotypes under resting conditions, following *in vitro* activation via the TCR or with cytokines, all MAIT cells had the capacity to respond to stimulation and this was largely independent of clonotype. Although this appears contradictory with my findings from resting MAIT cells, it may be explained by the strong stimuli used for activation in this study. It is plausible that under conditions of suboptimal stimulation, for example low concentrations of antigen or cytokines, or in the presence of weaker stimulatory ligands, differences between clonotypes in their TCR and/or cytokine responsiveness may be more evident. The clonality of MAIT cells was reduced in Experiment 2 relative to Experiment 1, perhaps due to the younger age of the donors in Experiment 2 or differences in the health status of the donors from the two experiments. Either way, the individuals in Experiment 1 have likely been exposed to a greater number or variety of MAIT cell activating stimuli. The reduced clonality of their TCR repertoires therefore supports clonotype-specific expansion of MAIT cells in response to microbial exposure, consistent with my findings that MAIT cell clonotypes differ in activation state when analysed *ex vivo*.

7.1. Limitations

Given the cost of single-cell sequencing, I was limited by the number of donors and conditions that I could study. This is a particular problem for findings that showed some inconsistency between donors. For example, the association between clonotypes and clusters was generally stronger in Experiment 1 than Experiment 2, but also showed variability within each experiment. This may be related to the experimental setup of the two experiments or to donor characteristics. Testing this would require a large number of samples, which is prohibitive due to cost. In

Experiment 2, the activation trajectories identified were largely consistent across donors. However, I was only able to define the MAIT cell functional response to TCR and cytokine stimulation for 20 hours. It is unclear what the trajectory would look like if MAIT cells were stimulated with bacteria, with both TCR and cytokine stimuli, and/or for different durations. Moreover, MAIT cells expressed very little IL-17 under the conditions tested. Thus, I am unable to comment on whether IL-17-producing MAIT cells comprise a distinct subset, although my data suggest that they most likely represent a particular state of activation.

My data indicate that DN, $CD4^+$ and $CD8^+$ MAIT cells are not fundamentally distinct subsets, consistent with recent mouse scRNA-seq data from multiple tissues (Constantinides et al., 2019). However, identifying coreceptor expression using scRNA-seq is imperfect due to gene dropouts (Stegle et al., 2015). For example, in some $CD4^+$ MAIT cells, expression of *CD4* would not have been detected and therefore these cells would have been classified as DN. This issue can now be easily overcome using cell hashing (Stoeckius et al., 2018). Even accounting for this limitation, I would expect all *CD4*-expressing MAIT cells to cluster together if they comprised a truly distinct subset and this was not observed.

Due to tissue availability and the feasibility of collecting a sufficient number of MAIT cells for in-depth single-cell profiling, I was only able to assess the transcriptional profile of MAIT cells in the blood and liver. Although MAIT cells did not comprise functionally distinct subsets within the liver, this does not preclude the existence of MAIT cell subsets in other tissues. In particular, in mucosal tissues such as the gut, MAIT cells will be continually exposed to a diversity of commensal bacteria, many of which will be capable of producing MAIT cell stimulatory riboflavin metabolites, and could influence the phenotype and function of MAIT cells. In some tissues, there is

evidence for altered MAIT cell functionality. For example, MAIT cells display a type 17-skewed phenotype in the female genital tract and oral mucosa (Gibbs et al., 2016; Sobkowiak et al., 2018), and show the unique capacity to produce IL-10 in adipose tissue (Carolan et al., 2015). Therefore, these sites would be of particular interest for further investigation into the potential presence of tissue-specific MAIT cell subsets.

As the technology was not readily available at the time of the experiment, ATAC-seq was performed at a bulk rather than a single-cell level. Therefore, it was not possible to assess the degree of heterogeneity in the chromatin accessibility profile of blood MAIT cells. Given my scRNA-seq findings, namely that MAIT cells do not appear to comprise truly distinct subsets, this is unlikely to have considerably impacted on my conclusions. Nevertheless, given the heterogeneity of conventional T_{mem} cells, single-cell ATAC-seq would have allowed a more detailed comparison of MAIT cells with different memory subsets. For example, MAIT cells showed increased accessibility at a number of type 17 loci relative to T_{mem} cells. It would be interesting to address how the accessibility of these loci and enrichment of TF motifs compares between MAIT cells and Tc17 cells.

7.2. Future experiments

I am keen to perform scRNA-seq and scTCR-seq on MAIT cells in additional tissues. This would allow me to address whether my findings from the liver, namely that liver MAIT cells do not comprise distinct subsets and that the TCR repertoire is highly shared between the blood and liver, are applicable to other tissues. Now that spatial scRNA-seq technologies are more widely available, it would be very

interesting to explore how MAIT cell clusters and clonotypes are localised in three-dimensional space within particular tissues. Given the altered activation states between clonotypes within the liver, it is plausible that these clonotypes may be differentially spatially located within the tissue and thus exposed to a different microenvironment.

Exploring longitudinal changes in the transcriptional profile and clonal landscape of MAIT cells during development or disease would be a useful follow-up to this study. Given the reduced clonality of the MAIT cell TCR repertoire in donors from Experiment 2 relative to Experiment 1, I hypothesise that the MAIT cell TCR repertoire becomes more restricted with repeated microbial encounters due to expansion of particular MAIT cell clonotypes depending on the specific environmental conditions, for example combination of microbial metabolites and cytokines. Expansion and contraction of MAIT cell clonotypes has so far only been observed *in vivo* in human *S*. Paratyphi A infection (Howson et al., 2018). Thus, how the MAIT cell TCR repertoire changes during other infections, particular during chronic infection, is unknown. Longitudinal analyses would also allow me to address whether the differences I observed in the activation state of clonotypes amongst resting MAIT cells reflect the time since their most recent activation *in vivo*.

As well as longitudinally studying changes in the transcriptome and TCR repertoire, I am interested to explore how the chromatin accessibility landscape of MAIT cells changes under different stimulation conditions, for example under type 17-promoting conditions (Tang et al., 2013), or following prolonged stimulation when MAIT cells acquire type 2 functionality (Kelly et al., 2019). Interesting areas to explore include the enrichment of TF motifs in newly accessible regions, the time taken for changes in chromatin accessibility, and the stability of altered chromatin states. This may

help to address the reversibility of observed changes in MAIT cell functionality during disease.

7.3. Concluding remarks

In summary, I have demonstrated that MAIT cells are distinct from conventional T_{mem} cells in their transcriptome and chromatin accessibility profile, and have identified some TFs with potentially important roles in regulating this unique T cell population. Heterogeneity in the phenotype and TCR repertoire of MAIT cells has previously been identified, both within and between different tissues. However, it was unknown whether human MAIT cells comprise multiple distinct subsets. I have shown through scRNA-seq of resting blood and liver MAIT cells, and TCR- and cytokine-stimulated MAIT cells, that human MAIT cells fundamentally comprise a single subset, at least in the blood and the liver. Heterogeneity within the MAIT cell population appears to reflect variation in MAIT cell activation state that depends in part on both their clonal origin and tissue localisation. Finally, this study has comprehensively described the features of the MAIT cell TCR repertoire, including a highly restricted TCRα chain and diverse TCRβ usage, resulting in a paired TCRαβ repertoire that is essentially unique to each individual. In this thesis, I have provided the first description of the chromatin accessibility landscape of human MAIT cells, and have provided the most detailed description of the phenotype, TCR composition and functionality of human MAIT cells. These data can serve as the basis for future investigations of MAIT cell biology in health and disease.

References

Alexeyev, M., Shokolenko, I., Wilson, G., and LeDoux, S. (2013). The maintenance of mitochondrial DNA integrity--critical analysis and update. Cold Spring Harb Perspect Biol *5*, a012641–a012641.

Amemiya, H.M., Kundaje, A., and Boyle, A.P. (2019). The ENCODE Blacklist: Identification of Problematic Regions of the Genome. Sci. Rep. *9*, 9354.

Amir, M., Chaudhari, S., Wang, R., Campbell, S., Mosure, S.A., Chopp, L.B., Lu, Q., Shang, J., Pelletier, O.B., He, Y., et al. (2018). REV-ERBα Regulates TH17 Cell Development and Autoimmunity. Cell Rep *25*, 3733–3749.e3738.

Arase, H., Arase, N., and Saito, T. (1996). Interferon γ production by natural killer (NK) cells and NK1.1+ T cells upon NKR-P1 cross-linking. J Exp Med *183*, 2391–2396.

Bacher, R., and Kendziorski, C. (2016). Design and computational analysis of single-cell RNA-sequencing experiments. Genome Biol *17*, 63.

Baev, D.V., Peng, X.-H., Song, L., Barnhart, J.R., Crooks, G.M., Weinberg, K.I., and Metelitsa, L.S. (2004). Distinct homeostatic requirements of CD4+ and CD4- subsets of Vα24-invariant natural killer T cells in humans. Blood *104*, 4150–4156.

Bailey, T.L., Bodén, M., Buske, F.A., Frith, M., Grant, C.E., Clementi, L., Ren, J., Li, W.W., and Noble, W.S. (2009). MEME SUITE: tools for motif discovery and searching. Nucleic Acids Res. *37*, W202–W208.

Barnitz, R.A., Imam, S., Yates, K., and Haining, W.N. (2013). Isolation of RNA and the synthesis and amplification of cDNA from antigen-specific T cells for genome-wide expression analysis. Methods Mol. Biol. *979*, 161–173.

Bánki, Z., Krabbendam, L., Klaver, D., Leng, T., Kruis, S., Mehta, H., Müllauer, B., Orth-Höller, D., Stoiber, H., Willberg, C.B., et al. (2019). Antibody opsonization enhances MAIT cell responsiveness to bacteria via a TNF dependent mechanism. Immunol Cell Biol.

Ben Youssef, G., Tourret, M., Salou, M., Ghazarian, L., Houdouin, V., Mondot, S., Mburu, Y., Lambert, M., Azarnoush, S., Diana, J.-S., et al. (2018). Ontogeny of human mucosal-associated invariant T cells and related T cell subsets. J Exp Med *151*, jem.20171739–21.

Bendelac, A., Killeen, N., Littman, D.R., and Schwartz, R.H. (1994). A subset of CD4+ thymocytes selected by MHC class I molecules. Science *263*, 1774–1778.

Bennett, M.S., Trivedi, S., Iyer, A.S., Hale, J.S., and Leung, D.T. (2017). Human mucosal-associated invariant T (MAIT) cells possess capacity for B-cell help. J. Leukoc. Biol. jlb.4A0317–116R.

Beura, L.K., Fares-Frederickson, N.J., Steinert, E.M., Scott, M.C., Thompson, E.A., Fraser, K.A., Schenkel, J.M., Vezys, V., and Masopust, D. (2019). CD4+ resident

memory T cells dominate immunosurveillance and orchestrate local recall responses. J Exp Med jem.20181365.

Bianchi, T., Gasser, S., Trumpp, A., and MacDonald, H.R. (2006). c-Myc acts downstream of IL-15 in the regulation of memory CD8 T-cell homeostasis. Blood *107*, 3992–3999.

Billerbeck, E., Kang, Y.-H., Walker, L., Lockstone, H., Grafmueller, S., Fleming, V., Flint, J., Willberg, C.B., Bengsch, B., Seigel, B., et al. (2010). Analysis of CD161 expression on human CD8+ T cells defines a distinct functional subset with tissue-homing properties. Pnas *107*, 3006–3011.

Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics *30*, 2114–2120.

