

Universität Hamburg

DER FORSCHUNG | DER LEHRE | DER BILDUNG

# Single-cell RNA sequencing data analysis reveals human T cell adaptation

Dissertation zur Erlangung des Doktorgrades
an der Fakultät für Mathematik, Informatik und Naturwissenschaften
Fachbereich Informatik
der Universität Hamburg
vorgelegt von Tianran Zhang
Hamburg, Mai 2024

**Vorsitzender der Prüfungskommission**

Professor. Dr. Matthias Rarey

**Gutachter**

Professor Dr. Jan Baumbach

Professor Dr. Stefan Bonn

**Datum der mündlichen Prüfung**

09.10.2024

# Abstract

Single-cell RNA sequencing (scRNA-seq) has transformed our understanding of human immunology, revealing the transcriptional activity of individual cells within the tissue environments. Recent cross-tissue studies using this technology have contributed to the development of Human Cell Atlas, producing deep biological insights. In particular, the study of T cell adaptation across tissues using scRNA-seq stands at the forefront of developing innovative therapeutic approaches. However, the specific identity, function, and distribution of T cell populations within different tissue environments remain to be fully understood.

This study investigates the complex adaptation of human T cells across various tissues, including visceral adipose tissue (VAT), liver, small intestine, and peripheral blood by using scRNA-seq and TCR sequencing (TCR-seq). To analyze this high-dimensional data, numerous computational methods and analysis tools were applied. Clustering analysis revealed the cellular heterogeneity of immune cell populations for each tissue site. Differential gene expression analysis on the clonal-related cells between tissues uncovered both common and tissue-specific adaptation signatures within CD4+ and CD8+ T cell lineages, such as CD69 for both liver and VAT and BATF specifically for VAT. The study also elucidated the heterogeneity and residence of naive T cells in tissues. Together with the trajectory analysis, a model of tissue adaptation was introduced, showing that T cells adjust their gene expression profiles from naive states to adapt to specific tissues. Moreover, through cell-cell communication analysis, interactions of the IL6 signaling pathway between different T cell subsets and innate immune cells were found more enriched in VAT, as a potential explanation of *BATF*'s higher expression.

This research provided valuable perspectives on tissue adaptation of T cells, as well as offering a comprehensive analysis framework for the interpretation of cross-tissue data with scRNA-seq and TCR-seq. These findings emphasized the great importance of computational methods in dealing with large biomedical datasets. The use of appropriate algorithms and analyses is essential for revealing the true biological meanings within the data. Future research could integrate additional cell types and data modalities to uncover more fine-grained cell types and state annotations, fostering a comprehensive understanding of the T cell adaptation in tissues.

# Deutsche Zusammenfassung

Die Einzelzell-RNA-Sequenzierung (scRNA-seq) hat unser Verständnis der Humanimmunologie verändert, da sie die Transkriptionsaktivität einzelner Zellen in der Gewebeumgebung aufdeckt. Jüngste gewebeübergreifende Studien mit dieser Technologie haben zur Entwicklung des Human Cell Atlas beigetragen und tiefe biologische Einblicke ermöglicht. Insbesondere die Untersuchung der gewebeübergreifenden Anpassung von T-Zellen mit Hilfe von scRNA-seq steht an der Spitze der Entwicklung innovativer therapeutischer Ansätze. Die spezifische Identität, Funktion und Verteilung von T-Zellpopulationen in verschiedenen Gewebeumgebungen muss jedoch noch vollständig geklärt werden.

Diese Studie untersucht die komplexe Anpassung menschlicher T-Zellen in verschiedenen Geweben, darunter viszerales Fettgewebe (VAT), Leber, Dünndarm und peripheres Blut, unter Verwendung von scRNA-seq und TCR-Sequenzierung (TCR-seq). Zur Analyse dieser hochdimensionalen Daten wurden zahlreiche Berechnungsmethoden und Analysetools eingesetzt. Die Clustering-Analyse zeigte die zelluläre Heterogenität der Immunzellpopulationen für jede Gewebestelle. Die differenzielle Genexpressionsanalyse der klonal-verwandten Zellen zwischen den Geweben deckte sowohl gemeinsame als auch gewebespezifische Anpassungssignaturen innerhalb der CD4+ und CD8+ T-Zelllinien auf, wie z. B. CD69 sowohl für Leber als auch für VAT und BATF speziell für VAT. Die Studie gab auch Aufschluss über die Heterogenität und den Verbleib naiver T-Zellen in den Geweben. Zusammen mit der Trajektorienanalyse wurde ein Modell der Gewebeanpassung eingeführt, das zeigt, dass T-Zellen ihre Genexpressionsprofile vom naiven Zustand aus anpassen, um sich an bestimmte Gewebe anzupassen. Darüber hinaus wurde durch eine Analyse der Zell-Zell-Kommunikation festgestellt, dass die Interaktionen des IL6-Signalwegs zwischen verschiedenen T-Zell-Untergruppen und angeborenen Immunzellen in der Mehrwertsteuer stärker angereichert sind, was eine mögliche Erklärung für die höhere Expression von BATF darstellt.

Diese Forschung lieferte wertvolle Perspektiven zur Gewebeanpassung von T-Zellen und bot einen umfassenden Analyserahmen für die Interpretation gewebeübergreifender Daten mit scRNA-seq und TCR-seq. Diese Ergebnisse unterstreichen die große Bedeutung von Berechnungsmethoden im Umgang mit großen biomedizinischen Datensätzen. Die Verwendung geeigneter Algorithmen und Analysen ist unerlässlich,

um die wahren biologischen Bedeutungen der Daten zu erkennen. Zukünftige Forschungsarbeiten könnten weitere Zelltypen und Datenmodalitäten integrieren, um feinere Zelltypen und Zustandsannotationen aufzudecken und so ein umfassendes Verständnis der T-Zell-Anpassung in Geweben zu fördern.

# Contents


**1 Introduction** **3**

1.1 Immune basics . . . 3
1.1.1 Innate immunity . . . 4
1.1.2 Adaptive immunity . . . 6
1.2 T cells . . . 8
1.2.1 T cell receptors . . . 8
1.2.2 T cell heterogeneity . . . 9
1.2.3 T cell residence . . . 13
1.3 Tissue immunity . . . 14
1.3.1 Visceral adipose tissue . . . 14
1.3.2 Liver . . . 19
1.3.3 Small intestine . . . 22
1.4 Single-cell RNA sequencing analysis . . . 23
1.4.1 Single-cell RNA sequencing technique . . . 23
1.4.2 Single-cell RNA sequencing analysis workflow . . . 24
1.5 Single-cell TCR sequencing analysis . . . 37
1.6 Research question . . . 40

**2 Material and methods** **43**

2.1 Processing and sequencing of human tissue samples . . . 43
2.2 Single-cell sequencing data processing . . . 45
2.3 Single-cell RNA sequencing downstream analysis . . . 48
2.4 Single-cell TCR sequencing analysis . . . 53

**3 Human tissue T cell adaptation** **55**

3.1 Cross-tissue analysis of paired scRNA-seq and TCR-seq data . . . 55
3.1.1 T cell subsets across tissues . . . 57
3.1.2 T cell repertoires across tissues . . . 59

3.2 Tissue adaptation of CD4+ T cells . . . 62
3.2.1 Tissue adaptation signatures of CD4+ T cells . . . 62
3.2.2 Tissue adaptation model of CD4+ T cells . . . 72
3.2.3 IL6 signaling pathway in VAT CD4+ T cells . . . 82
3.3 Tissue adaptation of CD8+ T cells . . . 85
3.3.1 Tissue adaptation signatures of CD8+ T cells . . . 85
3.3.2 CD8 Trm signatures of the small intestine . . . 89
3.4 Tissue adaptation of myeloid cells and NK cells . . . 95
3.4.1 Tissue adaptation signatures of NK cells . . . 95
3.4.2 Tissue adaptation signatures of myeloid cells . . . 98

**4 Discussion** **105**
4.1 Tissue T cell adaptation . . . 105
4.1.1 Tissue T cell adaptation signatures revealed by DGE analysis . 105
4.1.2 Naive T cells adapt to tissues . . . 107
4.2 Challenges of scRNA-seq data analysis . . . 107
4.3 Single-cell multimodal data integration . . . 111
4.4 Summary and outlook . . . 114

**Publications** **118**

**Bibliography** **119**

**Acronyms** **141**

**List of Figures** **145**

**List of Tables** **149**

**Acknowledgments** **151**

# 1 Introduction

In the introduction section of this thesis, firstly the foundational principles of the immune system were introduced, containing the key players of both innate immunity and adaptive immunity. This was followed by an extensive overview of human T cells, including their distinct characteristics and the roles of their various subsets. Subsequently, in the section of tissue immunity, the unique immune landscapes of three human tissues—visceral adipose tissue, liver, and small intestine were highlighted, with a more detailed exploration of the first two tissues. Next, the analysis workflow for single-cell RNA sequencing and TCR sequencing data was introduced step by step, covering data preprocessing and downstream analysis. Finally, the main research question of this study was raised.

## 1.1 Immune basics

The immune system is a complex biological system in which various organs, tissues, cells, and molecules work together to protect the host against infections, diseases, and foreign substances [1]. The immune system is essential for human survival [2]. Without a working immune system, minor infections can become severe and even fatal [3]. Beyond defending against pathogens, the immune system plays an important role in maintaining body homeostasis and facilitating the repair of tissues after injury [4]. It helps to eliminate dead cells and supports the renewal of tissues throughout the body. A well-functioning immune system can recognize foreign cells from the body's own cells. This ability prevents it from attacking the body's own tissues, which can lead to autoimmune diseases.

Immunity refers to the body's protection from disease and, more specifically, infectious disease [1]. It can be classified into two major types: innate immunity and adaptive immunity. Innate immunity offers rapid, general protection against pathogens

without the need of prior exposure. It identifies a specific and limited group of molecular configurations, in total around 1000 byproducts from microbes and damaged cells. In contrast, adaptive immunity potentially can recognize millions of distinct molecular structures of microbes, along with self-originated and environmental antigens with high specificity. Innate and adaptive immunity are integrated together in host defense, with numerous cells and molecules functioning in coordination. Innate immunity triggers and shapes adaptive immunity, while adaptive responses in turn enhance innate defenses against pathogens [2].

### 1.1.1 Innate immunity

Innate immunity, also known as natural or native immunity, provides the first line of protection against pathogens and is present from birth [2]. Diverse elements are involved in innate immunity. They are: (1) chemical and physical barriers like skin membrane and gastric acid, (2) innate immune cells including phagocytic cells, dendritic cells, and natural killer cells, and (3) other soluble factors such as members of the complement system. There are two primary responses in innate immunity: inflammation and antiviral defense. Inflammation mobilizes and activates leukocytes and plasma proteins at infection sites to eliminate pathogens, and it responds to cell damage and abnormal substance accumulations in tissues. Antiviral defense inhibits virus replication within cells and increases susceptibility to killing by lymphocytes, removing viral infection sources. Different cell types from innate immunity play different roles in defense against microorganisms.

**Phagocytes** including neutrophils, monocytes, and macrophages, mainly function by engulfing and destroying pathogens like bacteria, viruses, and infected or dead cells. This process is known as phagocytosis [5]. Neutrophils constitute the largest cell population within white blood cells. Generated in the bone marrow, neutrophils circulate in the blood and can migrate to areas of infection rapidly. In addition to neutrophils, eosinophils, and basophils also belong to the granulocyte family, which are both implicated in defending against parasites.

**Monocytes** circulate in the blood as well. Different subsets of monocytes have been uncovered in human, which can be distinguished by their surface markers [6]. Classical monocytes highly expressing CD14 but not CD16 (CD14++CD16–), comprising about 80-90% of the total monocyte cell group in the blood, are involved in the pro-

duction of inflammatory mediators in addition to phagocytosis. Non-classical monocytes, instead, highly expressing CD16 (CD14+CD16++), assist in tissue repair after injury, anti-inflammatory responses, and maintenance of endothelial health and integrity. Besides, an intermediate subset of monocytes (CD14++CD16+) found in humans, participates in antigen presentation and the generation of pro-inflammatory cytokines.

**Macrophages** originate from monocytes and mature upon entering tissues. Human macrophages are defined according to the expression of HLA-DR and CD11b, along with the differential expression of receptors CD64 and CD16 [7]. In infected tissues, macrophages are usually the first cells to react to microbes by ingesting and killing them as well as dead cells. Activated macrophages release cytokines that recruit neutrophils and other white blood cells from the blood into the infected region. Based on their location, function, and signals received from the environment, macrophages can broadly be described as either "classically activated" (M1) or "alternatively activated" (M2). M1 macrophages, activated by interferon-gamma, participate in pro-inflammatory responses, while M2 macrophages, induced by interleukin-4 (IL-4) and interleukin-13 (IL-13), are correlated with anti-inflammatory responses, tissue repair, and remodeling.

**Dendritic cells** (DCs) are known as the most important antigen-presenting cells (APCs), that are responsible for capturing, displaying antigens to lymphocytes, and therefore stimulating their proliferation and differentiation. Dendritic cells serve as a critical connection between the innate and adaptive immune systems by interacting with T cells. They are heterogeneous and can reside in tissues [8]. Classical or conventional dendritic cells (cDCs), expressing CD11c, are the most abundant in skin, mucosa, and organ parenchyma. Plasmacytoid Dendritic Cells (pDCs) with the marker of CD123, respond to viral infections through producing substantial quantities of type I interferons.

**Natural killer cells** (NK cells) are the most described innate lymphoid cells and their primary function is to kill infected cells [1]. NK cells originate from bone marrow and perform effector functions without expressing diverse receptors found on T or B cells. These cells can activate macrophages by producing IFN$\gamma$, with a similar mechanism as cytotoxic T cells. By the marker gene CD56 and the absence of CD3 expression, NK cells are defined. Furthermore, two subsets of human NK cells can be divided by their surface marker CD56– CD56(bright) and CD56(dim) NK cells. CD56dim NK

cells, expressing highly CD16, constitute most of NK cells in peripheral blood and are highly cytotoxic. Conversely, CD56(bright) NK cells have low cytotoxicity and are capable of producing cytokines, particularly IFN$\gamma$ upon activation [9].