Booth, J.S., Salerno-Goncalves, R., Blanchard, T.G., Patil, S.A., Kader, H.A., Safta, A.M., Morningstar, L.M., Czinn, S.J., Greenwald, B.D., and Sztein, M.B. (2015). Mucosal-Associated Invariant T Cells in the Human Gastric Mucosa and Blood: Role in Helicobacter pylori Infection. Front. Immunol. *6*, 742–14.

Boyle, A.P., Davis, S., Shulha, H.P., Meltzer, P., Margulies, E.H., Weng, Z., Furey, T.S., and Crawford, G.E. (2008). High-resolution mapping and characterization of open chromatin across the genome. Cell *132*, 311–322.

Böttcher, K., Rombouts, K., Saffioti, F., Roccarina, D., Rosselli, M., Hall, A., Luong, T., Tsochatzis, E.A., Thorburn, D., and Pinzani, M. (2018). MAIT cells are chronically activated in patients with autoimmune liver disease and promote profibrogenic hepatic stellate cell activation. Hepatology *68*, 172–186.

Brandt, D., and Hedrich, C.M. (2018). TCRαβ+CD3+CD4-CD8- (double negative) T cells in autoimmunity. Autoimmun Rev *17*, 422–430.

Breiman, L. (2001). Random Forests. Machine Learning *45*, 5–32.

Buck, M.D., O'Sullivan, D., and Pearce, E.L. (2015). T cell metabolism drives immunity. J Exp Med *212*, 1345–1360.

Buenrostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y., and Greenleaf, W.J. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat Meth *10*, 1213–1218.

Buenrostro, J.D., Wu, B., Chang, H.Y., and Greenleaf, W.J. (2015a). ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. Curr Protoc Mol Biol *109*, 21.29.1–.29.9.

Buenrostro, J.D., Wu, B., Litzenburger, U.M., Ruff, D., Gonzales, M.L., Snyder, M.P., Chang, H.Y., and Greenleaf, W.J. (2015b). Single-cell chromatin accessibility reveals principles of regulatory variation. Nature *523*, 486–490.

Butler, A., Hoffman, P., Smibert, P., Papalexi, E., and Satija, R. (2018). Integrating single-cell transcriptomic data across different conditions, technologies, and species. Nat. Biotechnol. *36*, 411–420.

Cannoodt, R., Saelens, W., Sichien, D., Tavernier, S., Janssens, S., Guilliams, M., Lambrecht, B., De Preter, K., and Saeys, Y. (2016). SCORPIUS improves trajectory inference and identifies novel modules in dendritic cell development. bioRxiv *17*, 079509.

Carolan, E., Tobin, L.M., Mangan, B.A., Corrigan, M., Gaoatswe, G., Byrne, G., Geoghegan, J., Cody, D., O'Connell, J., Winter, D.C., et al. (2015). Altered distribution and increased IL-17 production by mucosal-associated invariant T cells in adult and childhood obesity. J. Immunol. *194*, 5775–5780.

Castro, G., Liu, X., Ngo, K., De Leon-Tabaldo, A., Zhao, S., Luna-Roman, R., Yu, J., Cao, T., Kuhn, R., Wilkinson, P., et al. (2017). RORγt and RORα signature genes in human Th17 cells. PLoS ONE *12*, e0181868.

Chen, Z., Wang, H., D'Souza, C., Sun, S., Kostenko, L., Eckle, S.B.G., Meehan, B.S., Jackson, D.C., Strugnell, R.A., Cao, H., et al. (2017). Mucosal-associated invariant T-cell activation and accumulation after in vivo infection depends on microbial riboflavin synthesis and co-stimulatory signals. Mucosal Immunol *10*, 58–68.

Cheroutre, H., and Lambolez, F. (2008). Doubting the TCR coreceptor function of CD8αα. Immunity *28*, 149–159.

Chiba, A., Tajima, R., Tomi, C., Miyazaki, Y., Yamamura, T., and Miyake, S. (2012). Mucosal-associated invariant T cells promote inflammation and exacerbate disease in murine models of arthritis. Arthritis Rheum. *64*, 153–161.

Chitadze, G., Oberg, H.-H., Wesch, D., and Kabelitz, D. (2017). The Ambiguous Role of γδ T Lymphocytes in Antitumor Immunity. Trends in Immunol *38*, 668–678.

Consortium, T.E.P. (2012). An integrated encyclopedia of DNA elements in the human genome. Nature *489*, 57–74.

Constantinides, M.G., Link, V.M., Tamoutounour, S., Wong, A.C., Perez-Chaparro, P.J., Han, S.-J., Chen, Y.E., Li, K., Farhat, S., Weckel, A., et al. (2019). MAIT cells are imprinted by the microbiota in early life and promote tissue repair. Science *366*, eaax6624–15.

Corbett, A.J., Eckle, S.B.G., Birkinshaw, R.W., Liu, L., Patel, O., Mahony, J., Chen, Z., Reantragoon, R., Meehan, B., Cao, H., et al. (2014). T-cell activation by transitory neo-antigens derived from distinct microbial pathways. Nature *509*, 361–365.

Corces, M.R., Buenrostro, J.D., Wu, B., Greenside, P.G., Chan, S.M., Koenig, J.L., Snyder, M.P., Pritchard, J.K., Kundaje, A., Greenleaf, W.J., et al. (2016). Lineage-specific and single-cell chromatin accessibility charts human hematopoiesis and leukemia evolution. Nat Genet *48*, 1193–1203.

Corces, M.R., Granja, J.M., Shams, S., Louie, B.H., Seoane, J.A., Zhou, W., Silva, T.C., Groeneveld, C., Wong, C.K., Cho, S.W., et al. (2018). The chromatin accessibility landscape of primary human cancers. Science *362*, eaav1898.

Corces, M.R., Trevino, A.E., Hamilton, E.G., Greenside, P.G., Sinnott-Armstrong, N.A., Vesuna, S., Satpathy, A.T., Rubin, A.J., Montine, K.S., Wu, B., et al. (2017). An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. Nat Meth *14*, 959–962.

Cosgrove, C., Ussher, J.E., Rauch, A., Gärtner, K., Kurioka, A., Hühn, M.H., Adelmann, K., Kang, Y.-H., Fergusson, J.R., Simmonds, P., et al. (2013). Early and nonreversible decrease of CD161++ /MAIT cells in HIV infection. Blood *121*, 951–961.

Croxford, J.L., Miyake, S., Huang, Y.-Y., Shimamura, M., and Yamamura, T. (2006). Invariant Vα19i T cells regulate autoimmune inflammation. Nat Immunol *7*, 987–994.

Cui, Y., Franciszkiewicz, K., Mburu, Y.K., Mondot, S., Le Bourhis, L., Premel, V., Martin, E., Kachaner, A., Duban, L., Ingersoll, M.A., et al. (2015). Mucosal-associated invariant T cell-rich congenic mouse strain allows functional evaluation. J. Clin. Invest. *125*, 4171–4185.

Cusanovich, D.A., Daza, R., Adey, A., Pliner, H.A., Christiansen, L., Gunderson, K.L., Steemers, F.J., Trapnell, C., and Shendure, J. (2015). Multiplex single-cell profiling of chromatin accessibility by combinatorial cellular indexing. Science *348*, 910–914.

D'Souza, C., Pediongco, T., Wang, H., Scheerlinck, J.-P.Y., Kostenko, L., Esterbauer, R., Stent, A.W., Eckle, S.B.G., Meehan, B.S., Strugnell, R.A., et al. (2018). Mucosal-Associated Invariant T Cells Augment Immunopathology and Gastritis in Chronic Helicobacter pylori Infection. J. Immunol. ji1701512.

Dash, P., McClaren, J.L., Oguin, T.H., Rothwell, W., Todd, B., Morris, M.Y., Becksfort, J., Reynolds, C., Brown, S.A., Doherty, P.C., et al. (2011). Paired analysis of TCRα and TCRβ chains at the single-cell level in mice. J. Clin. Invest. *121*, 288–295.

Davey, M.S., Willcox, C.R., Baker, A.T., Hunter, S., and Willcox, B.E. (2018). Recasting Human Vδ1 Lymphocytes in an Adaptive Role. Trends in Immunol *39*, 446–459.

DePasquale, E.A.K., Schnell, D.J., Van Camp, P.-J., Valiente-Alandí, Í., Blaxall, B.C., Grimes, H.L., Singh, H., and Salomonis, N. (2019). DoubletDecon: Deconvoluting Doublets from Single-Cell RNA-Sequencing Data. Cell Rep *29*, 1718–.

Dias, J., Boulouis, C., Gorin, J.-B., van den Biggelaar, R.H.G.A., Lal, K.G., Gibbs, A., Loh, L., Gulam, M.Y., Sia, W.R., Bari, S., et al. (2018). The CD4-CD8- MAIT cell subpopulation is a functionally distinct subset developmentally related to the main CD8+ MAIT cell pool. Pnas *5*, 201812273.

Dias, J., Leeansyah, E., and Sandberg, J.K. (2017). Multiple layers of heterogeneity and subset diversity in human MAIT cell responses to distinct microorganisms and to innate cytokines. Pnas *114*, E5434–E5443.

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

Duan, M., Goswami, S., Shi, J.-Y., Wu, L.-J., Wang, X.-Y., Ma, J.-Q., Zhang, Z., Shi, Y., Ma, L.-J., Zhang, S., et al. (2019). Activated and Exhausted MAIT Cells Foster Disease Progression and Indicate Poor Outcome in Hepatocellular Carcinoma. Clin. Cancer Res. clincanres.3040.2018.

Durinck, S., Moreau, Y., Kasprzyk, A., Davis, S., De Moor, B., Brazma, A., and Huber, W. (2005). BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis. Bioinformatics *21*, 3439–3440.

Dusseaux, M., Martin, E., Serriari, N., Péguillet, I., Premel, V., Louis, D., Milder, M., Le Bourhis, L., Soudais, C., Treiner, E., et al. (2011). Human MAIT cells are xenobiotic-resistant, tissue-targeted, CD161hi IL-17–secreting T cells. Blood *117*, 1250–1259.

Dutta, M., Kraus, Z.J., Gomez-Rodriguez, J., Hwang, S.-H., Cannons, J.L., Cheng, J., Lee, S.-Y., Wiest, D.L., Wakeland, E.K., and Schwartzberg, P.L. (2013). A role for Ly108 in the induction of promyelocytic zinc finger transcription factor in developing thymocytes. J. Immunol. *190*, 2121–2128.

Eckle, S.B.G., Birkinshaw, R.W., Kostenko, L., Corbett, A.J., McWilliam, H.E.G., Reantragoon, R., Chen, Z., Gherardin, N.A., Beddoe, T., Liu, L., et al. (2014). A molecular basis underpinning the T cell receptor heterogeneity of mucosal-associated invariant T cells. J Exp Med *211*, 1585–1600.

Eltahla, A.A., Rizzetto, S., Pirozyan, M.R., Betz-Stablein, B.D., Venturi, V., Kedzierska, K., Lloyd, A.R., Bull, R.A., and Luciani, F. (2016). Linking the T cell receptor to the single cell transcriptome in antigen-specific human T cells. Immunol Cell Biol *94*, 604–611.

Emgård, J., Bergsten, H., McCormick, J.K., Barrantes, I., Skrede, S., Sandberg, J.K., and Norrby-Teglund, A. (2019). MAIT Cells Are Major Contributors to the Cytokine Response in Group A Streptococcal Toxic Shock Syndrome. Pnas *116*, 25923–25931.

Fergusson, J.R., Huhn, M.H., Swadling, L., Walker, L.J., Kurioka, A., Llibre, A., Bertoletti, A., Holländer, G., Newell, E.W., Davis, M.M., et al. (2015). CD161intCD8+ T cells: a novel population of highly functional, memory CD8+ T cells enriched within the gut. Mucosal Immunol *9*, 401–413.