**Residence of innate immune cells**
Many of the studies related to tissue residence of innate immune cells are tissue-resident macrophages. Macrophages can acquire different functions in order to adapt to the specific tissue environment [7]. Tissue-resident macrophages display a set of markers such as the scavenger receptors CD163 and MARCO, mannose receptor CD206, CD169, and CD68. The expression of these markers can differ across various tissues, such as the liver, lung, spleen, gut and heart [10]. DCs are also resident in tissues, but due to their tendency to migrate into secondary lymphoid organs, it is difficult to determine their residency status. Tissue-resident NK cells have been discovered in different tissues, for example, the lung and liver [11]. Similarly, tissue retention markers such as CD69, CD103, and others are generally used for their identification.

### 1.1.2 Adaptive immunity

Adaptive immunity, also known as specific or acquired immunity, delivers a highly specific defense mechanism that evolves through an individual's life by exposure to pathogens from diseases or vaccination [2]. Adaptive immunity is usually activated after the initial response of the innate immune system. It is characterized by its capacity to recognize and memorize specific pathogens, resulting in an enhanced response upon subsequent encounters. There are two categories of adaptive immune responses: humoral immunity and cellular immunity, which are mediated by different lymphocytes, the unique cells from adaptive immunity. Humoral immunity targets pathogens in body fluids using antibodies produced by B lymphocytes (B cells), while cellular immunity, also called cell-mediated immunity, is mediated by T lymphocytes (T cells). T and B lymphocytes are both specialized cells that identify and react to foreign antigens, though they differ in their specific functions and roles.

**B cells** are the only cells that are able to produce antibodies. Antibodies, also called immunoglobulins, are specialized proteins that bind to foreign substances (antigens). There are five different immunoglobulins, with each type equipped with a unique structure and function, allowing the immune system to effectively react to a broad

spectrum of pathogens and to provide various types of immunity [12]. B cells arise and mature in the bone marrow. After being activated by antigens, they proliferate and differentiate into plasma cells or memory B cells. Plasma cells typically reside in the bone marrow or mucosal tissues, where they produce abundant high-affinity antibodies. Memory B cells, that do not actively secrete antibodies, remain in circulation and mucosal tissues for the long term, ready to initiate rapid antibody production upon re-exposure to their specific antigen. Instead of antibody secretion, B cells can also work as antigen-presenting cells, facilitating the activation of helper T cells.

**T cells**, the mediators of cellular immunity, originate in the bone marrow and migrate to and mature in the thymus. T cells exhibit specific recognition for antigens. They identify peptides derived from foreign proteins when these are presented alongside the body's own major histocompatibility complex (MHC) molecules on the cell surfaces. Therefore, T cells only respond to antigens associated with cell surfaces, not those that are soluble [1]. T lymphocytes consist of functionally distinct populations. The two main subsets of T cells are CD4+ helper T cells and CD8+ cytolytic T cells (CTLs), both known as $\alpha\beta$ T cells because of their expression of $\alpha\beta$ T cell receptors (TCR). Besides, CD4+ regulatory T cells, which also possess $\alpha\beta$ TCRs, maintain immune tolerance by suppressing immune responses. Other T cells are innate-like T cells including NKT (natural killer T) cells, mucosal-associated invariant T (MAIT) cells, and $\gamma\delta$ T cells. The T cell groups are very heterogeneous and they can remain in tissues, playing a very important role in local immune surveillance and response [13]. The detailed characteristics of each T cell subset and their tissue residence were introduced in the next section.

Cells from both innate and adaptive immunity secret cytokines, e.g. macrophages, B cells, and T cells as well as non-immune cells endothelial cells and fibroblasts in different tissues. Cytokines represent a broad type of small proteins with varied structures and functions, that are crucial in cell signaling in the immune system. They play pivotal roles in the body's immune response, including the activation of immune cells (e.g. interleukin-1 (IL-1) and tumor necrosis factor (TNF)), cellular communication (e.g. interferons (IFNs)), cell growth and differentiation (e.g. IL-7), and regulation of inflammatory response (e.g. interleukin-1 (IL-10) and transforming growth factor-beta (TGF-$\beta$)) [14].

The immune system operates in a coordinated manner to counter various threats from the environment. It's crucial for addressing various health challenges, such as

autoimmunity, where the immune response targets the body itself, and transplantation, where managing immune reactions is key to preventing organ rejection. Insights into immunity also drive vaccine development and cancer immunotherapy, leveraging the body's natural defenses against diseases. Moreover, understanding resident immune cells in tissues is vital for uncovering their roles in local immunity. This knowledge not only advances medical science but also informs the development of targeted therapies, emphasizing the importance of in-depth immune system research for improving health outcomes [10]. Recent advancements in high-throughput technologies have enabled simultaneous measurement of multiple cell types, cytokines, and other immune function biomarkers not only in blood but also in some solid tissues [15]. These innovations provide the opportunity to explore variations in the human immune system on a global scale, considering the interplay between specific cell populations and proteins.

## 1.2 T cells

T lymphocytes derive from lymphoid progenitor cells in the bone marrow and migrate to the thymus for maturation, where they undergo rigorous positive and negative selection. Then these mature T cells move to the periphery, entering the pool of naive T cells that constantly circulate between secondary lymphoid organs and the bloodstream via the lymphatic system. When encountering specific antigens in the peripheral, naive T cells undergo proliferation and then develop into various effector or memory T cell types, capable of migrating to tissue sites. T cells coordinate multiple facets of adaptive immunity over a lifetime, including responses to pathogens, allergens, and tumors. T cells are also recognized as significant contributors to lots of inflammatory and autoimmune diseases [2].

### 1.2.1 T cell receptors

T cells recognize foreign antigens and protect against many different pathogens by expressing highly polymorphic surface receptors, called T cell receptors (TCR). These receptors mostly consist of a single pair of chains, with over 90% of human T cells displaying $\alpha$ and $\beta$ TCR chains ($\alpha\beta$ T cells), and a smaller fraction expressing $\gamma$ and $\delta$ TCR chains ($\gamma\delta$ T cells). The diversity of TCRs, particularly among $\alpha\beta$ TCRs, is

remarkable, with estimations suggesting up to 100 million distinct TCR sequences in humans [16].

Each TCR chain is structured with a variable (V) domain, a constant (C) domain, a hydrophobic transmembrane region, and a short cytoplasmic tail [1]. The variable region of TCRs is responsible for antigen recognition and is encoded through the assembly of several gene segments, including variable (V) and joining (J) segments for the $\alpha$ and $\gamma$ TCR chains, with the addition of diversity (D) segments for the $\beta$ and $\delta$ TCR chains. Through the process of VDJ recombination, a single allele from each gene segment is randomly selected and joined to construct the variable region. This recombination, along with the fusion to a constant gene segment, produces a functional TCR chain transcript. The addition or deletion of random nucleotides at the joining positions of gene segments introduces strong combinatorial diversity (based on the selection of gene segments for recombination) and junctional diversity (the largest contribution, resulting from the variability in nucleotide addition or deletion), creating an extensive and diverse TCR repertoire with the ability of recognizing a vast range of antigens. Further diversity is generated by the pairing of $\alpha$ and $\beta$ or $\gamma$ and $\delta$ TCR chains to assemble a complete T cell receptor.

The variability among different TCRs is primarily localized within the short amino acid sequences from the V regions of TCR chains, known as hypervariable or complementarity-determining regions (CDRs). Each TCR chain contains three CDRs. CDR1 and 2 are encoded by the V gene segments, playing a crucial role in TCR combination with the MHC complex. The CDR3 region, deriving from the highly variable junctional region between V and J (or D and J) gene segments, is thus highly variable. As a result, the likelihood of two T cells sharing common CDR3 sequence is extremely low, except that they are cells from the same clonal expansion. This specificity makes the CDR3 sequence together with TCR gene combinations, a critical marker for identifying T cell clonotypes.

### 1.2.2 T cell heterogeneity

T cells are heterogeneous and a variety of T cell subtypes have been identified, each with distinct functions, surface markers, and roles within the immune system. Within $\alpha\beta$ T cells, there are two primary lineages: CD4+ T cells and CD8+ T cells. CD4+ T cells, including T helper cells and regulatory T cells, recognize antigens presented

by class II MHC molecules. Conversely, CD8+ T cells are specialized in recognizing antigens presented by class I MHC molecules. Each of these T cell categories can be further classified into different subsets according to their activation states and distinct functions [1].

**Naive T cells** maintain their naive state functionally until they encounter their specific antigen, prompting them to differentiate into either effector or memory T cells. In humans, naive T cells are defined by the expression of CD45RA and the lack of memory-associated marker CD45RO. Together with other naive markers including the lymphoid homing receptors CC-chemokine receptor 7 (CCR7) and the L-selectin CD62 ligand (CD62L), these cells can be identified. In the periphery, naive T cells consistently circulate between secondary lymphoid organs and the bloodstream through the lymphatic system in search of their specific antigens. Besides, the vast array of T cell receptors on naive T cells enables the recognition of a broad spectrum of antigens. These features are crucial for early detection of pathogens. Naive T cells play a critical role in maintaining immune surveillance and diversity, which is essential for the effective functioning of the entire immune system.

Naive T cells are traditionally viewed as a quiescent and homogeneous group of antigen-inexperienced cells, and they have been primarily studied through peripheral blood samples, due to the challenge of obtaining tissue specimens. Recent advancements, however, have unveiled that naive T cells exhibit a much greater degree of compartmentalization and diversity than previously understood [17]. Their variability regarding phenotype, location, differentiation potential, cytokine production, and gene expression profiles highlights a complex heterogeneity of the naive T cell population. Such new insights can bring significant clinical implications for immune surveillance, therapeutic interventions, and deepen our understanding of naive T cell predifferentiation.

**Effector T cells** (Teff) are differentiated from naive T cells upon antigen recognition mainly in secondary lymphoid tissues. These cells perform specialized effector functions and can migrate to any site of infection or inflammation. Effector T cells of the CD4+ lineage include regulatory T cells and multiple types of helper T cells, which release cytokines and display surface molecules that stimulate other immune cells. Cytotoxic T cells (CTLs), the effector cells within the CD8+ lineage, kill infected and tumor cells while also releasing cytokines that activate macrophages and promote inflammation. Unlike their CD4+ counterparts, cytotoxic T cells are functionally ho-

mogeneous.

The activation mechanisms of naive CD4 and CD8 T cells are similar, which involve critical signals from the T cell receptor and co-receptor (CD4 or CD8), together with the co-stimulatory receptor CD28, that interacts with B7 molecules on dendritic cells. These combined signals then trigger significant alterations in the T cells' gene expression, resulting in the upregulation of activation markers like CD69, CD25, CD40 ligand, and CTLA-4. It also stimulates the production of cytokine IL-2, which is crucial for the proliferation and differentiation, as well as clonal expansion of T cells. Concurrently, T cells adjust their migratory properties: they downregulate the expression of molecules such as CD62L and CCR7, enabling them to exit lymphoid tissues, and upregulate the expression of molecules like integrins LFA1 and VLA4, ligands for E- and P-selectins, and multiple chemokine receptors, enhancing their ability to migrate to sites of infection or tissue damage [18].

**CD4 helper T cells** are diverse and there are three major subsets—Th1, Th2, and Th17 cells, distinguished by their signature cytokines and effector functions. Th1 cells produce IFN$\gamma$ and are crucial in responding to intracellular infections and their differentiation is regulated by the T-bet transcription factor. Th2 cells, under the regulation of the GATA3 transcription factor, secrete IL-4, IL-5, and IL-13, helping innate immune cells including mast cells, eosinophils, and basophils, as well as B cells, protect against parasite infections. Th17 cells, with their differentiation regulated by the transcription factor ROR$\gamma$T, produce IL-17 and IL-22, and they are involved in defending against fungal and bacterial infections and various autoimmune diseases. In addition, another subset of effector CD4 T cells is the follicular helper T (Tfh) cells, which are differentiated by the BCL6 transcription factor. Tfh cells primarily interact with B cells within the lymphoid follicles to promote their development into antibody-secreting plasma cells and memory B cells [19].

**Regulatory T cells** (Tregs) are a distinct subset of CD4 T cells, responsible for suppressing immune responses and preserving self-tolerance. These cells are identified by the high expression of the high-affinity IL2 receptor alpha chain (CD25) on their surface, along with the unique transcription factor FOXP3. Tregs arising from the thymus (tTregs) are selected based on their recognition of self-antigens. Peripheral regulatory T cells (pTregs) originate from naive CD4 T cells in the peripheral lymphoid organs following exposure to specific types of antigens. The adaptability of pTregs is vital in maintaining immune homeostasis, such as preventing autoimmu-

nity and tissue damage. The inhibitory cytokines TGF-β and IL-10 also contribute to the generation of regulatory T cells as well as the maintenance of their suppressive functions, which is now acknowledged as a fundamental mechanism for protecting the integrity of the body's tissues and organs [20].

**Memory T cells** are distinguished by their long lifespan and their capacity to quickly differentiate into effector T cells, providing rapid and enhanced protection against pathogens the immune system has previously encountered [18]. Memory T cells can develop from effector T cells along a linear pathway, or arise from activated naive T cells before or during the initial stages of differentiation into effector cells, through a divergent development model. These cells are heterogeneous, with subsets differing particularly related to their locations and migration patterns. Based on the expression of lymph node homing receptor CCR7 and the naive/memory marker CD45RA, CD4+ and CD8+ memory T cells were initially classified as: central memory T (Tcm) cells (CD45RA- CCR7+), which recirculate through the blood and routinely migrate through lymph nodes (LNs), and effector memory T (Tem) cells (CD45RA- CCR7-), which travel between tissues and the circulation [21]. Subsequently, a sub-population of resident memory T (Trm) cells was characterized which do not recirculate and remain in tissues [13]. These cells garnered increasing attention in immunological research due to their critical role in providing localized and long-lasting immunity in tissue sites.

**Innate-like T cells**, which include γδ T cells, NKT cells, and MAIT cells, are situated at the intersection of innate and adaptive immunity, bridging the two systems with their unique capabilities and functions. These cells have semi-invariant or invariant T cell receptors that recognize only a limited set of antigens. γδ T cells with TCR of γ and δ chains, consist of less than 5% of the total population of T cells. The V regions of their γ and δ chains have limited variability, and the junctional diversity is also little. γδ T cells are involved in a series of biological activities such as producing cytokines and killing infected cells. NKT cells are a small subset of T cells, which express markers that are found on NK cells like CD56. In humans, NKT cells typically express a TCR of the alpha rearrangement Vα24-Jα18, together with one of three beta chains, resulting in limited diversity. Their TCRs uniquely identify lipid antigens presented by the class I MHC-like, CD1 molecules. NKT cells are also able to produce cytokines (e.g. IL4 and IFNγ) quickly. In contrast, MAIT cells are an abundant T cell subset found in mucosal tissues, particularly the lungs. MAIT cells play a role in

microbial defense by using a limited TCR repertoire that is conserved across species. These T cell receptors recognize antigens presented by the MHC class I-like molecule, MR1. MAIT cells are marked by the expression of CD161 and TRAV1-2. In relation to the co-receptors CD4 and CD8, different subsets can be defined with the majority of these cells being CD8-positive [1].

### 1.2.3 T cell residence

Evidence for T cell residence has been constantly reported since the identification of Trm cells. Trm cells are the most abundant T cell lineage. In order to survive and maintain their functions in different tissue environments, these T cells make adaptations by acquiring tissue-specific characteristics, resulting in changes in their gene expression profiles. In different tissues, Trm cells express varying levels of resident markers and tissue-specific signatures [22]. The majority of human CD4+ and CD8+ Trm cells express CD69, which is a canonical Trm marker, but it is not a definitive marker to distinguish between recirculating cells and Trm cells since CD69- Trm cells have also been reported [23]. CD8+ Trm cells express more CD103 and CD101 compared to CD4 Trm cells, although a small group of CD103+CD4+ Trm cells were found in human skin [24], lung [25], and gut [26]. In addition to these relative universal tissue residence signatures, Trm cells in each tissue host unique signatures that can support their survival and functions in the tissue environment [27].

Besides tissue-resident memory T cells, naive T cells have been discovered within tissues, with the cell quantity reaching up to half of the resident T cells in the lung and more than half in the colon. The expression of Trm marker CD69 was found to be upregulated on a subset of human naive CD4+ T cells from lymph nodes and mucosal tissues, further suggesting their tissue retention [28]. Tissue-specific signatures of chemokine receptors and cytokines were also revealed in naive CD4+ and CD8+ T cells [29]. Additionally, analysis of TCR clonal distribution in naive T cells has shown tissue-specific clonal expansions with little overlap between clones from different lymphoid tissues. Collectively, these findings suggest the association of tissue residency with the sustained maintenance of naive T cells via local homeostasis [30], leading to the question of whether naive T cells adapt to the local tissue environment [17].

Tissue-resident T cells patrol tissues, providing the first-line defense by rapidly re-

sponding to pathogens at initial sites of encounter. Besides, these T cells contribute to maintaining tissue integrity and homeostasis. They regulate local inflammation, assist in tissue repair, and, by producing various cytokines and growth factors, modulate the activity of other immune cells. However, instead of the protective roles of resident T cells, they have been implicated in the development of autoimmunity and chronic inflammatory diseases. Consequently, novel therapies targeting Trm cells could offer effective treatments for such autoimmune diseases. Moreover, strategies for depleting Trm cells could be beneficial in transplantation settings, minimizing the risk of graft rejection and improving transplant outcomes [31]. Therefore, precisely identifying the characteristics and roles of resident T cells across various tissue locations is crucial for unlocking the wide range of potential therapeutic applications.

Much of the current knowledge about resident T cells as well as other resident immune cells, is concluded from mice studies, where peripheral tissues and organs are more accessible. Moreover, cellular properties such as migration are more readily manipulated and traced in mice compared to human samples. With the increasing availability of new technologies such as mass cytometry, single-cell RNA sequencing, and spatial imaging, we can now investigate human T cell residency with greater precision. Application of these approaches on the single-cell level is revealing new transcriptional and functional characteristics for human tissue-resident T cells, providing new perspectives into their ontogeny, maintenance, and functional roles within specific tissue environments [22, 32, 33].

## 1.3 Tissue immunity

### 1.3.1 Visceral adipose tissue

Adipose tissue (AT), known as body fat, is a loose connective tissue composed of various cell types including predominantly adipocytes along with stromal cells, endothelial cells, and a wide range of immune cells. Adipose tissue is crucial in managing the overall energy balance and glucose homeostasis of the body. As an energy storage organ, adipose tissue functions as an energy reservoir, storing lipids and managing their mobilization and distribution throughout the body. It also operates as an endocrine organ, secreting a variety of bioactive substances, including adipokines. These substances engage in communication with other organs, influencing and regu-

lating numerous metabolic processes [34, 35].

There are two major different adipose tissues in mammals according to their primary function: white adipose tissue (WAT) and brown adipose tissue (BAT). The brown adipose tissue generates body heat by the increased levels of uncoupling protein 1 (UCP1) in its mitochondria, providing ways to decrease body weight and restore glucose and lipid homeostasis [36]. In contrast, the white AT accounts for most of the adipose tissue in adults, and functions as the main energy storage tissue in the body, where the white adipocytes are mainly present. These white adipocytes are characterized by a high capacity of storing energy in the form of fatty acids and orchestrating systemic metabolic activities through the secretion of hormones and adipokines [37]. Two subtypes of WAT can be further classified: subcutaneous adipose tissue (SCAT), found beneath the dermal layer of skin, and visceral adipose tissue (VAT), situated within the abdominal cavity surrounding the inner organs. Visceral fat, including mesenteric, gonadal, epicardial, retroperitoneal, omental and peri-renal fat has been linked closely to the increased risk of metabolic diseases [38, 39], while subcutaneous fat is considered to pose a lower health risk compared to visceral fat and evidence suggests it might even have protective benefits [40].

VAT actively participates in body weight regulation, insulin sensitivity, inflammation control, and lipid metabolism through the excretion of cytokines and bioactive mediators. The high number of immune cells contained in VAT interact with adipocytes, maintaining the organ's metabolism and overall physiology. In VAT, most types of immune cells can be found, including macrophages, eosinophils, T cells, B cells, and innate lymphoid cells [41]. Many of them migrate to the tissue in the early stages of life and stay permanently resident in the adipose tissue. These tissue-resident immune cells are pivotal in maintaining adipose tissue homeostasis and function by modulating the inflammatory state of adipose tissue and contributing to the control of thermogenesis [42].

**T cells in AT**

T cells are the second most abundant group of immune cells in adipose tissue [43]. The existence of different subtypes of T cells has been uncovered including CD4+ T helper cells, CD8+ cytotoxic T cells, gamma delta T cells, and natural killer T cells in WAT [44]. Adipose T cells play indispensable roles in tissue immunometabolism, as the interaction of immunity and metabolism. The distinct landscapes of T cell subtypes have been reviewed in adipose tissue environment with a specific association

to obesity [41, 45, 46].

The presence of CD4+ T cells was first detected in human visceral adipose tissue, and it was found to correlate with body mass index [47]. Upon activation by antigen stimulation, naive CD4+ T cells can differentiate into different helper T cells (Th1, Th2, and Th17 cells) and regulatory T cells. AT contained fewer naive CD4+ T cells compared to second lymphoid tissues but a large amount of effector memory T cells [48]. Among all CD4+ T cells, regulatory T cells with suppressive and tissue-regulatory functions are the central population in maintaining the integrity of adipose tissue. They work to restrain inflammation, preventing the development of insulin resistance in AT. Tregs in VAT exhibit a different phenotype and transcriptional profile compared to those in lymphoid tissues. They have a high expression of the adipocyte transcription factor PPAR$\gamma$, transcription factor BATF, GATA3, and interferon regulatory factor IRF4, supporting their development and maintenance in VAT [49]. Additionally, IL33 signaling was required for survival and expansion by adipose Tregs [50]. However, less is understood about the role of other subsets of T cells in adipose tissue. Th1 cells, the first helper T cell subset identified in adipose tissue, produce proinflammatory cytokines, including TNF$\alpha$ and IFN$\gamma$, which were thought to be deleterious in obesity [51]. The other type of Th cells, Th2 cells, produce STAT6-responsive cytokines IL4 and IL13, shown to be beneficial in adipose tissue [51]. Th17 cells were implicated in proinflammatory processes in adipose tissue. Several studies have corroborated the increase in Th17 cells within the visceral adipose tissue of individuals with adiposity [52–54].

CD8+ T cells, similar to CD4+ T cells, infiltrated in adipose tissue, have been increasingly recognized for their involvement in the regulation of metabolic processes and inflammation. The increase of CD8+ T cells in VAT in obesity contributed to macrophage recruitment through the expression of IFN$\gamma$ [55–57], promoting an inflammatory response implicated in metabolic dysfunctions. Vijay et al identified a subpopulation of CD8+ T cells that express metallothionein with potential obesity indication [58]. However, a recently published study using scRNA-seq data provided another view of the functions of CD8+ T cells in adipose tissue [59]. They identified PD1+TIGIT+ CD8 T cells in healthy fat which expressed not only effector/tissue-resident markers (CD69, GZMB, CCL3, BATF) but also the tissue-regenerative markers (AREG, TNF, and IFNG), suggesting a tissue regenerative or remodeling program. Therefore, careful investigation is needed to decipher the role of CD8+ T cells in AT.

Besides the conventional alpha-beta T cells, gamma-delta T cells were also found in lean adipose tissue with similar abundance as CD8 T cells [60]. Primly, these cells were claimed to promote inflammation and insulin resistance in mice [61]. Other studies showed gamma-delta T cells expressing IL-17A regulate thermogenesis which could be linked to the IL33 signalling [62]. Finally, natural killer T cells were shown to be enriched in VAT [63]. Conclusions about the function of NKT cells in AT are controversial and both pro- and anti-inflammatory effects have been reported in mice and humans [64].

One of the essential factors to preserve the differentiation and continuous presence of T cells is T cell receptor signaling, which is necessary throughout every stage of T cell development. Compared to splenic T cells, TCR in VAT is less diversified and the restriction was more pronounced in obesity [56]. VAT Tregs harbor a unique TCR repertoire compared to those in lymphoid tissues [50], showing specific antigen recognition by antigen-presenting cells. Similarly, CD8+ T cells in adipose tissue displayed distinct TCR$\beta$ V and J gene usage and lower repertoire diversity compared to blood [65]. The restricted TCR repertoire supports the development and accumulation of T cells in adipose tissue.

**Other immune cells in AT**

Numerous innate immune cells also exist in adipose tissue. Myeloid cells are the dominant population among adipose tissue-resident immune cells, which contain macrophages, mast cells, neutrophils, eosinophils, and dendritic cells. Among these myeloid cells, macrophages, as the most abundant immune cells in VAT [43], are considered to be major mediators in driving adipose tissue inflammation and contributing to tissue homeostasis [66]. Different subpopulations of macrophages have been identified within VAT and their functions range from lipid scavenging, phagocytosis, and cytokine secretion to thermogenesis regulation, as reviewed in [67]. They can be both beneficial and harmful: M1-like macrophages are proinflammatory, increased in obese VAT [68] while M2-like macrophages are anti-inflammatory, promoting the differentiation of beige adipocytes [69]. Using single-cell RNA sequencing, researchers revealed more heterogeneity of adipose macrophages and found rarely overlap with the classical M1/M2 paradigm [70, 71].

Dendritic Cells in AT can be conventional DCs and plasmacytoid DCs. It has been observed that DCs are involved in adipogenesis and insulin sensitivity [72]. The heterogeneity of DCs in adipose tissue was presented in a single-cell study [73]. So far,

less is understood of the role of DC in AT and many questions remain to be answered. Other myeloid cells are less abundant and less studied in adipose tissue: Eosinophils, as the main cell source of IL-4 in the adipose tissue [74] support the activation of macrophages and promote adipose thermogenesis; Neutrophils contribute to adipose tissue inflammation [75]; Mast cells accumulate in VAT, exacerbating obesity and the metabolic syndrome [76].

Despite of T cells and myeloid cells, other immune cells including NK cells cells, type 2 innate lymphoid cells (ILCs), and B cells were all found in adipose tissue. Natural killer cells are found more in VAT than other adipose dots and they are implicated in the involvement of adipose tissue inflammation [77]. The high-dimensional flow cytometry performed on NK cells from different fat spots and blood has revealed the diversity of NK cells in adipose tissue and discovered the protein signatures of VAT-resident CD56bright cells characterized by the co-expression of CD63, CD69, and NKG2D [78]. Type 2 innate lymphoid cells were discovered in VAT, contributing to tissue homeostasis [79]. The number of B cells in VAT is comparable to that of T cells. B cells, which produce antibodies and cytokines, produce different IgG antibodies in obesity that can be either harmful or beneficial [80].

To date, most of the studies in the adipose immune system are related to dysfunctions like obesity or aging and their goal is to discover how different immune cells change and are influenced compared to lean subjects. Generally, a variety of immune cells from both innate and adaptive immune cells are recruited in obesity such as M1 macrophages, NK cells, Th1 cells, CD8+ T cells, and B cells. These cells infiltrated in AT engage in the inflammatory cascade that leads to insulin resistance and metabolic dysfunction [42]. Among them, the myeloid cells particularly macrophages have gained the largest attention due to their abundant presence and pivotal role in the inflammatory processes within adipose tissue. In terms of T cells, further studies are required to uncover their characteristics, functions, and underlying mechanisms in adipose tissue. Moreover, few studies in AT focus on the adaptation of immune cells especially T cells by comparing the profiles of these adipose-resident immune cells to other organs.

### 1.3.2 Liver

The liver, as the body's largest solid organ, stands at the forefront of several crucial biological functions, including the metabolism of nutrients, clearance of toxins, and protein production [81]. It's also a central immunological organ due to its anatomical organization. Positioned downstream of the gut where venous blood drains, the liver, not only processes nutrients supplied from the blood but also actively screens and responds to potential gut-draining pathogens [82]. With its specialized architecture, the liver's tissue homeostasis is maintained by dynamic cell type interactions, additionally supporting hepatic regeneration.

The major cell types composed of the liver are parenchymal cells or hepatocytes (HC), which constitute 60% - 70% of the total liver cell population. Other non-parenchymal cells include hepatic stellate cells (HSCs), cholangiocytes, liver sinusoidal endothelial cells (LSECs), and immune cells [82]. A wide array of immune cell types can be found in the liver such as macrophages and DCs from the innate immune system as well as different T cells and B cells from the adaptive immune system [82]. Among them, some cell types are either transient visitors or preferentially concentrated, while others are exclusively resident such as the Kupffer cells (KCs), known as resident liver macrophages. Those immune cells are crucial in maintaining the balance between immunotolerance and immunity within the liver in response to environmental signals.