Fergusson, J.R., Smith, K.E., Fleming, V.M., Rajoriya, N., Newell, E.W., Simmons, R., Marchi, E., Björkander, S., Kang, Y.-H., Swadling, L., et al. (2014). CD161 Defines a Transcriptional and Functional Phenotype across Distinct Human T Cell Lineages. Cell Rep *9*, 1075–1088.

Fernandez, C.S., Amarasena, T., Kelleher, A.D., Rossjohn, J., McCluskey, J., Godfrey, D.I., and Kent, S.J. (2015). MAIT cells are depleted early but retain functional cytokine expression in HIV infection. Immunol Cell Biol *93*, 177–188.

Fornes, O., Castro-Mondragon, J.A., Khan, A., van der Lee, R., Zhang, X., Richmond, P.A., Modi, B.P., Correard, S., Gheorghe, M., Baranašić, D., et al. (2019). JASPAR 2020: update of the open-access database of transcription factor binding profiles. Nucleic Acids Res. *172*, 650.

Garner, L.C., Klenerman, P., and Provine, N.M. (2018). Insights Into Mucosal-Associated Invariant T Cell Biology From Studies of Invariant Natural Killer T Cells. Front. Immunol. *9*, 1478.

Gehring, J., Park, J.H., Chen, S., Thomson, M., and Pachter, L. (2018). Highly Multiplexed Single-Cell RNA-seq for Defining Cell Population and Transcriptional Spaces. bioRxiv *8*, 315333.

Georgel, P., Radosavljevic, M., Macquin, C., and Bahram, S. (2011). The non-conventional MHC class I MR1 molecule controls infection by Klebsiella pneumoniae in mice. Mol Immunol *48*, 769–775.

Gerlach, C., Moseman, E.A., Loughhead, S.M., Alvarez, D., Zwijnenburg, A.J., Waanders, L., Garg, R., la Torre, de, J.C., and Andrian, von, U.H. (2016). The Chemokine Receptor CX3CR1 Defines Three Antigen-Experienced CD8 T Cell Subsets with Distinct Roles in Immune Surveillance and Homeostasis. Immunity *45*, 1270–1284.

Gérart, S., Sibéril, S., Martin, E., Lenoir, C., Aguilar, C., Picard, C., Lantz, O., Fischer, A., and Latour, S. (2013). Human iNKT and MAIT cells exhibit a PLZF-dependent proapoptotic propensity that is counterbalanced by XIAP. Blood *121*, 614–623.

Gherardin, N.A., Keller, A.N., Woolley, R.E., Le Nours, J., Ritchie, D.S., Neeson, P.J., Birkinshaw, R.W., Eckle, S.B.G., Waddington, J.N., Liu, L., et al. (2016). Diversity of T Cells Restricted by the MHC Class I-Related Molecule MR1 Facilitates Differential Antigen Recognition. Immunity *44*, 32–45.

Gherardin, N.A., Loh, L., Admojo, L., Davenport, A.J., Richardson, K., Rogers, A., Darcy, P.K., Jenkins, M.R., Prince, H.M., Harrison, S.J., et al. (2018a). Enumeration, functional responses and cytotoxic capacity of MAIT cells in newly diagnosed and relapsed multiple myeloma. Sci. Rep. *8*, 4159.

Gherardin, N.A., Souter, M.N.T., Koay, H.F., Mangas, K.M., Seemann, T., Stinear, T.P., Eckle, S.B.G., Berzins, S.P., d'Udekem, Y., Konstantinov, I.E., et al. (2018b). Human blood MAIT cell subsets defined using MR1 tetramers. Immunol Cell Biol *96*, 507–525.

Gibbs, A., Leeansyah, E., Introini, A., Paquin-Proulx, D., Hasselrot, K., Andersson, E., Broliden, K., Sandberg, J.K., and Tjernlund, A. (2016). MAIT cells reside in the female genital mucosa and are biased towards IL-17 and IL-22 production in response to bacterial stimulation. Mucosal Immunol *10*, 35–45.

Giresi, P.G., Kim, J., McDaniell, R.M., Iyer, V.R., and Lieb, J.D. (2007). FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements) isolates active regulatory elements from human chromatin. Genome Res. *17*, 877–885.

Giudicelli, V., and Lefranc, M.-P. (2011). IMGT/JunctionAnalysis: IMGT Standardized Analysis of the V-J and V-D-J Junctions of the Rearranged Immunoglobulins (IG) and T Cell Receptors (TR). Cold Spring Harb Protoc *2011*, pdb.prot5634–pdb.prot5634.

Godec, J., Tan, Y., Liberzon, A., Tamayo, P., Bhattacharya, S., Butte, A.J., Mesirov, J.P., and Haining, W.N. (2016). Compendium of Immune Signatures Identifies Conserved and Species-Specific Biology in Response to Inflammation. Immunity *44*, 194–206.

Godfrey, D.I., Koay, H.F., McCluskey, J., and Gherardin, N.A. (2019). The biology and functional importance of MAIT cells. Nat Immunol *20*, 1110–1128.

Gold, M.C., Eid, T., Smyk-Pearson, S., Eberling, Y., Swarbrick, G.M., Langley, S.M., Streeter, P.R., Lewinsohn, D.A., and Lewinsohn, D.M. (2013). Human thymic MR1-restricted MAIT cells are innate pathogen-reactive effectors that adapt following thymic egress. Mucosal Immunol *6*, 35–44.

Gold, M.C., Cerri, S., Smyk-Pearson, S., Cansler, M.E., Vogt, T.M., Delepine, J., Winata, E., Swarbrick, G.M., Chua, W.-J., Yu, Y.Y.L., et al. (2010). Human Mucosal Associated Invariant T Cells Detect Bacterially Infected Cells. PLOS Biol *8*, e1000407–e1000414.

Gold, M.C., McLaren, J.E., Reistetter, J.A., Smyk-Pearson, S., Ladell, K., Swarbrick, G.M., Yu, Y.Y.L., Hansen, T.H., Lund, O., Nielsen, M., et al. (2014). MR1-restricted MAIT cells display ligand discrimination and pathogen selectivity through distinct T cell receptor usage. J Exp Med *211*, 1601–1610.

Goldstein, I., and Hager, G.L. (2018). Dynamic enhancer function in the chromatin context. Wiley Interdiscip Rev Syst Biol Med *10*, e1390.

Gordy, L.E., Bezbradica, J.S., Flyak, A.I., Spencer, C.T., Dunkle, A., Sun, J., Stanic, A.K., Boothby, M.R., He, Y.-W., Zhao, Z., et al. (2011). IL-15 regulates homeostasis and terminal maturation of NKT cells. J Immunol *187*, 6335–6345.

Goronzy, J.J., and Weyand, C.M. (2017). Successful and Maladaptive T Cell Aging. Immunity *46*, 364–378.

Greenaway, H.Y., Ng, B., Price, D.A., Douek, D.C., Davenport, M.P., and Venturi, V. (2013). NKT and MAIT invariant TCRα sequences can be produced efficiently by VJ gene recombination. Immunobiology *218*, 213–224.

Griewank, K., Borowski, C., Rietdijk, S., Wang, N., Julien, A., Wei, D.G., Mamchak, A.A., Terhorst, C., and Bendelac, A. (2007). Homotypic interactions mediated by Slamf1 and Slamf6 receptors control NKT cell lineage development. Immunity *27*, 751–762.

Guo, C., Kong, W., Kamimoto, K., Rivera-Gonzalez, G.C., Yang, X., Kirita, Y., and Morris, S.A. (2019). CellTag Indexing: genetic barcode-based sample multiplexing for single-cell genomics. Genome Biol *20*, 90–13.

Guo, J., Hawwari, A., Li, H., Sun, Z., Mahanta, S.K., Littman, D.R., Krangel, M.S., and He, Y.-W. (2002). Regulation of the TCRα repertoire by the survival window of CD4(+)CD8(+) thymocytes. Nat Immunol *3*, 469–476.

Hafemeister, C., and Satija, R. (2019). Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. bioRxiv *3*, 576827.

Haga, K., Chiba, A., Shibuya, T., and Osada, T. (2016). MAIT cells are activated and accumulated in the inflamed mucosa of ulcerative colitis. Jgh *31*, 965–972.

Hama, I., Tominaga, K., Yamagiwa, S., Setsu, T., Kimura, N., Kamimura, H., Wakai, T., and Terai, S. (2019). Different distribution of mucosal-associated invariant T cells within the human cecum and colon. Cent Eur J Immunol *44*, 75–83.

Han, A., Glanville, J., Hansmann, L., and Davis, M.M. (2014). Linking T-cell receptor sequence to functional phenotype at the single-cell level. Nat. Biotechnol. *32*, 684–692.

Haque, M., Song, J., Fino, K., Wang, Y., Sandhu, P., Song, X., Norbury, C., Ni, B., Fang, D., Salek-Ardakani, S., et al. (2016). C-Myc regulation by costimulatory signals modulates the generation of CD8+ memory T cells during viral infection. Open Biol *6*, 150208.

Hariprakash, J.M., and Ferrari, F. (2019). Computational Biology Solutions to Identify Enhancers-target Gene Pairs. Computational and Structural Biotechnology Journal *17*, 821–831.

Hashimoto, K., Hirai, M., and Kurosawa, Y. (1995). A gene outside the human MHC related to classical HLA class I genes. Science *269*, 693–695.

Hegde, P., Weiss, E., Paradis, V., Wan, J., Mabire, M., Sukriti, S., Rautou, P.-E., Albuquerque, M., Picq, O., Gupta, A.C., et al. (2018). Mucosal-associated invariant T cells are a profibrogenic immune cell population in the liver. Nat Commun *9*, 2146.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol Cell *38*, 576–589.

Held, K., Beltrán, E., Moser, M., Hohlfeld, R., and Dornmair, K. (2015). T-cell receptor repertoire of human peripheral CD161hiTRAV1-2+ MAIT cells revealed by next generation sequencing and single cell analysis. Human Immunology *76*, 607–614.

Hengst, J., Strunz, B., Deterding, K., Ljunggren, H.-G., Leeansyah, E., Manns, M.P., Cornberg, M., Sandberg, J.K., Wedemeyer, H., and Björkström, N.K. (2016). Nonreversible MAIT cell-dysfunction in chronic hepatitis C virus infection despite successful interferon-free therapy. Eur J Immunol *46*, 2204–2210.

Hess, J., Angel, P., and Schorpp-Kistner, M. (2004). AP-1 subunits: quarrel and harmony among siblings. J. Cell. Sci. *117*, 5965–5973.

Hiejima, E., Kawai, T., Nakase, H., Tsuruyama, T., Morimoto, T., Yasumi, T., Taga, T., Kanegane, H., Hori, M., Ohmori, K., et al. (2015). Reduced Numbers and Proapoptotic Features of Mucosal-associated Invariant T Cells as a Characteristic Finding in Patients with Inflammatory Bowel Disease. Inflamm. Bowel Dis. *21*, 1529–1540.

Hinks, T.S.C., Marchi, E., Jabeen, M., Olshansky, M., Kurioka, A., Pediongco, T.J., Meehan, B.S., Kostenko, L., Turner, S.J., Corbett, A.J., et al. (2019). Activation and In Vivo Evolution of the MAIT Cell Transcriptome in Mice and Humans Reveals Tissue Repair Functionality. Cell Rep *28*, 3249–3262.e5.

Hosoya, T., Albanus, R.D., Hensley, J., Myers, G., Kyono, Y., Kitzman, J., Parker, S.C.J., and Engel, J.D. (2018). Global dynamics of stage-specific transcription factor binding during thymocyte development. Sci. Rep. *8*, 5605.

Howson, L.J., Napolitani, G., Shepherd, D., Ghadbane, H., Kurupati, P., Preciado-Llanes, L., Rei, M., Dobinson, H.C., Gibani, M.M., Teng, K.W.W., et al. (2018). MAIT cell clonal expansion and TCR repertoire shaping in human volunteers challenged with Salmonella Paratyphi A. Nat Commun *9*, 303.

Huang, C., and Kanagawa, O. (2001). Ordered and coordinated rearrangement of the TCRα locus: role of secondary rearrangement in thymic selection. J Immunol *166*, 2597–2601.

Huang, S., Shen, Y., Pham, D., Jiang, L., Wang, Z., Kaplan, M.H., Zhang, G., and Sun, J. (2017). IRF4 Modulates CD8+ T Cell Sensitivity to IL-2 Family Cytokines. ImmunoHorizons *1*, 92–100.

Hume, M.A., Barrera, L.A., Gisselbrecht, S.S., and Bulyk, M.L. (2015). UniPROBE, update 2015: new tools and content for the online database of protein-binding microarray data on protein-DNA interactions. Nucleic Acids Res. *43*, D117–D122.

Ilicic, T., Kim, J.K., Kolodziejczyk, A.A., Bagger, F.O., McCarthy, D.J., Marioni, J.C., and Teichmann, S.A. (2016). Classification of low quality cells from single-cell RNA-seq data. Genome Biol *17*, 29.

Intlekofer, A.M., Takemoto, N., Wherry, E.J., Longworth, S.A., Northrup, J.T., Palanivel, V.R., Mullen, A.C., Gasink, C.R., Kaech, S.M., Miller, J.D., et al. (2005). Effector and memory CD8+ T cell fate coupled by T-bet and eomesodermin. Nat Immunol *6*, 1236–1244.

Ishihara, S., Nieda, M., Kitayama, J., Osada, T., Yabe, T., Ishikawa, Y., Nagawa, H., Muto, T., and Juji, T. (1999). CD8+NKR-P1A+ T cells preferentially accumulate in human liver. Eur J Immunol *29*, 2406–2413.

Ivanov, I.I., McKenzie, B.S., Zhou, L., Tadokoro, C.E., Lepelley, A., Lafaille, J.J., Cua, D.J., and Littman, D.R. (2006). The Orphan Nuclear Receptor RORγt Directs the Differentiation Program of Proinflammatory IL-17+ T Helper Cells. Cell *126*, 1121–1133.

Jeffery, H.C., van Wilgenburg, B., Kurioka, A., Parekh, K., Stirling, K., Roberts, S., Dutton, E.E., Hunter, S., Geh, D., Braitch, M.K., et al. (2016). Biliary epithelium and

liver B cells exposed to bacteria activate intrahepatic MAIT cells through MR1. J. Hepatol. *64*, 1118–1127.

Jiang, J., Cao, Z., Shan, W., Liu, H., and Cheng, X. (2018). 4-1BB expression on MAIT cells is associated with enhanced IFN-γ production and depends on IL-2. Cell. Immunol.

Jiang, J., Wang, X., An, H., Yang, B., Cao, Z., Liu, Y., Su, J., Zhai, F., Wang, R., Zhang, G., et al. (2014). MAIT Cell Function is Modulated by PD-1 Signaling in Patients with Active Tuberculosis. Am. J. Respir. Crit. Care Med. 140630082918004.

Jo, J., Tan, A.T., Ussher, J.E., Sandalova, E., Tang, X.-Z., Tan-Garcia, A., To, N., Hong, M., Chia, A., Gill, U.S., et al. (2014). Toll-like receptor 8 agonist and bacteria trigger potent activation of innate immune cells in human liver. PLoS Pathog *10*, e1004210–e1004213.

Jolma, A., Yan, J., Whitington, T., Toivonen, J., Nitta, K.R., Rastas, P., Morgunova, E., Enge, M., Taipale, M., Wei, G., et al. (2013). DNA-Binding Specificities of Human Transcription Factors. Cell *152*, 327–339.

Juárez-Rebollar, D., Rios, C., Nava-Ruíz, C., and Méndez-Armenta, M. (2017). Metallothionein in Brain Disorders. Oxidative Medicine and Cellular Longevity *2017*.

Kang, H.M., Subramaniam, M., Targ, S., Nguyen, M., Maliskova, L., McCarthy, E., Wan, E., Wong, S., Byrnes, L., Lanata, C.M., et al. (2018). Multiplexed droplet single-cell RNA-sequencing using natural genetic variation. Nat. Biotechnol. *36*, 89–94.

Karabacak Calviello, A., Hirsekorn, A., Wurmus, R., Yusuf, D., and Ohler, U. (2019). Reproducible inference of transcription factor footprints in ATAC-seq and DNase-seq datasets via protocol-specific bias modeling. Genome Biol *20*, 42.

Kawachi, I., Maldonado, J., Strader, C., and Gilfillan, S. (2006). MR1-Restricted Vα19i Mucosal-Associated Invariant T Cells Are Innate T Cells in the Gut Lamina Propria That Provide a Rapid and Diverse Cytokine Response. J Immunol *176*, 1618–1627.

Keller, A.N., Eckle, S.B.G., Xu, W., Liu, L., Hughes, V.A., Mak, J.Y.W., Meehan, B.S., Pediongco, T., Birkinshaw, R.W., Chen, Z., et al. (2017). Drugs and drug-like molecules can modulate the function of mucosal-associated invariant T cells. Nat Immunol *486*, 554.

Kelly, J., Minoda, Y., Meredith, T., Cameron, G., Philipp, M.-S., Pellicci, D.G., Corbett, A.J., Kurts, C., Gray, D.H., Godfrey, D.I., et al. (2019). Chronically stimulated human MAIT cells are unexpectedly potent IL-13 producers. Immunol Cell Biol *200*, 1901.

Kempfer, R., and Pombo, A. (2019). Methods for mapping 3D chromosome architecture. Nat. Rev. Genet. *16*, 245–20.

Kjer-Nielsen, L., Patel, O., Corbett, A.J., Le Nours, J., Meehan, B., Liu, L., Bhati, M., Chen, Z., Kostenko, L., Reantragoon, R., et al. (2012). MR1 presents microbial vitamin B metabolites to MAIT cells. Nature *491*, 717–723.

Klemm, S.L., Shipony, Z., and Greenleaf, W.J. (2019). Chromatin accessibility and the regulatory epigenome. Nat. Rev. Genet. *20*, 207–220.

Koay, H.F., Su, S., Amann-Zalcenstein, D., Daley, S.R., Comerford, I., Miosge, L., Whyte, C.E., Konstantinov, I.E., d'Udekem, Y., Baldwin, T., et al. (2019a). A divergent transcriptional landscape underpins the development and functional branching of MAIT cells. Sci Immunol *4*, eaay6039.

Koay, H.F., Gherardin, N.A., Enders, A., Loh, L., Mackay, L.K., Almeida, C.F., Russ, B.E., Nold-Petry, C.A., Nold, M.F., Bedoui, S., et al. (2016). A three-stage intrathymic development pathway for the mucosal-associated invariant T cell lineage. Nat Immunol *17*, 1300–1311.

Koay, H.F., Gherardin, N.A., Xu, C., Seneviratna, R., Zhao, Z., Chen, Z., Fairlie, D.P., McCluskey, J., Pellicci, D.G., Uldrich, A.P., et al. (2019b). Diverse MR1-restricted T cells in mice and humans. Nat Commun *10*, 2243.

Komori, T., Yagi, H., Nomura, S., Yamaguchi, A., Sasaki, K., Deguchi, K., Shimizu, Y., Bronson, R.T., Gao, Y.H., Inada, M., et al. (1997). Targeted Disruption of Cbfa1 Results in a Complete Lack of Bone Formation owing to Maturational Arrest of Osteoblasts. Cell *89*, 755–764.

Koppejan, H., Jansen, D.T.S.L., Hameetman, M., Thomas, R., Toes, R.E.M., and van Gaalen, F.A. (2019). Altered composition and phenotype of mucosal-associated invariant T cells in early untreated rheumatoid arthritis. Arthritis Res. Ther. *21*, 3–7.

Korsunsky, I., Millard, N., Fan, J., Slowikowski, K., Zhang, F., Wei, K., Baglaenko, Y., Brenner, M., Loh, P.R., and Raychaudhuri, S. (2019). Fast, sensitive and accurate integration of single-cell data with Harmony. Nat Meth *16*, 1289–1296.

Kovalovsky, D., Uche, O.U., Eladad, S., Hobbs, R.M., Yi, W., Alonzo, E., Chua, K., Eidson, M., Kim, H.-J., Im, J.S., et al. (2008). The BTB-zinc finger transcriptional regulator PLZF controls the development of invariant natural killer T cell effector functions. Nat Immunol *9*, 1055–1064.

Krensky, A.M., and Clayberger, C. (2009). Biology and clinical relevance of granulysin. Tissue Antigens *73*, 193–198.

Kulakovskiy, I.V., Vorontsov, I.E., Yevshin, I.S., Sharipov, R.N., Fedorova, A.D., Rumynskiy, E.I., Medvedeva, Y.A., Magana-Mora, A., Bajic, V.B., Papatsenko, D.A., et al. (2018). HOCOMOCO: towards a complete collection of transcription factor binding models for human and mouse via large-scale ChIP-Seq analysis. Nucleic Acids Res. *46*, D252–D259.

Kurioka, A., Ussher, J.E., Cosgrove, C., Clough, C., Fergusson, J.R., Smith, K., Kang, Y.H., Walker, L.J., Hansen, T.H., Willberg, C.B., et al. (2015). MAIT cells are licensed through granzyme exchange to kill bacterially sensitized targets. Mucosal Immunol *8*, 429–440.

Kurioka, A., Jahun, A.S., Hannaway, R.F., Walker, L.J., Fergusson, J.R., Sverremark-Ekström, E., Corbett, A.J., Ussher, J.E., Willberg, C.B., and Klenerman, P. (2017). Shared and Distinct Phenotypes and Functions of Human CD161++ Vα7.2+ T Cell Subsets. Front. Immunol. *8*, 1031.

Kurioka, A., van Wilgenburg, B., Javan, R.R., Hoyle, R., van Tonder, A.J., Harrold, C.L., Leng, T., Howson, L.J., Shepherd, D., Cerundolo, V., et al. (2018). Diverse Streptococcus pneumoniae Strains Drive a Mucosal-Associated Invariant T-Cell Response Through Major Histocompatibility Complex class I-Related Molecule-Dependent and Cytokine-Driven Pathways. The Journal of Infectious Diseases *217*, 988–999.

Kurioka, A., Walker, L.J., Klenerman, P., and Willberg, C.B. (2016). MAIT cells: new guardians of the liver. Clin Transl Immunology *5*, e98.

Kuylenstierna, C., Björkström, N.K., Andersson, S.K., Sahlström, P., Bosnjak, L., Paquin-Proulx, D., Malmberg, K.-J., Ljunggren, H.-G., Moll, M., and Sandberg, J.K. (2011). NKG2D performs two functions in invariant NKT cells: direct TCR-independent activation of NK-like cytolysis and co-stimulation of activation by CD1d. Eur J Immunol *41*, 1913–1923.

Lafzi, A., Moutinho, C., Picelli, S., and Heyn, H. (2018). Tutorial: guidelines for the experimental design of single-cell RNA sequencing studies. Nat Protoc *13*, 2742–2757.