**Intrahepatic T cells**

T cells are critical in the hepatic immune system. They participate in resolving hepatotropic infections, maintaining immunity as well as the initiation and resolution of liver inflammation [83]. Many of the studies about intrahepatic T cells are from CD8+ T cells. The unique anatomy of the liver has made it possible for T cells to be directly accessible to antigen-presenting cells like hepatocytes. Studies have supported the hepatocellular priming of naive CD8+ T cells in the liver, which triggers local cells' initial proliferation and local activation [84, 85]. CD8+ Trm cells were identified first in murine liver [23] and human CD8 Trm cells were characterized by a population of $Tbet^{low}Eomes^{low}Blimp1^{high}Hobit^{low}$ T cells expressing liver-homing/retention markers (CD69,CD103,CXCR6 and CXCR3) [86]. The C-type lectin CD69 was expressed by most of the liver CD8 Trm cells but studies have shown that using only CD69 as a marker is insufficient and its expression is not required for CD8+ Trm in liver [87]. Similarly, CD103 is not always expressed in liver CD8+ Trm cells [88]. Transcriptional differences between species have been found in these cells: murine CD8 Trm

cells highly express transcriptional factor Hobit and Blimp1 while human CD8 Trm cells lack the expression of Hobit. Chemokine receptors CXCR6 and CXCR3 are both expressed highly in liver CD8+ T cells [89]. CXCR6 interacts with its ligand CXCL16 expressed by hepatocytes and LSECs, supporting the development and maintenance of memory CD8+ T cells in the liver [90]. In order to maintain a homeostatic state of tolerance, CD8+ Trm cells express high levels of immune checkpoint molecule PD-1 [91] and IL2 [86], which represents the key liver adaptations.

Much less has been studied about the phenotype and functions of liver-resident CD4+ T cells. Wiggins et al revealed two distinct populations of CD69-expressing intrahepatic CD4+ T cells in the human liver — $CD69^{high}$ CD4 Trm cells and a novel population $CD69^{Int}$ CD4 T cells [92]. The $CD69^{high}$ CD4 Trm cells are liver-resident expressing CD49a and CXCR6, which were also found in CD8 Trm cells. The $CD69^{Int}$ CD4 T cells contained both resident and recirculating T cells, characterized by high expression of CXCR3, CXCR1 and CX3CR1, which were required for hepatocyte homing. Compared to CD8+ Trm cells, CD103 is not expressed by most of the liver CD4+ Trm cells. The presence of naive-like CD4+ T cells in human livers have been demonstrated [93]. Interestingly, a population of tissue-resident naive-like CD4+ T cells was identified by a single-cell atlas study of intrahepatic T cells in human liver [94]. These cells showed a naive phenotype and expressed homing/tissue-resident gene markers such as CD69 and CXCR4, which were suggested to have a Th17 polarization state and associated with inflammatory liver diseases. Besides, Yu Zhao et. al reported the persistence of CD4+ T cells with the state of Th17 polarization in the development of fibrosis [95]. The role of different subsets of CD4+ T cells in liver inflammation and fibrosis has been actively discussed [96]. Normally, Th1 and Th17 cells are proinflammatory and profibrotic, while Th2 and Tregs, depending on the disease's setting and stage, can be either pro- or anti-fibrotic.

Except for conventional T cells, the liver also contains invariant T cells including MAIT cells, gamma-delta T cells and NKT cells. Based on their functional variance, they can be classified into two categories: type 1 (IFN$\gamma$ secreting) and Type 2 (IL17 secreting) [97]. The tissue signatures of these invariant T cells in the liver were similar to CD8 Trm cells, such as the transcriptional factor Hobit and the resident markers CD69 and CXCR6 [83]. Their involvement in various protective or pathological processes in liver has been implied, including defense against bacteria and reaction to injury.

In summary, a variety of studies on both conventional T cells and invariant T cells in liver uncovered several key liver-resident signatures and how they play their roles in the liver microenvironment. While to further deepen our insight into the liver T cell immunology, comprehensive analyses with novel techniques of various intrahepatic T cell subsets are still needed, especially the CD4+ T cells, which were much less studied.

**Other immune cells in the Liver**

Apart from T cells, the liver houses resident cells of the innate immune system, such as Kupffer cells, natural killer cells, dendritic cells, and innate immune cells. The Kupffer cells are liver-resident macrophages. As the largest population of tissue macrophages (80-90%), KCs account for around 35% of the non-parenchymal cells in the liver [98]. KCs are capable of detecting, binding, and engulfing pathogens and other associated molecules or cells from millions of red blood cells, platelets, and immune cells. They can activate T cells playing as antigen-presenting cells in the liver. Under the normal state, KCs exhibit a tolerogenic phenotype and can poorly stimulate the adaptive immune response. However, when stimulated by inflammatory molecules or cytokines, they can transform into potent activators of T cells [99]. The heterogeneity of liver macrophages has been revealed by scRNA-seq, which identified two distinct macrophage populations, one is inflammatory, and the other shows tolerogenic [100].

The liver also contains the body's largest population of NK cells. NK cells target and clear cells with the absence of self, aiding in cellular activation and the production of cytokines. Liver-resident NK cells can be distinguished by the surface marker CD49a according to a flow-cytometry study [101]. Additionally, multiple NK-ILC1 subsets within the liver were identified by the cross-organ scRNA-seq study [102], and common gene signatures of these cells across organs were uncovered. Myeloid DCs and plasmacytoid DCs can be both found in the liver. Similar to Kupffer cells, DCs serve as key regulator of the hepatic immune system, displaying a dual role that can be either protective or inflammatory [103].

The non-immune cells from the liver also play important roles in innate and adaptive immunity. LSECs are highly efficient APCs with both MHC class I and MHC class II. They can activate or induce tolerance of both CD4+ and CD8+ T cells in the liver. Hepatocytes express a wide range of immune receptors and they are able to detect pathogens and present antigens to naive T cells directly [82]. However, it remains un-

clear how hepatocytes and LSECs functionally contribute to the liver immune system in interacting with different types of T cells and other immune cell types.

### 1.3.3 Small intestine

The human small intestine (SI) is a critical player of the digestive system, playing a key role in the digestion and absorption of nutrients. There are three sections composed of human SI—duodenum, jejunum, and ileum. The duodenum serves as the initial segment, receiving digestive enzymes and bile to start the breakdown of food, while the jejunum and ileum further down are principally involved in nutrient absorption. It secretes various hormones that regulate digestive processes, such as secretin and cholecystokinin, which regulate digestive enzyme secretion and bile release. The small intestine also contributes to immune defense, with the GALT (gut-associated lymphoid tissue) contained in the intestine wall helping protect against pathogens.

The intestine stands as the largest compartment of the immune system, constantly encountering antigens and immune-regulating substances derived from the diet, as well as from the commensal microbiota. It serves as an entry point for numerous pathogens of clinical significance, influencing the progression of diseases in other parts of the body. This highlights the intestine's critical function not only in local defense mechanisms but also in the broader regulation of immune health across the entire organism. The mucosa, where most immune activity occurs, is made up of the epithelium, the lamina propria (LP) beneath it, and the muscularis mucosa, which is a slender muscle layer under the lamina propria. The lamina propria and epithelium function as the primary effector sites within the intestinal immune system, and they form distinct immunological compartments. The lamina propria is rich in a variety of immune cells, including B cells, T cells, dendritic cells, macrophages, eosinophils, and mast cells. In contrast, the epithelium is predominantly populated by T cells. Together, these layers house an extensive array of immune cells, making the intestinal mucosa the site with the highest concentration of macrophages, plasma cells, and T cells in the entire body [104].

The composition of T cell types varies across different compartments of the small intestine. The majority of human intraepithelial lymphocytes (IELs) are CD8+ T cells, with most of them are resident memory T cells showing tissue resident signatures. In

contrast, the lamina propria contains a large amount of functionally different CD4+ T cell subsets and plasma cells and a relatively smaller group of CD8+ T cells. MAIT cells and NKT cells were observed in the human jejunum, where they account for a minor fraction of IELs and lamina propria T cells [104].

The small intestine (SI) hosts a diverse array of innate immune cells. For instance, macrophages are the most enriched leukocytes in the healthy intestinal lamina propria. DCs in the SI present antigens from the gut lumen to T cells, regulating immune responses, such as tolerance to dietary antigens and commensal bacteria. NK cells contribute to the defense against infected or transformed cells in SI.

SI tissue-specific signatures have been reported and among different intestinal sites, unique tissue signatures including chemokine receptors, and transcriptional factors have been uncovered, which favors the recruitment of different types of immune cells. For example, CCL25 is expressed in the epithelium of the small intestine but not in the colon, which has been demonstrated to specifically promote the homing of T cells and plasmablasts to the small intestine. Among CD8+ Trm cells, most of them express high levels of both CD69 and CD103 in the SI. Besides, the transcription factors Eomes, Blimp1 and Hobit, etc, as well as the surface receptor CD160, are highly expressed in the SI intraepithelial Trm cells [105]. While there is an increasing interest in exploring the dynamics of immune cells within the small intestine, the comprehensive understanding remains challenging due to the complexity of its immune landscape and the specificity of its tissue structure.

## 1.4 Single-cell RNA sequencing analysis

### 1.4.1 Single-cell RNA sequencing technique

RNA sequencing (RNA-seq) is a deep sequencing technology for transcriptome profiling [106]. It allows scientists to obtain gene expression of a mixture of cells in samples, usually referred to as bulk RNA-seq. With this technology, scientists are able to identify changes in gene expressions under different conditions: healthy/diseased, wild-type/mutant, or control/treated samples. Besides, it can also be used for the detection of gene isoforms, gene fusions, single nucleotide variants, and many other interesting properties. However, bulk RNA-seq can only reveal the average expression levels for each gene across a population of cells, and therefore the heterogeneity

across individual cells remains unknown.

To uncover this question, single-cell RNA sequencing (scRNA-seq) technology was firstly developed in 2009 [107]. It enables the gene expression estimation of each gene for each cell, strongly promoting studies related to the detection of cell identity/state, function, and response. Since 2014, the technology has become more and more popular [108]. Compared to bulk RNA-seq, scRNA-seq is generally more expensive, while with the development of new protocols, the sequencing costs are getting lower and the sequencing accuracy is getting higher. Datasets involving scRNA-seq range from hundreds to millions of cells per study, and increase in size every year [109].

The main procedures of scRNA-seq technologies generally include single-cell isolation and capture, cell lysis, reverse transcription, cDNA amplification, and library preparation [110]. Based on the differences in cell isolation, these technologies can be classified into three types: (1) Microfluidic device-based such as Drop-seq and 10x Genomics Chromium, (2) well plate-based such as SMART-seq2, MARS-seq, and (3) Fluidigm C1 microfluidic chip-based. Different approaches vary in the enrichment of transcripts, the number of cells sequenced, the sequencing costs, etc. Besides, an alternative method of scRNA-seq, single-nucleus RNA sequencing (snRNA-seq), only captures the mRNAs in the nucleus of cells. snRNA-seq is quite useful on frozen samples or other samples that cannot be easily dissociated into single-cell suspensions. In this paper, the term scRNA-seq is used broadly to refer to either scRNA-seq or snRNA-seq. With the application of scRNA-seq, researchers are able to dissolve the transcriptome of cells from different tissue samples. For example, the creation of the human cell atlas, provides a detailed and high-resolution reference map of all human cells, serving as an essential resource for improving our understanding and developing treatments for diseases.

### 1.4.2 Single-cell RNA sequencing analysis workflow

Numerous analysis tools have been created for scRNA-seq data, aiming to expand the use of this technology in life and clinical sciences. By far, more than 1700 bioinformatics tools have been tracked and grouped into over 30 categories from the scRNA-tools database [111]. Therefore, conducting appropriate analysis on single-cell transcriptome data has become a major need, due to the increasing number of software tools, as well as the difference in protocols and the size of datasets. Several analysis tuto-

rials came up with detailed step-by-step guidance for the scRNA-seq analysis [112, 113]. Although most of the analysis steps are shared between them (Figure 1), new views of data analysis are still under development.

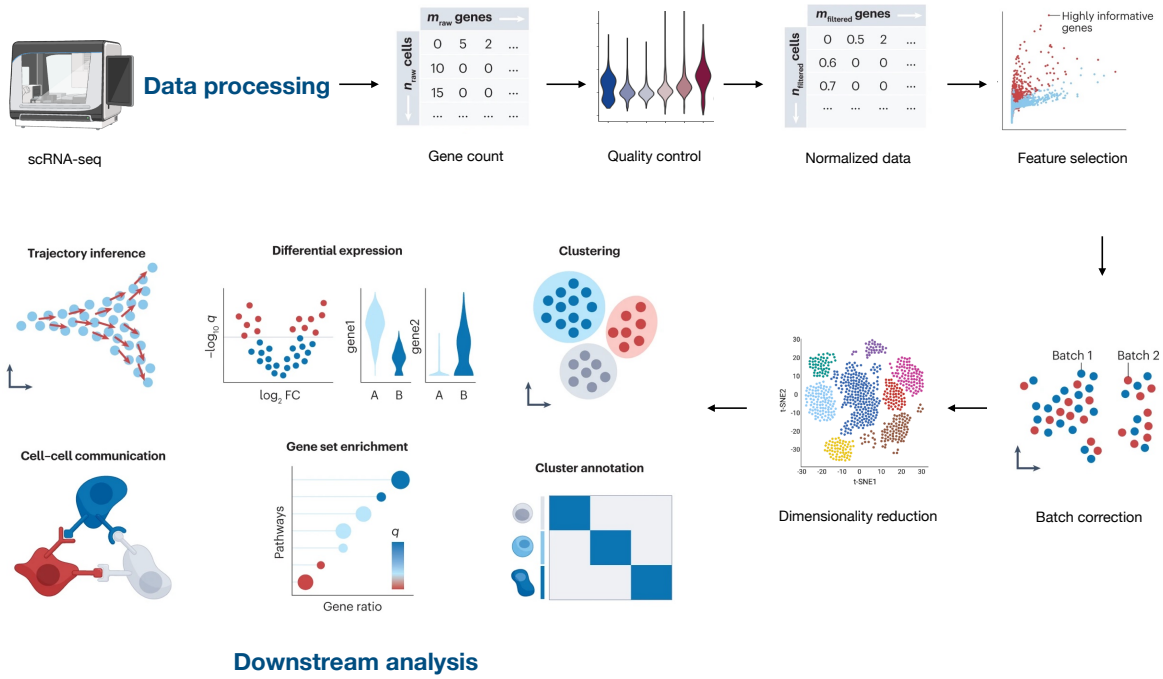



**Figure 1: Overview of scRNA-seq analysis steps.** Image modified from [113].

**Analysis frameworks**

Specialized analysis frameworks are required due to the size and complexity of the scRNA-seq data. The two most popular options are: R-based Seurat [114] and Python-based Scanpy [115]. Both Seurat and Scanpy are comprehensive toolkits for the quality control, analysis, and exploration of single-cell RNA-seq data, with well-written documentation. While Seurat is R-based, when dealing with very large datasets, it may encounter scalability challenges. Scanpy is a Python-based analysis framework, which is part of the scverse ecosystem [116]. It offers significant advantages in terms of scalability, extendability, and interoperability with the broader Python data science and machine learning ecosystem. Besides, there are other R-based Bioconductor packages, for example Scran [117], that store the data into the SingleCellExperiment object [118], providing various basic and other alternative advanced analyses for scRNA-seq data.