Lambert, S.A., Jolma, A., Campitelli, L.F., Das, P.K., Yin, Y., Albu, M., Chen, X., Taipale, J., Hughes, T.R., and Weirauch, M.T. (2018a). The Human Transcription Factors. Cell *172*, 650–665.

Lambert, S.A., Jolma, A., Campitelli, L.F., Das, P.K., Yin, Y., Albu, M., Chen, X., Taipale, J., Hughes, T.R., and Weirauch, M.T. (2018b). The Human Transcription Factors. Cell *172*, 650–665.

Lamichhane, R., Galvin, H., Hannaway, R.F., la Harpe, de, S.M., Munro, F., Tyndall, J.D., Vernall, A.J., McCall, J.L., Husain, M., and Ussher, J.E. (2019a). Type I interferons are important co-stimulatory signals during T cell receptor mediated human MAIT cell activation. Eur J Immunol eji.201948279.

Lamichhane, R., Schneider, M., la Harpe, de, S.M., Harrop, T.W.R., Hannaway, R.F., Dearden, P.K., Kirman, J.R., Tyndall, J.D.A., Vernall, A.J., and Ussher, J.E. (2019b). TCR- or Cytokine-Activated CD8+ Mucosal-Associated Invariant T Cells Are Rapid Polyfunctional Effectors That Can Coordinate Immune Responses. Cell Rep *28*, 3061–3076.e3065.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat Meth *9*, 357–359.

Lawson, V.J., Maurice, D., Silk, J.D., Cerundolo, V., and Weston, K. (2009). Aberrant selection and function of invariant NKT cells in the absence of AP-1 transcription factor Fra-2. J. Immunol. *183*, 2575–2584.

Le Bourhis, L., Dusseaux, M., Bohineust, A., Bessoles, S., Martin, E., Premel, V., Coré, M., Sleurs, D., Serriari, N.-E., Treiner, E., et al. (2013). MAIT cells detect and efficiently lyse bacterially-infected epithelial cells. PLoS Pathog *9*, e1003681.

Le Bourhis, L., Martin, E., Péguillet, I., Guihot, A., Froux, N., Coré, M., Lévy, E., Dusseaux, M., Meyssonnier, V., Premel, V., et al. (2010). Antimicrobial activity of mucosal-associated invariant T cells. Nat Immunol *11*, 701–708.

Lee, C.H., Zhang, H.H., Singh, S.P., Koo, L., Kabat, J., Tsang, H., Singh, T.P., and Farber, J.M. (2018). C/EBPδ drives interactions between human MAIT cells and endothelial cells that are important for extravasation. Elife *7*, 1–31.

Lee, O.-J., Cho, Y.-N., Kee, S.-J., Kim, M.-J., Jin, H.-M., Lee, S.-J., Park, K.-J., Kim, T.-J., Lee, S.-S., Kwon, Y.-S., et al. (2014). Circulating mucosal-associated invariant T cell levels and their cytokine levels in healthy adults. Exg *49*, 47–54.

Leeansyah, E., Ganesh, A., Quigley, M.F., Sönnerborg, A., Andersson, J., Hunt, P.W., Somsouk, M., Deeks, S.G., Martin, J.N., Moll, M., et al. (2013). Activation, exhaustion, and persistent decline of the antimicrobial MR1-restricted MAIT-cell population in chronic HIV-1 infection. Blood *121*, 1124–1135.

Leeansyah, E., Loh, L., Nixon, D.F., and Sandberg, J.K. (2014). Acquisition of innate-like microbial reactivity in mucosal tissues during human fetal MAIT-cell development. Nat Commun *5*, 3143.

Leeansyah, E., Svärd, J., Dias, J., Buggert, M., Nyström, J., Quigley, M.F., Moll, M., Sönnerborg, A., Nowak, P., and Sandberg, J.K. (2015). Arming of MAIT Cell Cytolytic Antimicrobial Activity Is Induced by IL-7 and Defective in HIV-1 Infection. PLoS Pathog *11*, e1005072.

Legoux, F., Bellet, D., Daviaud, C., Morr, El, Y., Darbois, A., Niort, K., Procopio, E., Salou, M., Gilet, J., Ryffel, B., et al. (2019a). Microbial metabolites control the thymic development of mucosal-associated invariant T cells. Science *366*, 494–499.

Legoux, F., Gilet, J., Procopio, E., Echasserieau, K., Bernardeau, K., and Lantz, O. (2019b). Molecular mechanisms of lineage decisions in metabolite-specific T cells. Nat Immunol *20*, 1244–1255.

Leng, T., Akther, H.D., Hackstein, C.-P., Powell, K., King, T., Friedrich, M., Christoforidou, Z., McCuaig, S., Neyazi, M., Arancibia-Carcamo, C.V., et al. (2019). TCR and Inflammatory Signals Tune Human MAIT Cells to Exert Specific Tissue Repair and Effector Functions. Cell Rep *28*, 3077–3091.e5.

Lepore, M., Kalinichenko, A., Calogero, S., Kumar, P., Paleja, B., Schmaler, M., Narang, V., Zolezzi, F., Poidinger, M., Mori, L., et al. (2017). Functionally diverse human T cells recognize non-microbial antigens presented by MR1. Elife *6*, e24476.

Lepore, M., Kalinichenko, A., Kalinicenko, A., Colone, A., Paleja, B., Singhal, A., Tschumi, A., Lee, B., Poidinger, M., Zolezzi, F., et al. (2014). Parallel T-cell cloning and deep sequencing of human MAIT cells reveal stable oligoclonal TCRβ repertoire. Nat Commun *5*, 3866.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079.

Li, Q.-J., Chau, J., Ebert, P.J.R., Sylvester, G., Min, H., Liu, G., Braich, R., Manoharan, M., Soutschek, J., Skare, P., et al. (2007). miR-181a is an intrinsic modulator of T cell sensitivity and selection. Cell *129*, 147–161.

Li, Q., Brown, J.B., Huang, H., and Bickel, P.J. (2011). Measuring reproducibility of high-throughput experiments. The Annals of Applied Statistics *5*, 1752–1779.

Li, Y., Huang, B., Jiang, X., Chen, W., Zhang, J., Wei, Y., Chen, Y., Lian, M., Bian, Z., Miao, Q., et al. (2018). Mucosal-Associated Invariant T Cells Improve Nonalcoholic Fatty Liver Disease Through Regulating Macrophage Polarization. Front. Immunol. *9*, 1994.

Li, Z., Schulz, M.H., Look, T., Begemann, M., Zenke, M., and Costa, I.G. (2019). Identification of transcription factor binding sites using ATAC-seq. Genome Biol *20*, 45–21.

Liao, Y., Smyth, G.K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics *30*, 923–930.

Linehan, J.L., Harrison, O.J., Han, S.-J., Byrd, A.L., Vujkovic-Cvijin, I., Villarino, A.V., Sen, S.K., Shaik, J., Smelkinson, M., Tamoutounour, S., et al. (2018). Non-classical Immunity Controls Microbiota Impact on Skin Immunity and Tissue Repair. Cell *172*, 784–796.e18.

Ling, L., Lin, Y., Zheng, W., Hong, S., Tang, X., Zhao, P., Li, M., Ni, J., Li, C., Wang, L., et al. (2016). Circulating and tumor-infiltrating mucosal associated invariant T (MAIT) cells in colorectal cancer patients. Sci. Rep. *6*, 20358–10.

Loh, L., Wang, Z., Sant, S., Koutsakos, M., Jegaskanda, S., Corbett, A.J., Liu, L., Fairlie, D.P., Crowe, J., Rossjohn, J., et al. (2016). Human mucosal-associated invariant T cells contribute to antiviral influenza immunity via IL-18-dependent activation. Pnas *113*, 10133–10138.

Luecken, M.D., and Theis, F.J. (2019). Current best practices in single-cell RNA-seq analysis: a tutorial. Mol. Syst. Biol. *15*, e8746.

Lynch, L., Michelet, X., Zhang, S., Brennan, P.J., Moseman, A., Lester, C., Besra, G., Vomhof-Dekrey, E.E., Tighe, M., Koay, H.F., et al. (2015). Regulatory iNKT cells lack expression of the transcription factor PLZF and control the homeostasis of Treg cells and macrophages in adipose tissue. Nat Immunol *16*, 85–95.

Madsen, J.G.S., Rauch, A., Van Hauwaert, E.L., Schmidt, S.F., Winnefeld, M., and Mandrup, S. (2018). Integrated analysis of motif activity and gene expression changes of transcription factors. Genome Res. *28*, 243–255.

Man, K., Gabriel, S.S., Liao, Y., Gloury, R., Preston, S., Henstridge, D.C., Pellegrini, M., Zehn, D., Berberich-Siebelt, F., Febbraio, M.A., et al. (2017). Transcription

Factor IRF4 Promotes CD8+ T Cell Exhaustion and Limits the Development of Memory-like T Cells during Chronic Infection. Immunity *47*, 1129–1141.e5.

Mark Benedict, X.Z. (2017). Non-alcoholic fatty liver disease: An expanded review. World Journal of Hepatology *9*, 715–732.

Martin, E., Treiner, E., Duban, L., Guerri, L., Laude, H., Toly, C., Premel, V., Devys, A., Moura, I.C., Tilloy, F., et al. (2009). Stepwise Development of MAIT Cells in Mouse and Human. PLOS Biol *7*, e1000054–12.

Martinet, L., and Smyth, M.J. (2015). Balancing natural killer cell activation through paired receptors. Nat Rev Immunol *15*, 243–254.

Matsuda, J.L., Gapin, L., Sidobre, S., Kieper, W.C., Tan, J.T., Ceredig, R., Surh, C.D., and Kronenberg, M. (2002). Homeostasis of Vα14i NKT cells. Nat Immunol *3*, 966–974.

McGinnis, C.S., Murrow, L.M., and Gartner, Z.J. (2019a). DoubletFinder: Doublet Detection in Single-Cell RNA Sequencing Data Using Artificial Nearest Neighbors. Cell Systems *8*, 329–337.e4.

McGinnis, C.S., Patterson, D.M., Winkler, J., Conrad, D.N., Hein, M.Y., Srivastava, V., Hu, J.L., Murrow, L.M., Weissman, J.S., Werb, Z., et al. (2019b). MULTI-seq: sample multiplexing for single-cell RNA sequencing using lipid-tagged indices. Nat Meth *16*, 619–626.

McInnes, L., Healy, J., and Melville, J. (2018). UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction.

McLeay, R.C., and Bailey, T.L. (2010). Motif Enrichment Analysis: a unified framework and an evaluation on ChIP data. BMC Bioinformatics 2010 11:1 *11*, 165.

McWilliam, H.E.G., Eckle, S.B.G., Theodossis, A., Liu, L., Chen, Z., Wubben, J.M., Fairlie, D.P., Strugnell, R.A., Mintern, J.D., McCluskey, J., et al. (2016). The intracellular pathway for the presentation of vitamin B-related antigens by the antigen-presenting molecule MR1. Nat Immunol *17*, 531–537.

Meermeier, E.W., Laugel, B.F., Sewell, A.K., Corbett, A.J., Rossjohn, J., McCluskey, J., Harriff, M.J., Franks, T., Gold, M.C., and Lewinsohn, D.M. (2016). Human TRAV1-2-negative MR1-restricted T cells detect S. pyogenes and alternatives to MAIT riboflavin-based antigens. Nat Commun *7*, 12506–12512.

Meierovics, A.I., and Cowley, S.C. (2016). MAIT cells promote inflammatory monocyte differentiation into dendritic cells during pulmonary intracellular infection. J Exp Med *213*, 2793–2809.

Meierovics, A., Yankelevich, W.-J.C., and Cowley, S.C. (2013). MAIT cells are critical for optimal mucosal immune responses during in vivo pulmonary bacterial infection. Pnas *110*, E3119–E3128.

Meyer, C.A., and Liu, X.S. (2014). Identifying and mitigating bias in next-generation sequencing methods for chromatin biology. Nat. Rev. Genet. *15*, 709–721.

Milner, J.J., Toma, C., Yu, B., Zhang, K., Omilusik, K., Phan, A.T., Wang, D., Getzler, A.J., Nguyen, T., Crotty, S., et al. (2017). Runx3 programs CD8+ T cell residency in non-lymphoid tissues and tumours. Nature *552*, 253–257.

Monod, M.Y., Giudicelli, V., Chaume, D., and Lefranc, M.P. (2004). IMGT/JunctionAnalysis: the first tool for the analysis of the immunoglobulin and T cell receptor complex V-J and V-D-J JUNCTIONs. Bioinformatics *20*, i379–i385.

Mootha, V.K., Lindgren, C.M., Eriksson, K.-F., Subramanian, A., Sihag, S., Lehar, J., Puigserver, P., Carlsson, E., Ridderstråle, M., Laurila, E., et al. (2003). PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. Nat Genet *34*, 1–7.

Moran, A.E., Holzapfel, K.L., Xing, Y., Cunningham, N.R., Maltzman, J.S., Punt, J., and Hogquist, K.A. (2011). T cell receptor signal strength in Treg and iNKT cell development demonstrated by a novel fluorescent reporter mouse. J Exp Med *208*, 1279–1289.

Moskowitz, D.M., Zhang, D.W., Hu, B., Le Saux, S., Yanes, R.E., Ye, Z., Buenrostro, J.D., Weyand, C.M., Greenleaf, W.J., and Goronzy, J.J. (2017). Epigenomics of human CD8 T cell differentiation and aging. Sci Immunol *2*, eaag0192.

Nagy, P.L., Cleary, M.L., Brown, P.O., and Lieb, J.D. (2003). Genomewide demarcation of RNA polymerase II transcription units revealed by physical fractionation of chromatin. Pnas *100*, 6364–6369.

Novak, J., Dobrovolny, J., Novakova, L., and Kozak, T. (2014). The Decrease in Number and Change in Phenotype of Mucosal-Associated Invariant T cells in the Elderly and Differences in Men and Women of Reproductive Age. Scand J Immunol *80*, 271–275.

O'Keeffe, J., Doherty, D.G., Kenna, T., Sheahan, K., O'Donoghue, D.P., Hyland, J.M., and O'Farrelly, C. (2004). Diverse populations of T cells with NK cell receptors accumulate in the human intestine in health and in colorectal cancer. Eur J Immunol *34*, 2110–2119.

Okada, S., Markle, J.G., Deenick, E.K., Mele, F., Averbuch, D., Lagos, M., Alzahrani, M., Al-Muhsen, S., Halwani, R., Ma, C.S., et al. (2015). Impairment of immunity to Candida and Mycobacterium in humans with bi-allelic RORC mutations. Science *349*, 606–613.

Okamoto, N., Kanie, O., Huang, Y.-Y., Fujii, R., Watanabe, H., and Shimamura, M. (2005). Synthetic alpha-mannosyl ceramide as a potent stimulant for an NKT cell repertoire bearing the invariant Valpha19-Jalpha26 TCR alpha chain. Chem. Biol. *12*, 677–683.

Olesin, E., Nayar, R., Saikumar-Lakshmi, P., and Berg, L.J. (2018). The Transcription Factor Runx2 Is Required for Long-Term Persistence of Antiviral CD8+ Memory T Cells. ImmunoHorizons *2*, 251–261.

Oo, Y.H., Shetty, S., and Adams, D.H. (2010). The role of chemokines in the recruitment of lymphocytes to the liver. Dig Dis *28*, 31–44.

Paquin-Proulx, D., Avelino-Silva, V.I., Santos, B.A.N., Silveira Barsotti, N., Siroma, F., Fernandes Ramos, J., Coracini Tonacio, A., Song, A., Maestri, A., Barros Cerqueira, N., et al. (2018). MAIT cells are activated in acute Dengue virus infection and after in vitro Zika virus infection. PLoS Negl Trop Dis *12*, e0006154.

Paquin-Proulx, D., Santos, B.A.N., Barsotti, N.S., Marinho, A.K.B.B., Kokron, C.M., Carvalho, K.I., Barros, M.T., Kalil, J., Elmacken, M., Sandberg, J.K., et al. (2017). Loss of Circulating Mucosal-Associated Invariant T Cells in Common Variable Immunodeficiency Is Associated with Immune Activation and Loss of Eomes and PLZF. ImmunoHorizons *1*, 142–155.

Patel, O., Kjer-Nielsen, L., Le Nours, J.E.R.O.M., Eckle, S.B.G., Birkinshaw, R., Beddoe, T., Corbett, A.J., Liu, L., Miles, J.J., Meehan, B., et al. (2013). Recognition of vitamin B metabolites by mucosal-associated invariant T cells. Nat Commun *4*, 1–9.

Pauken, K.E., Sammons, M.A., Odorizzi, P.M., Manne, S., Godec, J., Khan, O., Drake, A.M., Chen, Z., Sen, D.R., Kurachi, M., et al. (2016). Epigenetic stability of exhausted T cells limits durability of reinvigoration by PD-1 blockade. Science *354*, 1160–1165.

Pearce, E.L., Mullen, A.C., Martins, G.A., Krawczyk, C.M., Hutchins, A.S., Zediak, V.P., Banica, M., DiCioccio, C.B., Gross, D.A., Mao, C.-A., et al. (2003). Control of effector CD8+ T cell function by the transcription factor Eomesodermin. Science *302*, 1041–1043.

Pellicci, D.G., Hammond, K.J.L., Uldrich, A.P., Baxter, A.G., Smyth, M.J., and Godfrey, D.I. (2002). A Natural Killer T (NKT) Cell Developmental Pathway Involving a Thymus-dependent NK1.1−CD4+ CD1d-dependent Precursor Stage. J Exp Med *195*, 835–844.

Perez-Muñoz, M.E., Arrieta, M.-C., Ramer-Tait, A.E., and Walter, J. (2017). A critical assessment of the "sterile womb" and “in utero colonization” hypotheses: implications for research on the pioneer infant microbiome. Microbiome *5*, 48.

Pique-Regi, R., Degner, J.F., Pai, A.A., Gaffney, D.J., Gilad, Y., and Pritchard, J.K. (2011). Accurate inference of transcription factor binding from DNA sequence and chromatin accessibility data. Genome Res. *21*, 447–455.

Porcelli, S., Yockey, C.E., Brenner, M.B., and Balk, S.P. (1993). Analysis of T cell antigen receptor (TCR) expression by human peripheral blood CD4-8- α/β T cells demonstrates preferential use of several Vβ genes and an invariant TCR α chain. J Exp Med *178*, 1–16.

Provine, N.M., Amini, A., Garner, L.C., Dold, C., Hutchings, C., FitzPatrick, M.E.B., Reyes, L.S., Chinnakannan, S., Oguti, B., Raymond, M., et al. (2019). Activation of MAIT cells plays a critical role in viral vector vaccine immunogenicity. bioRxiv *7*, 661397.

Provine, N.M., Binder, B., FitzPatrick, M.E.B., Schuch, A., Garner, L.C., Williamson, K.D., van Wilgenburg, B., Thimme, R., Klenerman, P., and Hofmann, M. (2018). Unique and Common Features of Innate-Like Human Vδ2+ γδT Cells and Mucosal-Associated Invariant T Cells. Front. Immunol. *9*, 756.

Quach, B., and Furey, T.S. (2016). DeFCoM: analysis and modeling of transcription factor binding sites using a motif-centric genomic footprinter. Bioinformatics btw740.

Quigley, M., Pereyra, F., Nilsson, B., Porichis, F., Fonseca, C., Eichbaum, Q., Julg, B., Jesneck, J.L., Brosnahan, K., Imam, S., et al. (2010). Transcriptional analysis of HIV-specific CD8+ T cells shows that PD-1 inhibits T cell function by upregulating BATF. *16*, 1147–1151.

Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841–842.

Raberger, J., Schebesta, A., Sakaguchi, S., Boucheron, N., Blomberg, K.E.M., Berglöf, A., Kolbe, T., Smith, C.I.E., Rülicke, T., and Ellmeier, W. (2008). The transcriptional regulator PLZF induces the development of CD44 high memory phenotype T cells. Pnas *105*, 17919–17924.

Rahimpour, A., Koay, H.F., Enders, A., Clanchy, R., Eckle, S.B.G., Meehan, B., Chen, Z., Whittle, B., Liu, L., Fairlie, D.P., et al. (2015). Identification of phenotypically and functionally heterogeneous mouse mucosal-associated invariant T cells using MR1 tetramers. J Exp Med *212*, 1095–1108.

Ramirez, R.N., El-Ali, N.C., Mager, M.A., Wyman, D., Conesa, A., and Mortazavi, A. (2017). Dynamic Gene Regulatory Networks of Human Myeloid Differentiation. Cell Systems *4*, 416–429.e3.

Ramírez, F., Ryan, D.P., Grüning, B., Bhardwaj, V., Kilpert, F., Richter, A.S., Heyne, S., Dündar, F., and Manke, T. (2016). deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Res. *44*, W160–W165.

Ranson, T., Vosshenrich, C.A.J., Corcuff, E., Richard, O., Laloux, V., Lehuen, A., and Di Santo, J.P. (2003). IL-15 availability conditions homeostasis of peripheral natural killer T cells. Pnas *100*, 2663–2668.

Reantragoon, R., Corbett, A.J., Sakala, I.G., Gherardin, N.A., Furness, J.B., Chen, Z., Eckle, S.B.G., Uldrich, A.P., Birkinshaw, R.W., Patel, O., et al. (2013). Antigen-loaded MR1 tetramers define T cell receptor heterogeneity in mucosal-associated invariant T cells. J Exp Med *210*, 2305–2320.

Reantragoon, R., Kjer-Nielsen, L., Patel, O., Chen, Z., Illing, P.T., Bhati, M., Kostenko, L., Bharadwaj, M., Meehan, B., Hansen, T.H., et al. (2012). Structural insight into MR1-mediated recognition of the mucosal associated invariant T cell receptor. J Exp Med *209*, 761–774.

Riegert, P., Wanner, V., and Bahram, S. (1998). Genomics, isoforms, expression, and phylogeny of the MHC class I-related MR1 gene. J Immunol *161*, 4066–4077.

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics *26*, 139–140.

Rosati, E., Dowds, C.M., Liaskou, E., Henriksen, E.K.K., Karlsen, T.H., and Franke, A. (2017). Overview of methodologies for T-cell receptor repertoire analysis. BMC Biotechnol. *17*, 61.

Rouxel, O., Da Silva, J., Beaudoin, L., Nel, I., Tard, C., Cagninacci, L., Kiaf, B., Oshima, M., Diedisheim, M., Salou, M., et al. (2017). Cytotoxic and regulatory roles of mucosal-associated invariant T cells in type 1 diabetes. Nat Immunol *18*, 1321–1331.

Sag, D., Krause, P., Hedrick, C.C., Kronenberg, M., and Wingender, G. (2014). IL-10-producing NKT10 cells are a distinct regulatory invariant NKT cell subset. The Journal of Clinical Investigation *124*, 3725–3740.