**Raw data processing and quality control**

Alignment or mapping is a process of aligning the sequenced reads to a reference genome, which is a fundamental step in single-cell raw data processing. The raw sequence data obtained usually refer to the FASTQ files. The sequencing quality of the raw FASTQ files can be checked by software like FastQC [119]. The output contains a QC report for each input FASTQ file, where the quality score, base content, and other information are included. Then to process the FASTQ files, we aim to convert them to the gene count matrix, which can then be applied for downstream analysis. Several tools can be used to achieve this such as Cell Ranger (10x Genomics), zUMIs [120], and STARsolo [121], with each suited to specific protocols.

Various errors, including empty droplets, doublets, and sequence errors like PCR amplification mistakes, can significantly affect the number of cell barcodes obtained after the mapping process. Cells with a low count depth, few detected genes, and high fractions of mitochondrial counts are often empty droplets or cells with broken membranes whose cytoplasmic mRNA have leaked out, and as a result, only mRNAs located in the mitochondria are still preserved. Conversely, cells with a high count depth and a large number of detected genes are potentially doublets. Several elegant tools have been widely applied to detect doublets. The benchmark study by Xi et. al [122] compared several computational doublet-detection methods for scRNA-seq data and they revealed diverse performances and distinct advantages for different methods. Within them, DoubletFinder [123] demonstrated the highest accuracy, how-

ever, it did not stand out for computational efficiency and stability. Scrublet [124] was outstanding in removing heterotypic (doublets formed from different cell types) and homotypic (doublets formed from the same cell type) doublets.

Selecting suitable thresholds or tools for the QC of scRNA-seq data is crucial for successful downstream analysis. For instance, keeping many low-quality cells or droplets in the count data can complicate cell clustering and annotation, whereas over filtering might result in the loss of rare cell populations. In real cases, the QC process often requires iterative reassessment and adjustment in order to determine the optimal steps. It is usually advantageous to begin with permissive QC thresholds and investigate the effects of them, then refine the process with more rigorous QC. Besides, the QC step can also be applied to the transcripts. For human or mouse samples, raw count matrices typically contain around 30,000 genes, though not all of them are informative. Non-informative genes can be filtered out by setting a threshold for the minimum number of cells that express the gene. The choice of threshold should be adjusted based on the number of cells in the dataset and the objectives of the downstream analysis.

**Data normalization and transformation**

The "normalization" step is carried out on the raw count data after quality control to adjust for variable sampling effects. This process aims to transform the counts so that the variance falls within a specified range. It facilitates the use of generic statistical methods in subsequent analyses by reducing the bias from sequencing depth, or other technical factors. Different normalization methods are used in scRNA-seq data including the methods developed for bulk RNA-seq data and other scRNA-seq-specific normalization methods [125]. One of the most commonly used methods is the count depth scaling (e.g. CPM normalization), which can effectively uncover latent dataset structures (especially when followed by principal component analysis) as well as stabilize variance for other downstream analyses. There are other non-linear normalization methods by applying the parametric modeling like the negative binomial model [126], that's useful for more complex variations. After normalization, the data is typically log-transformed with a pseudo-count (commonly using 1). This transformation mitigates the mean-variance relationship and reduces the skewness in the scRNA-seq data, making it more closely to the normal distribution assumed by lots of downstream analysis methods.

The technical and biological covariates can still remain after data normalization, in-

cluding batch, dropout, or cell cycle effects. A simple linear regression can be performed to regress out the effects of these covariates as implemented both in Seurat and Scanpy. When multiple covariates such as cell cycle and count depth need to be corrected, the regression should be conducted on all covariates in one step to account for dependence between them. For batch effect removal, though this approach is applicable, it's more common to consider other specific methods discussed in the next section.

Normalization methods performed on the cellular level make the data more comparable between cells. Similarly, the normalized data can be further scaled on the gene level to improve comparisons between genes. Z-score is a method for gene scaling with a zero mean and a unit variance. This is commonly used for principle component analysis, while other than that, it's usually not suggested as much of the information would be lost [112].

**Batch correction and data integration**

Handling cells in distinct groups or "batches" is a common challenge in scRNA-seq data analysis. The batch effect can occur in many situations, such as different sample handling, experimental protocols, or sequencing depths, which are technical effects. It can also come from biological effects, for example, different donor or tissue sources, time points, etc. Whether or not to consider these biological factors as batch effects should depend on the study design and research question.

The removal of batch effects in scRNA-seq data is called "batch correction", and the term, "data integration", is also commonly used especially when dealing with complex data from different datasets where cell identities may not be shared. Various methods for batch correction have been developed: some directly adjust the gene expression data, while others operate within the dimensionality reduction space. Several benchmark papers have evaluated the performance of these batch correction methods. Buttner et al [127] highlighted the strong performance of Combat [128], originally developed for bulk RNA-seq data, which returns an expression matrix corrected for batch effects. Additionally, Harmony [129] and Seurat v3 [130], two popular methods specifically designed for scRNA-seq data, were shown to outperform other approaches by two recent benchmark papers [131, 132]. Recently, the development of deep learning-based methods has demonstrated good performance in tackling complex tasks, as highlighted in the benchmark study [133]. Although integration methods have been thoroughly evaluated across various benchmarks, it's

important to acknowledge that no single method proves optimal for all scenarios.

**Feature selection**

After filtering the non-expressed genes in the QC step, the number of genes for a single-cell dataset can still reach around 15,000. The "Feature selection" step aims to lower the dimensionality of scRNA-seq datasets, thereby reducing the computational demands as well as the noise in the data. The informative genes selected from this step are called highly variable genes (HVG). Depending on the complexity (e.g. cell heterogeneity) of the data, the number of HVGs can be decided, typically ranging between 1,000 and 5,000. Classical approaches include the dispersion-based method, which is implemented in Seurat [134] and Cell Ranger [135]. In the "Seurat" method, bins have equal width, whereas in the Cell Ranger method, the bins are determined by the percentile of the mean gene expression. Besides, Germain et al. proposed deviance [136] to select highly informative genes by fitting a gene-wise model that assumes constant expression across all cells and assess genes whether they violate this assumption.

**Dimensionality reduction**

As a high-throughput sequencing technology, scRNA-seq generates datasets with large numbers of cells and genes, leading to high-dimensional data. Feature selection helps exclude most uninformative genes, reducing data dimensionality. However, for visualization purposes, additional dimensionality reduction algorithms are still necessary. They are also helpful for downstream analysis as these methods reduce the data to its essential dimensions, potentially revealing its intrinsic dimensionality.

Principle Component Analysis (PCA) is a commonly used technique for dimensionality reduction in scRNA-seq data. This method reduces dimensionality by linearly transforming the data into a set of orthogonal variables known as principal components (PCs). The goal of PCA is to preserve most of the data's variations with fewer dimensions, removing components that show minimal variation. Although PCA does not capture the structure of the scRNA-seq data with fewer dimensions as well as non-linear methods, its high interpretability and computational efficiency make it a common initial step for non-linear dimensionality reduction methods. By selecting the first PCs with the largest possible variance, usually between 10 to 50 PCs are used for other downstream tasks.

A comparison study of 10 dimensionality reduction methods concluded that the t-

distributed stochastic neighbor embedding (t-SNE) method provided the highest accuracy and computing cost, while the uniform manifold approximation and projection (UMAP) algorithm demonstrated the greatest stability [137]. Both tSNE and UMAP are graph-based non-linear dimensionality reduction techniques, which are often used to visualize complex datasets into 2D or 3D spaces. tSNE [138] tries to optimize the similarity measure of data points in the lower dimensional space closely mirrors that in the high dimensional space. tSNE is advantageous in preserving the local structure of the data, while it suffers from high computational costs and inaccurate global data structure preservation. In contrast, UMAP [139] is faster and more effective at preserving the data's global structure as well as the local structure. In sum, both UMAP and tSNE are powerful tools for dimensionality reduction and visualization in scRNA-seq data analysis. While UMAP's advantages, including the increased speed, better preservation of global structure, and more intuitive parameters have made it a more effective and widely used tool for visualizing high-dimensional data.

**Cell clustering and annotation**

Cell clustering is commonly performed at the beginning of downstream analysis for scRNA-seq data. Cells are grouped into clusters based on the similarity of their gene expression profiles, allowing for the identification of distinct cell types or cell states. The gene expression similarity is measured by distance metrics from dimensionality-reduced representations. Clusters are then formed by minimizing the intra-cluster distance. K-means clustering is among the simplest and most popular machine learning techniques. It divides cells into k clusters by identifying k centroids and allocating each cell to the nearest cluster. In single-cell data, different distance metrics are used in k-means clustering, such as Euclidean distances, cosine similarity, and correlation-based distance metrics [140].

In addition to unsupervised clustering techniques, community detection methods like Louvain and Leiden are frequently used to generate clusters in scRNA-seq data. These methods are designed for analyzing the structure of large and complex networks and are based on the K-Nearest Neighbour (KNN) graph. Usually, K is set between 5 and 100 depending on the size of the dataset. The Louvain algorithm has been demonstrated to outperform other clustering techniques for single-cell RNA-seq data [141, 142]. However, this method may yield arbitrarily poorly connected communities, that are internally disconnected. The Leiden algorithm [143], as an improved version of Louvain, provides explicit guarantees for the connection of com-

munities and improves both the speed and quality of clustering results. By subsetting the KNN graph, the algorithm also allows the sub-clustering of specific cell groups, facilitating the detailed identification of cell states and enabling more precise cell type labeling.

Following clustering, cell clusters are annotated with different labels indicating their identities, which refer to cell types or cell states including cell subtypes or other unknown novel cell types. The cell annotation process can reveal the cellular composition of the sequenced sample, enhancing our understanding of the biological systems under investigation. There are multiple ways of cell annotation, either manual annotation by marker genes or applying automated annotation tools. For the first way, a reference marker gene list from literature or databases is needed. CellMarker [144] and PanglaoDB [145] are two databases containing a variety of cell type-specific marker genes with many of them identified from single-cell studies. Cell clusters can then be annotated by comparing data-derived marker genes with the reference or by directly visualizing these reference markers on the dimensionality-reduced space like UMAP. To identify data-derived marker genes, differential gene expression (DGE) analysis is performed between the cells in one cluster and all other cells in the dataset (see "Differential gene expression analysis"). Typically, the up-regulated marker genes of each cluster are used, and the comparison with the reference can be made using enrichment tests, the Jaccard index, or other overlap statistics. The manual annotation approach is knowledge-dependent with final labels manually decided by researchers. Therefore, it's time-consuming and may subject to human bias such as inconsistent cell type labels, however, this method provides high flexibility, particularly in identifying previously unknown cell types.

Alternatively, automated annotation tools provide a rapid way in cell type identification by using different strategies [146]. Tools like Garnett [147] reply on a predefined set of marker genes to transfer annotations between the reference and the dataset. CellTypist [148], initially developed for exploring tissue adaptation of cell types, offers built-in models for various immune cells trained on previously annotated datasets or atlases. Besides, by mapping the dataset to an existing and annotated reference dataset, cluster labels can be transferred based on their joint embedding, e.g. the transfer annotation implemented in Seurat [114] and scArches [149]. While in the presence of batch effects or other variations between two datasets, this query-to-reference mapping might not show well-integrated results.

**Trajectory inference**

Trajectory inference (TI) is to solve cellular differentiation and lineage determination at the level of individual cells. The accomplishment of this task involves ordering observed cells along the development trajectory based on similarities in their expression patterns. Trajectory inference methods are also called pseudotime analysis with pseudotime referring to an abstract unit of progress along the dynamic process. TI allows studying the progression of cells evolving from one state to another, and subsequently when and how cell fate decisions are made.

To apply TI methods properly, several general assumptions need to be satisfied, including (1) the biological process under investigation is dynamic, and the appropriate cells are sampled; (2) the biological data are sampled with a sufficient depth including transient states, ensuring the representation of the entire developmental trajectory; and (3) gene expression changes gradually throughout the developmental process. Additional assumptions or prior knowledge are required specifically by each method, such as the starting or ending state of the development process. TI is usually conducted after clustering and annotation, with data projected onto a lower dimensional representation, such as PCA.

The simplest developmental process can be described using the linear trajectory, such as the progression from immature cells, through intermediary stages, ultimately culminating in a terminal state. However, cells don't often follow linear processes. They make a commitment at some branching points and follow one of several potential routes. To suit different cases, many methods have been proposed. Saelens et al.[150] compared 45 TI methods in terms of their accuracy, scalability, stability, and usability. In general, a broad spectrum of performance was observed across the four evaluation criteria. Deconinck et al.[151] provides an exhaustive overview of current algorithms and their characteristics. These approaches include clustering-based, graph-based, manifold learning-based, and probability-based methods. Clustering-based methods identify the connections between cell clusters to form a trajectory structure and the minimum weighted spanning tree (MST) [152] is usually constructed to connect clusters. Graph-based methods find the connections between the lower dimensional representations of the cells and build a graph based on the clusters. One example of this method is the Partition-based graph abstraction (PAGA) [153], which expands the modeling of trajectory topologies to general graphs, facilitating the incorporation of loops or even multiple independent trajectories. Manifold-learning-based meth-

ods use principle curves or graphs to connect clusters and estimate the underlying trajectory structure. Slingshot [154] is one of the most widely used manifold-learning-based approaches. Finally, probabilistic methods model the developmental path by a probabilistic framework, providing a measure of certainty of a cell belonging to a specific branch of the trajectory.

While trajectory inference or pseudotime analysis can offer valuable insights, delving deeper into these biological processes often requires additional fine-grained analysis combined with other techniques. For example, to accurately capture the actual biological developments, it's suggested to incorporate RNA velocity information into TI, which is discussed in the next section. The PAGA graph, initially undirected, can then be extended to a directed transition graph with oriented edges. To discover which transcriptional factors drive and regulate the development process along the trajectory, differential trajectory analysis can be performed. TI method like Monocle [155] has incorporated this downstream analysis in its pipeline by fitting a generalized linear model. Other generic statistical frameworks have been suggested such as tradeSeq [156], capable of working with various (TI) methods. These methods provide a framework for testing differential expressed genes not only between different branches, but also between different conditions within the same branch.

**RNA velocity**

To improve the biological relevance of the inferred trajectories, additional resources of information rather than the gene expression counts are used. Among these, RNA velocity [157] stands out as the most widely adopted source of information. RNA velocity—the time derivative of the gene expression state—has introduced new ways to study cellular dynamics in single-cell RNA sequencing data. It can reveal both the descriptive state of a cell and its direction and speed of movement in the transcriptome space.

During the transcription process, DNA is first transcribed into unspliced precursor messenger RNA (pre-mRNA), which contains not only coding regions (exons) but also non-coding regions (introns). Then the introns are spliced out, turning into spliced mature mRNA. With single-cell RNA sequencing data, it's possible to reveal the abundance of unspliced and spliced mRNAs, and the change is called RNA velocity [157]. For each gene, when cells show a higher abundance of unspliced mRNAs than expected in the steady state, the RNA velocity is positive, meaning the gene is up-regulated, while negative velocity indicates the gene is down-regulated. Then

through the combination of RNA velocities across genes, a cell's future state can be estimated. The first method for estimating RNA velocities, known as the steady-state model, works under the assumptions that all genes have a uniform splicing rate and that the system reaches equilibrium during the experiment. However, these assumptions are often not true. In response to these limitations, the dynamic model, implemented in scVelo [158], has emerged to be the most widely used extension. This dynamic model doesn't assume the common splicing rate and the steady states of the system, allowing for a more nuanced analysis of RNA velocities, thereby providing deeper insights into cellular dynamics and developmental trajectories.

**Differential gene expression analysis**
Originating from bulk RNA-seq analysis, differential gene expression (DGE) is a fundamental question as well as in scRNA-seq data analysis. DGE analysis reveals genes that are differentially expressed between groups. When testing on different cell clusters, DGE analysis results in marker gene identification. When different conditions, like treatment groups, are present, DGE analysis is employed to uncover changes in gene expression profiles attributable to these conditions. Additionally with cell type annotation, cell type-specific differentially expressed genes can be identified by performing the test for each cell type. differentially expressed genes can be used for other downstream analyses, such as pathway analysis, to explore the underlying biological mechanisms.

Different test methods or tools used in bulk RNA-seq data can be applied in single-cell data, including the popular Student's t-test, Wilcoxon rank-sum test, and tools like edgeR, and DEseq2. The t-test and Wilcoxon rank-sum tests are simple naive tests for DE analysis in scRNA-seq data, showing high efficacy in marker gene selection [159]. However, these methods are not applicable to data with complex experimental designs as they cannot account for batch effects. Bulk DE tools such as edgeR and DEseq2 are used with "pseudobulk" by aggregating the expression for each sample-condition combination. Despite of the technical noise contained in scRNA-seq data such as dropouts, these bulk DE methods have shown favorable performance compared to methods explicitly designed for scRNA-seq data [160, 161]. Additionally, the single-cell method MAST [162] demonstrated the best performance among single-cell DE methods [161].

DE analysis generates results for each gene including log2 fold-change values and P values. log2 fold-change quantifies the magnitude of expression difference between

two conditions, where positive values indicate upregulation, and negative values denote downregulation. P values represent the statistical significance of the changes. To correct for multiple testing, instead, adjusted P values should be used. Methods like the Bonferroni correction, which divides the significance threshold by the number of tests, and the Benjamini-Hochberg procedure, which controls the rate of false discoveries, are commonly employed. Therefore, significant DE genes are obtained by filtering on the adjusted p values or in combination with log2 fold-change values.

**Enrichment analysis**

Following differential gene expression analysis, interpretation of the DE genes intends to discover gene sets with shared gene programs, such as biological processes and regulatory pathways. For this purpose, enrichment analysis is performed to explore pathways enriched related to specific condition changes. There are many public databases that archive the relationships between genes and their roles in biological functions or pathways. The Molecular Signatures Database (MSigDB) [163] is the most comprehensive database consisting of 9 collections of gene sets. The Gene Ontology [164] and KEGG [165] are the most frequently used databases. Gene Ontology (GO) categorizes gene functions into three aspects: Biological Process (BP), Molecular Function (MF), and Cellular Component (CC) with each GO term organized in a hierarchical parent-child structure. KEGG comprises a series of manually curated pathway maps depicting molecular interaction and reaction networks. Reactome [166] is another notable pathway and it provides the most nonredundant terms in terms of human data [167]. The choice to use specific databases often depends on their ability to provide meaningful information that enhances the understanding of the study's findings.

ORA and GSEA are two types of enrichment analyses commonly used in this process. ORA [168] is used for analyzing lists of significantly altered genes to find enriched pathways or functions by comparing them against a background set of genes. The gene list as input needs to be filtered before analysis (e.g., a P value cut-off for differential expression). ORA relies on statistical tests like the hypergeometric test and Fisher's exact test to identify specific pathways or functions that are over-represented in the list of genes. Unlike the over-representation analysis, GSEA [169] evaluates entire gene expression profiles to identify coordinated changes in the expression of predefined gene sets. Genes used for GSEA don't need to be filtered but are pre-ranked by test statistics, allowing the detection of subtle changes.

**Cell-cell communication analysis**

Cell-cell communication (CC) refers to the process by which cells interact with each other to coordinate their actions and maintain homeostasis. It's crucial for many biological processes, such as cell development, apoptosis, and migration. CCC can occur through several different ways: (1) Autocrine signaling, referring to intracellular communication by which cells communicate with themselves, (2) Paracrine signaling, when cells release signals to their nearby, but different, target cells, (3) Juxtacrine signaling, representing the communication by direct contact and (4) Endocrine signaling, by which signals (known as hormones) travel through the bloodstream to reach distant cells [170]. Previously, CC analysis was restricted to in vitro experiments involving one or two cell types and a small selection of genes. Since the development of single-cell technology, a number of computation tools have emerged to infer CC from scRNA-seq datasets. These methods use gene expression of the ligand and receptor as a proxy of protein abundance to infer intercellular crosstalk between a group of sender and receiver cells. CC analysis is typically performed after cell clustering in scRNA-seq data. By calculating an interaction score with different approaches, the communication can be quantified between a pair of cell groups, with one being the source and the other being the receiver.

Generally, CC methods are capable of capturing relevant biological signals, as demonstrated by the concordance with other modalities, however, they show relatively low agreement with each other [171]. Cellchat [172] and CellPhoneDB [173], two popular CCC tools, both have been reported for their high agreement with cytokine activities and robustness to noise [171]. Besides, another comparison study of CC tools claimed the better performance of statistical-based methods than network-based methods and ST-based methods by integrating the scRNA-seq data with spatial transcriptomics data [174]. Particularly, CellChat [172] was identified to perform the best in terms of consistency with spatial information, which might be attributed to its incorporation of regulatory information like cofactors with ligand-receptor interactions. Moreover, CellChat uses a comprehensive signaling database, called CellChatDB, which contains manually curated ligand-receptor complexes, which are multimeric ligands and receptors, together with cofactors, including soluble agonists, antagonists, co-stimulatory and co-inhibitory membrane-bound receptors. Recently, A new tool called Scriabin [175] has been introduced for performing CCC analysis at the single-cell level, potentially revealing communication that are obsured by these agglomerative methods. However, it does not take into account multi-subunit situa-

tions.

In sum, single-cell RNA sequencing data can be analyzed through a variety of approaches. Beyond the previously mentioned methods, gene regulatory network (GRN) analysis is also frequently applied, which aims to uncover gene regulatory interactions based on the co-expression patterns in scRNA-seq data. Additionally, more and more innovative algorithms are under development in this field, each bringing unique perspectives to the analysis. However, no single method universally fits all datasets, the choice of analytical approach must be carefully tailored to suit the specific characteristics and requirements of each dataset.

## 1.5 Single-cell TCR sequencing analysis

T cell receptors are heterogeneous surface antigen receptors expressed on T cells, which are responsible for antigen recognition by interacting with antigen MHC molecules. The TCR repertoire refers to the total profile of TCRs of one individual. It has been always of great interest in the characterization of TCR repertoire since it accurately describes T cell dynamics, reflecting the immune status of T cells under different biological conditions, such as cancer, autoimmune, and infectious diseases [176]. TCR sequencing facilitates the high-throughput analysis of TCR arrangement, and its combination with single-cell technology, enables the detailed examination of the diversity and clonality of T cell populations.

The development of high-throughput sequencing technologies has enabled deep and comprehensive immune repertoire profiling at the single sequence level, providing a high-resolution view and extensive data on the dynamic changes in the clonality of immune cells [177]. Before the single-cell approach, bulk TCR sequencing was limited to the analysis of single TCR chains. However, the application of single-cell techniques has made it possible to profile paired TCR chains within individual cells. Moreover, the joint profiling of single-cell TCR sequencing (scTCR-seq) and scRNA-seq is now available by many platforms (e.g. 10X Genomics), which further allows advanced integrative analysis of these two methodologies. Several bioinformatics tools and pipelines have been developed accordingly, providing various analyses including clonal diversity, clonal overlap, and gene usage analyses on TCR-seq data, as well as the interaction with scRNA-seq data, such as scRepertoire [178] and Scirpy

[179].

**Preprocessing and quality control**
Similar to single-cell RNA sequencing, data pre-processing and quality control are also required for scTCR-seq data. Firstly, the alignment of TCR raw sequencing reads are performed to a reference set of V, D, and J gene sequences. For data using 10X Genomics, this process can be done through 10X Genomics Cell Ranger software. As a result, the amino acid and nucleotide sequences for the complementarity determining regions (CDRs), together with annotated VDJ genes are returned for each T cell receptor chain.

The productive TCR contigs represent those that are likely to correspond to functional T cell receptors, thus non-productive TCRs are usually removed. Besides, examination of TCR chain compositions can also help to ensure the data quality. Since most T cells only contain one pair of chains ($\alpha\beta$ or $\gamma\delta$) and less than one third of T cells can have dual TCRs (two pair of receptors originating from different alleles) [180], T cells with multi-chains (more than two receptor pairs) probably are doublets, which should be excluded from downstream analysis.

**Clonal analysis**
After quality control, the T cell clonotype can be decided typically by the unique sequence combination of V gene, J gene, and CDR3 nucleotide sequence or amino acid sequence. With the identification of clonotype, a variety of clonal analyses can be performed, such as clonal expansion, diversity and overlap analysis.

Clonal expansion analysis is to quantify the clonotype distribution and reveal the expansion of clonotypes. When a specific antigen is recognized, T cells undergo rapid division and expansion, producing a large number of cells with the same clonotype, thus the same antigen specificity. By analyzing these clonal expansions, the strength and specificity of the immune response can be uncovered. Additionally, by comparing the clonal expansion between different samples or conditions (e.g. tissues), the variability and dynamics of T cell immune responses under various physiological and pathological states can be revealed.

Similarly, clonal diversity analysis can also reveal the immune response of T cells, with high diversity indicating a healthy immune system capable of responding to a wide array of pathogens, whereas low diversity suggests immune exhaustion or clonal dominance. Clonal diversity can be measured for samples or other group

variables by different indices, including Shannon, inverse Simpson, and Gini indices [178].

To measure clonal similarity between samples, the clonal overlap analysis is conducted. One method is to calculate the clonotype overlap that is scaled to the length of unique clonotypes in the smaller sample. Two other indices can be used as well: Morisita index and Jaccard index. By revealing clonal overlap between different cell groups, we can find out common immune responses and how these responses may be influenced by specific factors.

Besides, clustering of TCR profiles are possible based on the computation of TCR sequence similarity or distance as implemented by tcrdist [181]. Gene usage analysis on V and J genes can provide insights for the gene alterations in the repertoire, which may be caused by diseases or other special conditions. The distribution of the length of CDR3 sequences among different conditions can be biased as well. For example, Sureshchandra et. al. applied scTCR-seq and discovered the vaccination preferentially enriched distinct TCRs with an increased CDR3 length [182].

**Integrative analysis with scRNA-seq**

The integration of the paired single-cell RNA-seq and TCR-seq can be achieved by matching the cell barcodes from both modalities. Then T cells kept for integrative analysis are available for both gene expression and TCR profile. One of the most valuable information provided by scRNA-seq is the cell type annotation. With that, different T cell subsets can be compared in terms of their clonal expansion, diversity and overlap. For example, by comparing the clonal expansion between different subsets of T cells, the hyper-expanded cell groups can be revealed, which may elucidate the mechanism of specific cell types associated with diseases [183]. Another straightforward integrative analysis is to visualize the clonal properties on the dimensionality-reduced representation such as UMAP from the scRNA-seq data.

On the other way, TCR information can also facilitate the analysis used for scRNA-seq data. Differential gene expression analysis can be applied between clonally expanded T cells versus non-expanded T cells or on specific T cell clones. TCR-seq data can also combine with cell trajectory analysis, as implemented in the pipeline LRT [184], which provides a computational framework to identify clusters of clonotypes showing distinct skewed distribution along the differentiation path. Moreover, unbiased correlation analysis between gene expression and TCR profiles is supported

by Clonotype Neighbor Graph Analysis (CoNGA) [185]. It builds two graphs: one based on the similarity of TCR sequences and the other on gene expression data, establishing a correlation between the two.