Salio, M., Gasser, O., Gonzalez-Lopez, C., Martens, A., Veerapen, N., Gileadi, U., Verter, J.G., Napolitani, G., Anderson, R., Painter, G., et al. (2017). Activation of Human Mucosal-Associated Invariant T Cells Induces CD40L-Dependent Maturation of Monocyte-Derived and Primary Dendritic Cells. J. Immunol. ji1700615.

Salou, M., Legoux, F., Gilet, J., Darbois, A., Halgouet, du, A., Alonso, R., Richer, W., Goubet, A.-G., Daviaud, C., Menger, L., et al. (2018). A common transcriptomic program acquired in the thymus defines tissue residency of MAIT and NKT subsets. J Exp Med jem.20181483.

Sattler, A., Dang-Heine, C., Reinke, P., and Babel, N. (2015). IL-15 dependent induction of IL-18 secretion as a feedback mechanism controlling human MAIT-cell effector functions. Eur J Immunol *45*, 2286–2298.

Sattler, A., Thiel, L.G., Ruhm, A.H., Souidi, N., Seifert, M., Herberth, G., and Kotsch, K. (2019). The TL1A-DR3 Axis Selectively Drives Effector Functions in Human MAIT Cells. J. Immunol. *203*, 2970–2978.

Savage, A.K., Constantinides, M.G., Han, J., Picard, D., Martin, E., Li, B., Lantz, O., and Bendelac, A. (2008). The transcription factor PLZF directs the effector program of the NKT cell lineage. Immunity *29*, 391–403.

Schoenfelder, S., and Fraser, P. (2019). Long-range enhancer-promoter contacts in gene expression control. Nat. Rev. Genet. *20*, 437–455.

Schraml, B.U., Hildner, K., Ise, W., Lee, W.-L., Smith, W.A.E., Ben Solomon, Sahota, G., Sim, J., Mukasa, R., Cemerski, S., et al. (2009). The AP-1 transcription factor Batf controls T H 17 differentiation. Nature *460*, 405–409.

Schwessinger, R., Suciu, M.C., McGowan, S.J., Telenius, J., Taylor, S., Higgs, D.R., and Hughes, J.R. (2017). Sasquatch: predicting the impact of regulatory SNPs on transcription factor binding from cell- and tissue-specific DNase footprints. Genome Res. *27*, 1730–1742.

Seach, N., Guerri, L., Le Bourhis, L., Mburu, Y., Cui, Y., Bessoles, S., Soudais, C., and Lantz, O. (2013). Double Positive Thymocytes Select Mucosal-Associated Invariant T Cells. J Immunol *191*, 6002–6009.

Seiler, M.P., Mathew, R., Liszewski, M.K., Spooner, C., Barr, K., Meng, F., Singh, H., and Bendelac, A. (2012). Elevated and sustained expression of the transcription factors Egr1 and Egr2 controls NKT lineage differentiation in response to TCR signaling. Nat Immunol *13*, 264–271.

Sen, D.R., Kaminski, J., Barnitz, R.A., Kurachi, M., Gerdemann, U., Yates, K.B., Tsao, H.-W., Godec, J., LaFleur, M.W., Brown, F.D., et al. (2016). The epigenetic landscape of T cell exhaustion. Science *354*, 1165–1169.

Serriari, N.E., Eoche, M., Lamotte, L., Lion, J., Fumery, M., Marcelo, P., Chatelain, D., Barre, A., Nguyen-Khac, E., Lantz, O., et al. (2014). Innate mucosal-associated invariant T (MAIT) cells are activated in inflammatory bowel diseases. Clin Exp Immunol *176*, 266–274.

Shaler, C.R., Choi, J., Rudak, P.T., Memarnejadian, A., Szabo, P.A., Tun-Abraham, M.E., Rossjohn, J., Corbett, A.J., McCluskey, J., McCormick, J.K., et al. (2017a). MAIT cells launch a rapid, robust and distinct hyperinflammatory response to bacterial superantigens and quickly acquire an anergic phenotype that impedes their cognate antimicrobial function: Defining a novel mechanism of superantigen-induced immunopathology and immunosuppression. PLOS Biol *15*, e2001930.

Shaler, C.R., Tun-Abraham, M.E., Skaro, A.I., Khazaie, K., Corbett, A.J., Mele, T., Hernandez-Alejandro, R., and Haeryfar, S.M.M. (2017b). Mucosa-associated invariant T cells infiltrate hepatic metastases in patients with colorectal carcinoma but are rendered dysfunctional within and adjacent to tumor microenvironment. Cancer Immunol. Immunother. *66*, 7–13.

Sharma, P.K., Wong, E.B., Napier, R.J., Bishai, W.R., Ndung'u, T., Kasprowicz, V.O., Lewinsohn, D.A., Lewinsohn, D.M., and Gold, M.C. (2015). High expression of CD26 accurately identifies human bacteria-reactive MR1-restricted MAIT cells. Immunology *145*, 443–453.

Shi, Y., Glynn, J.M., Guilbert, L.J., Cotter, T.G., Bissonnette, R.P., and Green, D.R. (1992). Role for c-myc in activation-induced apoptotic cell death in T cell hybridomas. Science *257*, 212–214.

Shin, D., Lee, W., Lee, J.H., and Bang, D. (2019). Multiplexed single-cell RNA-seq via transient barcoding for simultaneous expression profiling of various drug perturbations. Sci Adv *5*, eaav2249.

Slichter, C.K., McDavid, A., Miller, H.W., Finak, G., Seymour, B.J., McNevin, J.P., Diaz, G., Czartoski, J.L., McElrath, M.J., Gottardo, R., et al. (2016). Distinct activation thresholds of human conventional and innate-like memory T cells. JCI Insight *1*, e86292.

Smith, A.D., Foss, E.D., Zhang, I., Hastie, J.L., Giordano, N.P., Gasparyan, L., VinhNguyen, L.P., Schubert, A.M., Prasad, D., McMichael, H.L., et al. (2019). Microbiota of MR1 deficient mice confer resistance against Clostridium difficile infection. PLoS ONE *14*, e0223025.

Sobkowiak, M.J., Davanian, H., Heymann, R., Gibbs, A., Emgård, J., Dias, J., Aleman, S., Krüger-Weiner, C., Moll, M., Tjernlund, A., et al. (2018). Tissue-resident MAIT cell populations in human oral mucosa exhibit an activated profile and produce IL-17. Eur J Immunol.

Song, L., and Crawford, G.E. (2010). DNase-seq: a high-resolution technique for mapping active gene regulatory elements across the genome from mammalian cells. Cold Spring Harb Protoc *2010*, pdb.prot5384–pdb.prot5384.

Stegle, O., Teichmann, S.A., and Marioni, J.C. (2015). Computational and analytical challenges in single-cell transcriptomics. Nat. Rev. Genet. *16*, 133–145.

Stoeckius, M., Zheng, S., Houck-Loomis, B., Hao, S., Yeung, B.Z., Mauck, W.M., Smibert, P., and Satija, R. (2018). Cell Hashing with barcoded antibodies enables multiplexing and doublet detection for single cell genomics. Genome Biol *19*, 58.

Street, K., Risso, D., Fletcher, R.B., Das, D., Ngai, J., Yosef, N., Purdom, E., and Dudoit, S. (2018). Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. BMC Genomics *19*, 477.

Stritesky, G.L., Jameson, S.C., and Hogquist, K.A. (2012). Selection of Self-Reactive T Cells in the Thymus. Annu. Rev. Immunol. *30*, 95–114.

Stubbington, M.J.T., Lönnberg, T., Proserpio, V., Clare, S., Speak, A.O., Dougan, G., and Teichmann, S.A. (2016). T cell fate and clonality inference from single-cell transcriptomes. Nat Meth *13*, 329–332.

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., et al. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Pnas *102*, 15545–15550.

Sullivan, B.M., Juedes, A., Szabo, S.J., Herrath, von, M., and Glimcher, L.H. (2003). Antigen-driven effector CD8 T cell function regulated by T-bet. Pnas *100*, 15818–15823.

Sun, Z., Unutmaz, D., Zou, Y.R., Sunshine, M.J., Pierani, A., Brenner-Morton, S., Mebius, R.E., and Littman, D.R. (2000). Requirement for RORgamma in thymocyte survival and lymphoid organ development. Science *288*, 2369–2373.

Sundström, P., Ahlmanner, F., Akéus, P., Sundquist, M., Alsén, S., Yrlid, U., Börjesson, L., Sjöling, Å., Gustavsson, B., Wong, S.B.J., et al. (2015). Human Mucosa-Associated Invariant T Cells Accumulate in Colon Adenocarcinomas but Produce Reduced Amounts of IFN-γ. J. Immunol. *195*, 3472–3481.

Sung, M.-H., Baek, S., and Hager, G.L. (2016). Genome-wide footprinting: ready for prime time? Nat Meth *13*, 222–228.

Surh, C.D., and Sprent, J. (2008). Homeostasis of Naive and Memory T Cells. Immunity *29*, 848–862.

Tang, X.-Z., Jo, J., Tan, A.T., Sandalova, E., Chia, A., Tan, K.C., Lee, K.H., Gehring, A.J., De Libero, G., and Bertoletti, A. (2013). IL-7 licenses activation of human liver intrasinusoidal mucosal-associated invariant T cells. J. Immunol. *190*, 3142–3152.

Tastan, C., Karhan, E., Zhou, W., Fleming, E., Voigt, A.Y., Yao, X., Wang, L., Horne, M., Placek, L., Kozhaya, L., et al. (2018). Tuning of human MAIT cell activation by commensal bacteria species and MR1-dependent T-cell presentation. Mucosal Immunol *117*, 1250–1605.

Thomas, S.Y., Scanlon, S.T., Griewank, K.G., Constantinides, M.G., Savage, A.K., Barr, K.A., Meng, F., Luster, A.D., and Bendelac, A. (2011). PLZF induces an intravascular surveillance program mediated by long-lived LFA-1-ICAM-1 interactions. J Exp Med *208*, 1179–1188.

Thomsen, M.K., Bakiri, L., Hasenfuss, S.C., Hamacher, R., Martinez, L., and Wagner, E.F. (2013). JUNB/AP-1 controls IFN-γ during inflammatory liver disease. J. Clin. Invest. *123*, 5258–5268.

Tilloy, F., Treiner, E., Park, S.-H., Garcia, C., Lemonnier, F., la Salle, de, H., Bendelac, A., Bonneville, M., and Lantz, O. (1999). An Invariant T Cell Receptor α Chain Defines a Novel TAP-independent Major Histocompatibility Complex Class Ib–restricted α/β T Cell Subpopulation in Mammals. J Exp Med *189*, 1907–1921.

Todeschini, A.-L., Georges, A., and Veitia, R.A. (2014). Transcription factors: specific DNA binding and specific gene regulation. Trends Genet. *30*, 211–219.

Tominaga, K., Yamagiwa, S., Setsu, T., Kimura, N., Honda, H., Kamimura, H., Honda, Y., Takamura, M., Yokoyama, J., Suzuki, K., et al. (2017). Possible involvement of mucosal-associated invariant T cells in the progression of inflammatory bowel diseases. Biomedical Research *38*, 111–121.

Toubal, A., Nel, I., Lotersztajn, S., and Lehuen, A. (2019). Mucosal-associated invariant T cells and disease. Nat Rev Immunol *178*, 1–15.

Treiner, E., Duban, L., Bahram, S., Radosavljevic, M., Wanner, V., Tilloy, F., Affaticati, P., Gilfillan, S., and Lantz, O. (2003). Selection of evolutionarily conserved mucosal-associated invariant T cells by MR1. Nature *422*, 164–169.

Tsompana, M., and Buck, M.J. (2014). Chromatin accessibility: a window into the genome. Epigenetics Chromatin *7*, 33.

Tsukumo, S.-I., Unno, M., Muto, A., Takeuchi, A., Kometani, K., Kurosaki, T., Igarashi, K., and Saito, T. (2013). Bach2 maintains T cells in a naive state by suppressing effector memory-related genes. Pnas *110*, 10735–10740.

Turtle, C.J., Delrow, J., Joslyn, R.C., Swanson, H.M., Basom, R., Tabellini, L., Delaney, C., Heimfeld, S., Hansen, J.A., and Riddell, S.R. (2011). Innate signals overcome acquired TCR signaling pathway regulation and govern the fate of human CD161(hi) CD8α^{+} semi-invariant T cells. Blood *118*, 2752–2762.

Ussher, J.E., Bilton, M., Attwod, E., Shadwell, J., Richardson, R., de Lara, C., Mettke, E., Kurioka, A., Hansen, T.H., Klenerman, P., et al. (2014). CD161++ CD8+ T cells, including the MAIT cell subset, are specifically activated by IL-12+IL-18 in a TCR-independent manner. Eur J Immunol *44*, 195–203.

Vaillant, F., Blyth, K., Andrew, L., Neil, J.C., and Cameron, E.R. (2002). Enforced expression of Runx2 perturbs T cell development at a stage coincident with beta-selection. J Immunol *169*, 2866–2874.

van Wilgenburg, B., Loh, L., Chen, Z., Pediongco, T.J., Wang, H., Shi, M., Zhao, Z., Koutsakos, M., Nussing, S., Sant, S., et al. (2018). MAIT cells contribute to protection against lethal influenza infection in vivo. Nat Commun *9*, 4706.

van Wilgenburg, B., Scherwitzl, I., Hutchinson, E.C., Leng, T., Kurioka, A., Kulicke, C., de Lara, C., Cole, S., Vasanawathana, S., Limpitikul, W., et al. (2016). MAIT cells are activated during human viral infections. Nat Commun *7*, 11653.

Varelias, A., Bunting, M.D., Ormerod, K.L., Koyama, M., Olver, S.D., Straube, J., Kuns, R.D., Robb, R.J., Henden, A.S., Cooper, L., et al. (2018). Recipient mucosal-associated invariant T cells control GVHD within the colon. J. Clin. Invest. *128*, 1919–1936.

Varshney, A., Scott, L.J., Welch, R.P., Erdos, M.R., Chines, P.S., Narisu, N., Albanus, R.D., Orchard, P., Wolford, B.N., Kursawe, R., et al. (2017). Genetic regulatory signatures underlying islet gene expression and type 2 diabetes. Pnas *114*, 2301–2306.

Voillet, V., Buggert, M., Slichter, C.K., Berkson, J.D., Mair, F., Addison, M.M., Dori, Y., Nadolski, G., Itkin, M.G., Gottardo, R., et al. (2018). Human MAIT cells exit peripheral tissues and recirculate via lymph in steady state conditions. JCI Insight *3*, e98487.

Wainer-Katsir, K., and Linial, M. (2019). BIRD: Identifying Cell Doublets via Biallelic Expression from Single cells. bioRxiv *10*, 709451.

Walch, M., Dotiwala, F., Mulik, S., Thiery, J., Kirchhausen, T., Clayberger, C., Krensky, A.M., Martinvalet, D., and Lieberman, J. (2014). Cytotoxic Cells Kill Intracellular Bacteria through Granulysin-Mediated Delivery of Granzymes. Cell *157*, 1309–1323.

Walker, L.J., Tharmalingam, H., and Klenerman, P. (2014). The rise and fall of MAIT cells with age. Scand J Immunol *80*, 462–463.

Walker, L.J., Kang, Y.-H., Smith, M.O., Tharmalingham, H., Ramamurthy, N., Fleming, V.M., Sahgal, N., Leslie, A., Oo, Y., Geremia, A., et al. (2012). Human MAIT and CD8αα cells develop from a pool of type-17 precommitted CD8+ T cells. Blood *119*, 422–433.

Wang, H., D'Souza, C., Lim, X.Y., Kostenko, L., Pediongco, T.J., Eckle, S.B.G., Meehan, B.S., Shi, M., Wang, N., Li, S., et al. (2018). MAIT cells protect against pulmonary Legionella longbeachae infection. Nat Commun *9*, 3350.

Wang, H., Kjer-Nielsen, L., Shi, M., D'Souza, C., Pediongco, T.J., Cao, H., Kostenko, L., Lim, X.Y., Eckle, S.B.G., Meehan, B.S., et al. (2019). IL-23 costimulates antigen-specific MAIT cell activation and enables vaccination against bacterial infection. Sci Immunol *4*, eaaw0402.

Watford, W.T., Hissong, B.D., Bream, J.H., Kanno, Y., Muul, L., and O'Shea, J.J. (2004). Signaling by IL-12 and IL-23 and the immunoregulatory roles of STAT4. Immunol. Rev. *202*, 139–156.

Webster, K.E., Kim, H.-O., Kyparissoudis, K., Corpuz, T.M., Pinget, G.V., Uldrich, A.P., Brink, R., Belz, G.T., Cho, J.-H., Godfrey, D.I., et al. (2014). IL-17-producing NKT cells depend exclusively on IL-7 for homeostasis and survival. Mucosal Immunol *7*, 1058–1067.

Willing, A., Jäger, J., Reinhardt, S., Kursawe, N., and Friese, M.A. (2018). Production of IL-17 by MAIT Cells Is Increased in Multiple Sclerosis and Is Associated with IL-7 Receptor Expression. J. Immunol. ji1701213.

Willinger, T., Freeman, T., Herbert, M., Hasegawa, H., McMichael, A.J., and Callan, M.F.C. (2006). Human naive CD8 T cells down-regulate expression of the WNT pathway transcription factors lymphoid enhancer binding factor 1 and transcription factor 7 (T cell factor-1) following antigen encounter in vitro and in vivo. J Immunol *176*, 1439–1446.

Wilson, R.P., Ives, M.L., Rao, G., Lau, A., Payne, K., Kobayashi, M., Arkwright, P.D., Peake, J., Wong, M., Adelstein, S., et al. (2015). STAT3 is a critical cell-intrinsic regulator of human unconventional T cell numbers and function. J Exp Med *212*, 855–864.

Winter, S.J., Schumacher, H.K., Imelmann, E., Grewers, Z., Osthues, T., and Krueger, A. (2019). MicroRNA miR-181a/b-1 controls MAIT cell development. Immunol Cell Biol *97*, 190–202.

Wolock, S.L., Lopez, R., and Klein, A.M. (2019). Scrublet: Computational Identification of Cell Doublets in Single-Cell Transcriptomic Data. Cell Systems *8*, 281–291.e289.

Won, E.J., Ju, J.K., Cho, Y.-N., Jin, H.-M., Park, K.-J., Kim, T.-J., Kwon, Y.-S., Kee, H.J., Kim, J.-C., Kee, S.-J., et al. (2016). Clinical relevance of circulating mucosal-associated invariant T cell levels and their anti-cancer activity in patients with mucosal-associated cancer. Oncotarget *7*, 76274–76290.

Wong, W.F., Kohu, K., Chiba, T., Sato, T., and Satake, M. (2011). Interplay of transcription factors in T-cell differentiation and function: the role of Runx. Immunology *132*, 157–164.

Yan, J., Allen, S., McDonald, E., Das, I., Mak, J.Y.W., Liu, L., Fairlie, D.P., Meehan, B.S., Chen, Z., Corbett, A.J., et al. (2019). MAIT cells promote tumor initiation, growth and metastases via tumor MR1. Cancer Discov CD–19–0569.

Yu, G., and He, Q.-Y. (2016). ReactomePA: an R/Bioconductor package for reactome pathway analysis and visualization. Mol Biosyst *12*, 477–479.

Zabijak, L., Attencourt, C., Guignant, C., Chatelain, D., Marcelo, P., Marolleau, J.-P., and Treiner, E. (2015). Increased tumor infiltration by mucosal-associated invariant T cells correlates with poor survival in colorectal cancer patients. Cancer Immunol. Immunother. *64*, 1601–1608.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., et al. (2008). Model-based Analysis of ChIP-Seq (MACS). Genome Biol *9*, R137.

Zuber, J., Rosen, S., Shonts, B., Sprangers, B., Savage, T.M., Richman, S., Yang, S., Lau, S.P., DeWolf, S., Farber, D., et al. (2015). Macrochimerism in Intestinal Transplantation: Association With Lower Rejection Rates and Multivisceral Transplants, Without GVHD. American Journal of Transplantation *15*, 2691–2703.

Zumwalde, N.A., Haag, J.D., Gould, M.N., and Gumperz, J.E. (2018). Mucosal associated invariant T cells from human breast ducts mediate a Th17-skewed response to bacterially exposed breast carcinoma cells. Breast Cancer Res. *20*, 111.

List of publications

FitzPatrick, M.E.B.*, Provine, N.M.*, **Garner, L.C.***, Powell, K., Amini, A., Irwin, S., Ferry, H., Ambrose, T., Friend, P., Vrakas, G., et al. (2019). Human intestinal tissue-resident memory CD8+ T cells comprise transcriptionally and functionally distinct subsets. (Under review)
***co-first author**

Slevin, S.M., Tan, M., Lahiff, C., **Garner, L.C.**, Wang, L.M., Ferry, H., Greenaway, B., Lynch, K., Geremia, A., Hughes, S., Leavens, K., Krull, D., Marks, D.J.B., Nevin, K., Page, K., Srinivasan, N., Tarzi, R., Klenerman, P., Travis, S., Arancibia-Cárcamo, C.V., Keshav, S. Lymphocyte activation gene (LAG)-3 is associated with mucosal inflammation and disease activity in ulcerative colitis. (Under review)

McCuaig, S., Barras, D., Mann, E., Friedrich, M., Bullers, S., Janney, A., **Garner, L.C.**, Domingo, E., Koelzer, V.H., Oxford IBD Cohort Investigators, Delorenzi M., Tejpar, S., Maughan, T., West, N.R., Powrie, F. The interleukin 22 pathway interacts with mutant KRAS to promote poor prognosis in colon cancer. (Under review)

Hagel, J.P., **Garner, L.C.**, Bilton, M., Mehta, H., Leng, T., Hackstein, C.-P., Phalora, P., Amini, A., Akther, H.D., Provine, N.M., et al. (2020). Human MAIT Cell Activation In Vitro. Methods Mol. Biol. *2098*, 97–124.

Provine, N.M., Amini, A., **Garner, L.C.**, Dold, C., Hutchings, C., FitzPatrick, M.E.B., Reyes, L.S., Chinnakannan, S., Oguti, B., Raymond, M., et al. (2019). Activation of MAIT cells plays a critical role in viral vector vaccine immunogenicity. bioRxiv *7*, 661397.

Garner, L.C.*, Klenerman, P., and Provine, N.M.* (2018). Insights Into Mucosal-Associated Invariant T Cell Biology From Studies of Invariant Natural Killer T Cells. Front. Immunol. *9*, 1478.
*** corresponding author**

Provine, N.M., Binder, B., FitzPatrick, M.E.B., Schuch, A., **Garner, L.C.**, Williamson, K.D., van Wilgenburg, B., Thimme, R., Klenerman, P., and Hofmann, M. (2018). Unique and Common Features of Innate-Like Human Vδ2+ γδT Cells and Mucosal-Associated Invariant T Cells. Front. Immunol. *9*, 756.

Llibre, A., **Garner, L.**, Partridge, A., Freeman, G.J., Klenerman, P., and Willberg, C.B. (2016). Expression of lectin-like transcript-1 in human tissues. F1000Research *5*, 2929.