Typically, the integrative analysis of scRNA-seq and scTCR-seq is conducted following the individual downstream analyses of each dataset. As multi-modal data integration techniques continue to evolve, the development of innovative methods is expected to offer fresh perspectives and enhance the integrative analysis of TCR-seq data with transcriptome data as well as with other omics datasets.

## 1.6 Research question

Recently, the advent of single-cell technologies has begun to deepen our understanding of the heterogeneity and adaptation of immune cells in human tissues, including the three tissues of interest —visceral adipose tissue, liver and small intestine [58, 94, 186]. Several cross-tissue studies using single-cell RNA sequencing technique have been published, which aroused great impact in this field [33, 148, 187, 188].

Han, X. et al. [187] have created a comprehensive single-cell atlas of humans, encompassing not only immune cells but also stromal and epithelial cells across over 50 tissues. While this atlas marks a significant milestone in the development of the Human Cell Atlas, it lacks a detailed comparison of gene expression profiles among immune cells, particularly T cells, across different tissues. Furthermore, the representation of T cells is limited in tissues where immune cells constitute a minor proportion. C. Domínguez Conde et al. [148] examined all immune cells extracted from 16 donor-matched human tissues, including omentum, liver, and several different segments of SI using scRNA-seq. This study highlighted the diverse immune cell compositions and signatures across tissues. However, the small number of T cells identified in the omentum (fewer than 400) made it difficult to derive reliable conclusions about the expression patterns of tissue-resident T cells in human adipose tissue. Additionally, another recent cross-tissue study [33] interrogated T cells across major barrier sites (skin, lung, jejunum) and lymphoid organs by integrating single-cell protein and transcriptome profiling, revealing tissue-specific adaptation and clonal segregation of Trm cells. These cross-tissue studies have great value in uncovering the shared or tissue-specific cell states and expression modules. Nevertheless, it's

still unclear for tissue T cell populations regarding their cell identity, function, and location.

Based on these contexts, this study aims to uncover the adaptation of human T cells in different organs including visceral adipose tissue, liver, and small intestine, together with blood as the peripheral reference. Using the single-cell RNA sequencing and TCR sequencing techniques, both the transcriptome landscape and the T cell receptor repertoire could be captured. Additionally, the scRNA-seq data contains innate immune cells of myeloid cells and natural killer cells for the purpose of comparison and cell interaction analysis. With this valuable data resource, several key questions are expected to be addressed: (1) the cellular composition of T cell subsets within each tissue and the potential difference. (2) the common and unique T cell adaptation gene signatures (3) the functions or mechanisms underlying T cell adaptation in different tissues, such as distinct signaling pathways involved. Accordingly, different computational methods were applied for various analyses of the cross-tissue high-throughput dataset in this study, offering a comprehensive understanding of T cell adaptation.

# 2 Material and methods

## 2.1 Processing and sequencing of human tissue samples

Four human tissue biopsies were obtained from six donors and the basic donor information was shown in Table 1. These human tissue samples were processed and sequenced with the following methods, which were performed and provided by my collaborators, mainly Jonas Wagner and Guido Rattay from the research group of Nicola Gagliani at the University Medical Center Hamburg-Eppendorf.

**Table 1:** Donor information

| Donor_ID | Age | Sex | BMI |
|---|---|---|---|
| 1 | 42 | m | 38.68 |
| 2 | 39 | m | 40.91 |
| 3 | 32 | f | 42.61 |
| 4 | 51 | f | 43.71 |
| 5 | 61 | f | 44.19 |
| 6 | 29 | f | 48.05 |

**Cell isolation**
All procedures were conducted at room temperature unless otherwise specified. Fresh organ samples were weighed, washed once in PBS, and then finely minced to about one cubic millimeter. Small intestine samples underwent a 20-minute incubation in DTT buffer at 37 °C with shaking to break disulfide bonds, followed by centrifugation at 400 RCF for 5 minutes. Liver and SI samples were then digested with collagenase for 30 minutes at 37 °C with shaking. Subsequently, liver samples were filtered through a 40 µm filter, while SI samples used a 100 µm filter, followed by a wash in MACS buffer and another centrifugation at 400 RCF for 5 minutes. A density gradient centrifugation was then performed to purify the leukocytes: 4ml of 60% Percoll was

placed at the bottom of a 15ml tube, with the cell suspension in 4ml of 40% Percoll layered carefully on top. This setup was centrifuged at 400 RCF for 20 minutes without acceleration or brakes, and the leukocyte layer collected, washed in MACS buffer, and prepped for further analysis. For VAT samples, due to different tissue properties, they were minced and directly digested in collagenase buffer II for 45 minutes at 37 °C with shaking. Filtration was through a 70 µm then a 40 µm filter, washed in MACS buffer, and centrifuged at 400 RCF for 5 minutes. Erythrocytes were lysed with ACK buffer, followed by a wash in MACS buffer, preparing cells for subsequent steps. As for blood cell isolation, EDTA-diluted blood was mixed 1:1 with DPBS, layered over 4ml Pancoll, and centrifuged at 400 RCF for 20 minutes. The plasma layer was removed, cells washed thrice in DPBS, and then processed further, including for flow cytometry.

**Single-cell sequencing**

The single-cell RNA sequencing (scRNA-seq) and T-cell receptor (TCR) sequencing were applied with the droplet-based approach from 10X Genomics, following the manufacturer's protocol. The 5' v1.1 kit was used for parallel TCR and transcriptome profiling. In short, for each tissue sample (blood, liver, VAT, SI), Live/Dead-CD45+CD19- cells were sorted. These isolated cells were individually loaded into wells on a Chromium chip along with partitioning oil and a master mix that included RT Enzyme Mix B, RT Reagent B, Poly-dT RT Primer, and Additive A. This setup was then processed in the Chromium Controller to generate Gel Beads-in-emulsion (GEMs), where individual cells were encapsulated with barcoded gel beads within oil droplets, facilitating the capture of mRNA from each cell and its subsequent reverse transcription into cDNA. Following GEM generation, a critical cleanup step was executed to separate the cDNA from the partitioning oil and recovery agent. This involved a phase separation, where the aqueous phase containing the cDNA was isolated and further purified using a Dynabeads cleanup mix with magnetic silane beads. This allowed for the selective binding and subsequent washing of the cDNA, which was then eluted and subjected to amplification to increase the quantity of cDNA available for sequencing.

For the generation of TCR libraries, the targeted enrichment of V(D)J segments from the amplified cDNA is necessary to ensure the library's specificity to TCR sequences. This process involved additional rounds of amplification and cleanup to prepare the enriched cDNA for library construction, which was then validated for quality before

sequencing. The gene expression libraries were constructed through a similar meticulous process, ensuring that each step—from initial cell lysis and cDNA synthesis to the final library preparation—was optimized for high-quality sequencing outcomes. Both TCR and gene expression libraries underwent a final quality assessment before being submitted for sequencing on an Illumina 6000 platform, facilitated by Novogene.

**Flow cytometry**

Flow cytometry analyzes cellular characteristics via laser light, measuring cell size with forward scatter and granularity with side scatter. Cells are labeled with fluorescent antibodies specific to proteins, with fluorescence excited by lasers for signal detection. To address various research needs, different fluorescent antibody cocktails were prepared, with compensation applied to correct fluorescence overlap. For surface staining, cells were incubated with these antibodies at 4 °C in the dark, washed, and prepared for analysis in FACS tubes. Intracellular cytokine staining involved fixing cells in formalin, permeabilizing with NP-40, then staining with antibodies, all performed in the dark to preserve fluorescence. Similarly, transcription factor staining used a specific staining kit for fixation and permeabilization before antibody incubation. Fluorescence-activated cell sorting (FACS) sorts cells post-staining using a BD Aria Fusion Sorter. This streamlined process—from staining protocols to sorting—allows for detailed cellular analysis and sorting based on fluorescent markers.

## 2.2 Single-cell sequencing data processing

**Alignment**

The FASTQ file for each sample was obtained as the raw sequencing result. 10X Genomics Cellranger software (v 4.0.0) was used to align and quantify these raw sequencing reads from both single-cell RNA sequencing and the paired single-cell TCR sequencing. Specifically, the "cellranger count" pipeline was used to generate the gene expression count matrix from raw FASTQ files of scRNA-seq. Moreover, the "cellranger vdj" pipeline was applied to T cell receptor profiling data for sequence assembly and clonotype calling. The resulting feature counts as well as the annotated contigs for TCR repertoire were contacted across all samples for quality control.

**Quality control**

The raw gene count matrix with all samples was ingested into the Scanpy (v 1.8.2) [115] framework for quality control and the basic downstream analyses. As the first step of quality control (QC) in scRNA-seq data analysis, cells of low quality were removed based on the cell QC metrics: (1) the number of counts per barcode (count depth), (2) the number of genes per barcode, and (3) the fraction of counts from mitochondrial genes per barcode [112]. The thresholds of these three QC covariates were jointly considered based on their distributions. As a result, cells with fewer than 300 detected genes or cells with more than 3000 gene counts or 7% mitochondrial gene counts were filtered out. To further remove doublets from the dataset, Scrublet [124] was used with the default parameters on each sample. Briefly, Scrublet simulates doublets from the data by merging random pairs of observed transcriptomes to construct a k-nearest-neighbor classifier. Then it assigns a score for each cell, which are then automatically thresholded to identify doublets. With the returning boolean array "predicted_doublets" from Scrublet, cells identified as doublets were removed from the dataset. Finally, after quality control, around 144,000 cells from scRNA-seq data were considered of high quality, which could be used for downstream analysis.

**Data normalization and transformation**

Data normalization was performed using the count depth scaling method, which scaled the count depth by a size factor. The size factor can be determined with different approaches, which are usually calculated by:

$$s_c = \frac{\sum_g y_{gc}}{L} \tag{2.1}$$

here, $y$ means the raw count, $c$ represents the cell index, $g$ represents the gene index, and $L$ is the target sum. Using the "scanpy.pp.normalize_total" method, the dataset was normalized with the target sum $L = 10000$. Then the normalized data was log-transformed with a pseudo count by $log(\frac{y}{s} + 1)$, implemented with "scanpy.pp.log1p" method. Additionally, z-score transformation was applied to scale the normalized expression on the gene level, which was mainly used for the principle component analysis.

**Feature selection**

To select the highly variable genes (HVG), the dispersion-based method was used. Specifically, the dispersion is defined based on the mean and variance of genes among

cells:

$$dispersion = log(variance/mean) \tag{2.2}$$

The normalized dispersion is obtained by scaling with the mean and standard deviation of the dispersions for genes falling into a given bin based on their mean expression levels. HVGs are then ranked and selected by this normalized dispersion. Based on this method, for each subject, "scanpy.pp.highly_variable_genes" returned a list of HVG by setting the cut-off values (min_mean=0.0125, max_mean=3, min_disp=0.5). Those genes that were identified as the HVG in at least three subjects were selected to be the final highly variable genes for the dataset.

**Batch correction**

Harmony [129] was used for integrating data samples from different subjects. This algorithm starts with a low-dimensional cell embedding space, typically PCA. It tries to adjust this embedding so that cells can group according to their true functional differences (e.g.cell types) rather than the batch effects. To achieve this goal, Harmony uses fuzzy clustering to assign each cell to multiple clusters and a penalty term to maximize the diversity of datasets (batches) represented within each cluster. After that, the global and dataset-specific centroids for each cluster are calculated and within each cluster, the dataset-specific correction factor is derived from these centroids. Subsequently, each cell is adjusted using a cell-specific correction factor, which is a linear combination of dataset correction factors, weighted according to the cell's soft cluster assignments. Given that each cell might be in multiple clusters, Harmony iterates these steps until cell cluster assignments are stable. To implement Harmony algorithm, firstly the top 50 principal components were obtained using "scanpy.pp.pca" on the scaled data matrix. Then the Harmony batch correction was conducted on this PC space by "scanpy.external.pp.harmony_integrate" with "donor" as the batch key. As a result, the batch-corrected principal components were returned, to replace the original principal components for downstream analysis.

**Dimensional reduction**

In addition to PCA, the UMAP algorithm [139] was applied to further reduce the dimensionality of the scRNA-seq data and facilitate visualizations on a two-dimensional space. UMAP works by optimizing a low-dimensional graph to be structurally close to the high-dimensional graph. Based on the theories of Riemannian geometry and algebraic topology, the algorithm defines a fuzzy topological representation, also known as "Čech complex", from the high-dimensional data space. The constructed

representation is a weighted graph, where edge weights indicate the likelihood of two points being connected. After building the final fuzzy simplicial complex, UMAP uses a force-directed graph layout algorithm to project the data into a lower-dimensional space, efficiently preserving both local and global structures. There are several parameters influencing the final shape of the UMAP figure. The most important parameter is "n_neighbors", which means the number of neighbors used for the high-dimensional graph. It plays a key role in balancing local versus global structure in UMAP—a lower value highlights local structure while a higher value emphasizes the global data structure at the expense of losing some details. Besides, the parameter "min_dist" is also commonly used, representing the minimum distance between points in low-dimensional space.

To perform UMAP in Scanpy workflow, the neighborhood graph was computed first by "sc.pp.neighbors" (n_neighbors=15) on the adjusted PC space from Harmony. Different numbers of PCs were used in this study depending on their explained variance: 20 PCs for global cell lineages as well as CD4+ T cells and CD8+ T cells, and 30 PCs for all the T cell clusters. Next, using the "scanpy.tl.umap" method, with the default parameters (min_dist=0.5), the UMAP coordinates were calculated for different cell groups.

## 2.3 Single-cell RNA sequencing downstream analysis

**Cell clustering and annotation**

To group cells into different cell clusters with similar gene expression profiles, the community detection method Leiden algorithm [143] was used. This method relies on a graph representation, the neighborhood graph and it consists of three steps: (1) local moving of nodes, (2) refinement of the partition, and (3) aggregation of the network based on the refined partition. Starting with a singleton partition, where each node forms its own community, the algorithm iterates through these steps until no further improvements can be made, and the final partition is obtained. In the implementation of the Leiden algorithm, the "resolution" parameter controls the scale of the cluster partition, where a higher value leads to more clusters and a lower value results in fewer clusters.

In this dataset, several rounds of clustering were performed using the Leiden algo-

rithm by "scanpy.tl.leiden" with different resolutions in order to reveal both the global cell lineages and the detailed cell subsets. Specifically, a range of resolutions was explored during each clustering step to ensure optimal classification of cell clusters. The final resolution value was determined based on the appropriate cluster numbers and the performance of distinguishing biologically different cell identities. Importantly, cell clustering was iteratively refined in cases where cell groups were inaccurately combined. For example, NK cells might initially co-cluster with CD8+ T cells during the preliminary clustering process and they were separated after the sub-clustering.

The top-ranked cluster-specific marker genes were identified by using the "scanpy.tl.rank_genes_groups" method with the Wilcoxon algorithm. They were compared with the known public marker genes for immune cells. Taking together, the marker genes for annotating different T cell subsets in this study were obtained and shown in the Table 2.

**Table 2:** Core marker genes of T cell types

| Cell type | Marker genes |
|---|---|
| Naive CD4 T cells | CD4, SELL, CCR7, KLF2, LEF1, TCF7, MAL |
| Memory CD4 T cells | CD4, LTB, TXNIP, ITGB1, S100A4, AQP3, MAL |
| Intermediate CD4 T cells | CD4, CXCR4, LTB, CREM, FOS, CCR7 |
| CD4 Trm cells | CD4, CD69, PRDM1, CXCR4, CXCR6, CD44, S100A4 |
| Th17 cells | CD4, CD69, KLRB1, NFKBIA/Z, CCR6, CXCR6, CCL20 |
| FOXP3+ Tregs | CD4, FOXP3, CTLA4, TIGIT, IL2RA |
| Naive CD8 T cells | CD8A, SELL, CCR7, KLF2, LEF1, NELL2 |
| CD8 Temra/Teff cells | CD8A, FGFBP2, CX3CR1, GZMH, FCGR3A, KLRG1, S1PR5 |
| CD8 Tem cells | CD8A, GZMK, GZMH, PRF1, CD74, CCL4 |
| CD8 Trm cells | CD8A, CD69, CD160, XCL2, CCL5, ITGAE |
| MAIT cells | CD8A, RORA, TRAV1-2, KLRB1, CCL20, CCR6, ZBTB16 |
| CD8 NKT cells | CD8A, NKG7, FCGR3A, NCR3, CD247 |
| MT+ T cells | MT2A, MT1X, MT1E, MT1F |
| HSP+ T cells | HSPA1A, HSPA1B, HSPA6, HSP90AA1 |
| TNF+ T cells | TNF, IFNG, FOS, FOSB, CD69 |
| gd T cells | TRDC, KIR2DL4, CD247 |

**Differential gene expression analysis**

Differential gene expression analysis was performed between different tissues to identify tissue common and specific adaptation signatures. The MAST algorithm [162] was applied to the cross-tissue dataset. MAST uses the hurdle model, a two-component,

generalized linear model, to compare the gene expression patterns in scRNA-seq data. The first component models the discrete expression rate of each gene across cells by logistic regression, and the second component models the conditional expression level by Gaussian. The model can be fitted as follows:

$$logit(Pr(Z_{i_g} = 1)) = X_i \beta_g^D \tag{2.3}$$

$$Pr(Y_{i_g} = y | Z_{i_g} = 1) = N(X_i \beta_g^C, \sigma_g^2) \tag{2.4}$$

Here, $Z = [Z_{i_g}]$ is the indicator, meaning whether the gene $g$ is expressed in cell $i$. $Y_{i_g}$ is the normalized scRNA-seq expression for gene $g$ in cell $i$. $X$ is the design matrix of the model, including the different conditions the model aims to test as well as the covariance the model wants to exclude. Based on this hurdle model, MAST is able to account for the bimodality of scRNA-seq data. The results combined from these two parts of the components are used to jointly infer differential expressed genes.

Using the normalized gene expression matrix as input, the cellular detection rate was additionally included in the MAST model as the covariate in order to remove the potential confoundings from the detection of gene expressions in cells. The returned P values were adjusted using Bonferroni correction. Consequently, genes were identified as significantly differentially expressed between tissues with adjusted P value < 0.05 and abs(log2(fold-change)) > 0.5.

**Pathway analysis**

The human Reactome database [166] was used for the pathway enrichment analysis of differentially expressed gene sets. To obtain the enriched pathways, the over-representation test was performed within the package "clusterProfiler" (v4.8.3) [189], where it applied the simple hypergeometric test. The P value of the enrichment for a specific pathway is computed as:

$$p = 1 - \sum_{i=0}^{k-1} \frac{\binom{M}{i}\binom{N-M}{n-i}}{\binom{N}{n}} \tag{2.5}$$

here, $N$ is the total number of genes in the background, $M$ is the size of the gene set annotated to the pathway in the database, $n$ is the size of the input gene list and $k$ is the number of input genes that are found in the gene set for the pathway. The P values were adjusted for multiple hypothesis testing, which were used to determine the significant pathways.

**Velocity and trajectory analysis**

The spliced and unspliced expression matrices for each sample were firstly generated from the "cellranger" output using Velocyto (v 0.17.15) by "velocyto run10x" [157]. Then the merged expression matrices were preprocessed following the standard steps in scVelo (v 0.2.4) [158]. To estimate the RNA velocity, a basic gene-specific model was used:

$$\frac{du(t)}{dt} = \alpha(t) - \beta(t)u(t) \tag{2.6}$$

$$\frac{ds(t)}{dt} = \beta(t)u(t) - \gamma(t)s(t) \tag{2.7}$$

Here, $\alpha(t)$ is the transcription rate, $\beta(t)$ is the splicing rate and $\gamma(t)$ is the degradation rate. To remove the potential noise, the unspliced counts $U$ and spliced counts $S$ are normalized separately by their size factors: $u = U/N_u$, $s = S/N_s$, where $N_u$ and $N_s$ are the total numbers of unspliced and spliced molecules observed in a given cell, respectively. Based on the above model, the steady-state model can be applied to estimate the RNA velocities, which requires three strong assumptions: (1) kinetics reached their equilibrium, (2) rates are constant across all cells, and (3) the splicing rate is common across all genes. This model can effectively capture the developmental direction in certain systems, while these assumptions are limited in circumstances where a population consists of diverse subpopulations with varying kinetics.

To address the constraints of the steady-state model, the dynamic model was applied with scVelo by "scv.tl.recover_dynamics" and the velocity was computed by "scv.tl.velocity". Under this likelihood-based dynamical model, the assumptions of steady states as well as the common splicing rate across genes are no longer valid. The transcription rate $\alpha$ is thought to be state-dependent, which can be denoted as $\alpha^{(k)}$, where $k$ is a latent variable representing the cell's transcriptional state. Four states of $k$ were considered: induction ($k = 1$), repression ($k = 0$) and their associated steady states ($k = ss_1 and k = ss_0$). The two sets of parameters were estimated using all data points: (1) the gene-specific reaction rates of transcription $\alpha$, splicing $\beta$ and degradation $\gamma$, (2) cell-specific latent variables including the discrete transcriptional state $k_i$ and the continuous latent time $t_i$, where $i$ represents a single cell. The inference of these parameters was based on the expectation-maximization (EM) framework. Specifically, for the model estimate of the unspliced/spliced phase trajectory: $\hat{X}(t) = (\hat{u}(t), \hat{s}(t))$, in the expectation step, the latent time $t_i$ is assigned by minimizing the distance to the phase trajectory and the transcriptional states $k_i$ are decided

by associating a likelihood to corresponding segments on the phase trajectory. In the maximization step, the parameters of the reaction rates are updated to optimize the overall likelihood.

To facilitate visualization, the obtained high-dimensional velocity vectors were projected onto a low-dimensional space UMAP by "scv.pl.velocity_embedding_stream". Additionally, the dynamical model recovered the latent time of the underlying cellular processes ("scv.tl.latent_time"), representing the cell's internal clock of their differentiation. The cluster-specific top-likelihood genes were identified with "scv.tl.rank_dynamical_genes" method. Briefly, genes were ranked based on their goodness of fit by aggregating across cells to derive the overall gene likelihood. This approach enables the identification of genes demonstrating notable dynamic behavior, with genes ranked among the top exhibiting clear signs of splicing dynamics, whereas the expression of low-ranked genes is mostly governed by noise or nonexisting transient states.

Next, the PAGA graph [153] was used to construct the cells' developmental trajectory. With the RNA velocity, the directed PAGA graph was built with "scv.tl.paga". PAGA uses cell clustering to partition the graph and then reveal the connectivity of the manifold partitions by a statistical model. Specifically, the single-cell data is denoted as a neighborhood graph in manifold learning techniques: $G = (V, E)$, where each vertex in set $V$ represents a cell and each edge in set $E$ corresponds to a neighborhood relation. After partitioning or clustering of cells, the graph can be simplified to a PAGA graph $G^*$, where the nodes represent cell groups and the weights of the edges correspond to the "PAGA connectivity" between groups. The PAGA connectivity measure is a statistical measure of confidence, defined as the ratio of the number of inter-edges between the clusters, normalized by the expected number of inter-edges under random assignment of edges. It allows the identification of the connected and disconnected regions of the data. The distance measure derived from random walks on the single-cell graph can be integrated with the identified high-confidence paths in the PAGA graph. Subsequently, cells within each partition can be ordered based on their distances from a designated root cell.

Besides, in order to compare the gene expression patterns along the trajectory between different tissues, a simple multiple linear model was established for each gene signature $g$:

$$Y_g = \beta_{0_g} + Tissue\beta_{1_g} + LatentTime\beta_{2_g} + Donor\beta_{3_g} \tag{2.8}$$

Here, $Y_g$ is the normalized gene expression, $LatentTime$ is the cells' latent time derived from scVelo, $Tissue$ is the variate of interests and $Donor$ is the covariate that the model tries to exclude. By examining the P values from the parameter $\beta_{1_g}$, which were adjusted for joint comparison of multiple genes, the significance of the gene on the expression differences between tissues were revealed.

**Cell-cell communication analysis**
Cellchat (v 1.4.0) [190] was applied to infer cell-cell interactions across all recovered immune cell types. The human database curated in CellChat including the "Secreted Signaling", "Cell-cell Contact" and "ECM-Receptor" were used. Within the database, there are 2021 validated interactions, that are further grouped into 229 functionally related signaling pathways, enabling more biologically meaningful interpretation. CellChat quantifies the communication between cell groups by assigning each interaction with a probability value. Firstly, significant communications are identified for each cell group by the DE analysis with the Wilcoxon rank-sum test. Then the average gene expressions of ligands or receptors are projected onto a protein-protein network through a random walk-based network propagation approach. With that, the intercellular communication probability is calculated according to the law of mass action. Notably, the communication probability is not exactly a probability in statistics but only a measure of the interaction strength. Finally, the significance of interactions is given by probability P values using a permutation test.

Since there are four tissue samples in the dataset, Cellchat was applied to each of them separately, and then the comparison analysis between the liver and VAT was performed according to the Cellchat comparison analysis workflow. By comparing the overall information flow (defined by the sum of communication probability among all pairs of cell groups) of each signaling pathway, CellChat identified the conserved signaling pathways in the specific tissue.

## 2.4 Single-cell TCR sequencing analysis

The annotated TCR contigs from the "cellranger" software were analyzed within the scRepertoire (v 1.7.1) [178] framework. To filter out poor-quality reads from the TCR-seq data, only productive TCR contigs are kept. Besides, cells with multi-chains TCR were filtered out, which were probably doublets. Given that the majority of T cells

have $\alpha\beta$ TCR chains, this study only focused on analyzing those TCR sequences. The TCR clonotypes were defined by their V(D)JC genes together with the nucleotide sequence of their CDR3 regions. Cells with the same clonotype were considered as clonally expanded cells. To combine the scRNA-seq data of T cells and scTCR-seq data, cell barcodes were matched and only those that existed in both data modalities were used for further analysis.

TCR clonal analysis was performed on the integrated data. Based on the defined clonotypes, the clonal expansion scale was measured and then cells were classified into five groups: Hyperexpanded (100 < X), Large (20 < X <= 100), Medium (5 < X <= 20), Small (1 < X <= 5), and Single (0 < X <= 1). Clonal expansion analysis compared the distribution of these cell clonal groups in each tissue for different T cell sub-groups identified from scRNA-seq analysis. Besides, clonal overlap analysis was used to examine how many clonotypes were shared between different tissues. The results were shown by the circle plot with the width representing the number of clonotypes shared between two tissues.

# 3 Human tissue T cell adaptation

## 3.1 Cross-tissue analysis of paired scRNA-seq and TCR-seq data

Human biopsy samples from blood, liver, visceral adipose tissue (VAT), and small intestine (SI) were collected from 6 donors aged 25–65 years old (Table 1). Single cells obtained after FACS sorting were applied for droplet-based single-cell RNA sequencing (scRNA-seq) coupled with single-cell TCR sequencing (scTCR-seq) (Figure 2).

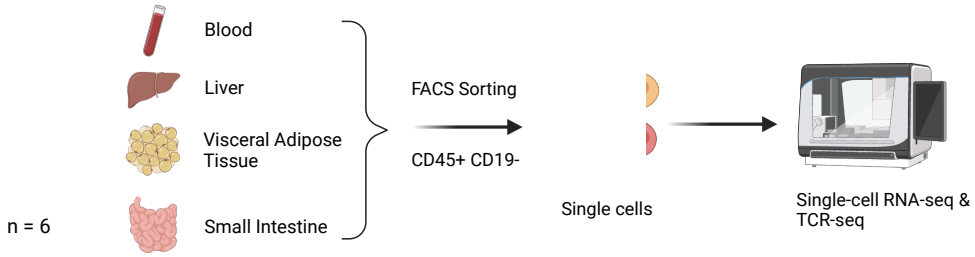



**Figure 2: Overview of single-cell experimental design**

A total of 142,840 cells were retained after quality control including 3 major cell lineages: T cells, Myeloid cells, and NK/Innate cells (Figure 3.A). The mean expression of three main lineage markers (*CD3D*, *CD68* and *NCAM1*) were utilized for global cell annotation. Among these, T cells presented the highest quantity in each of the four tissues examined (Figure 3.B). NK cells were mainly accumulated in liver and blood samples with fewer observed in VAT and SI. Myeloid cells were primarily found in VAT and blood. Due to some experimental problems, two donors (donor 1 and donor 6) contributed a relatively small number of cells as shown in Figure 3.B. Consequently, cells from these two were only included in cell clustering and annotation and excluded from all other downstream analyses.

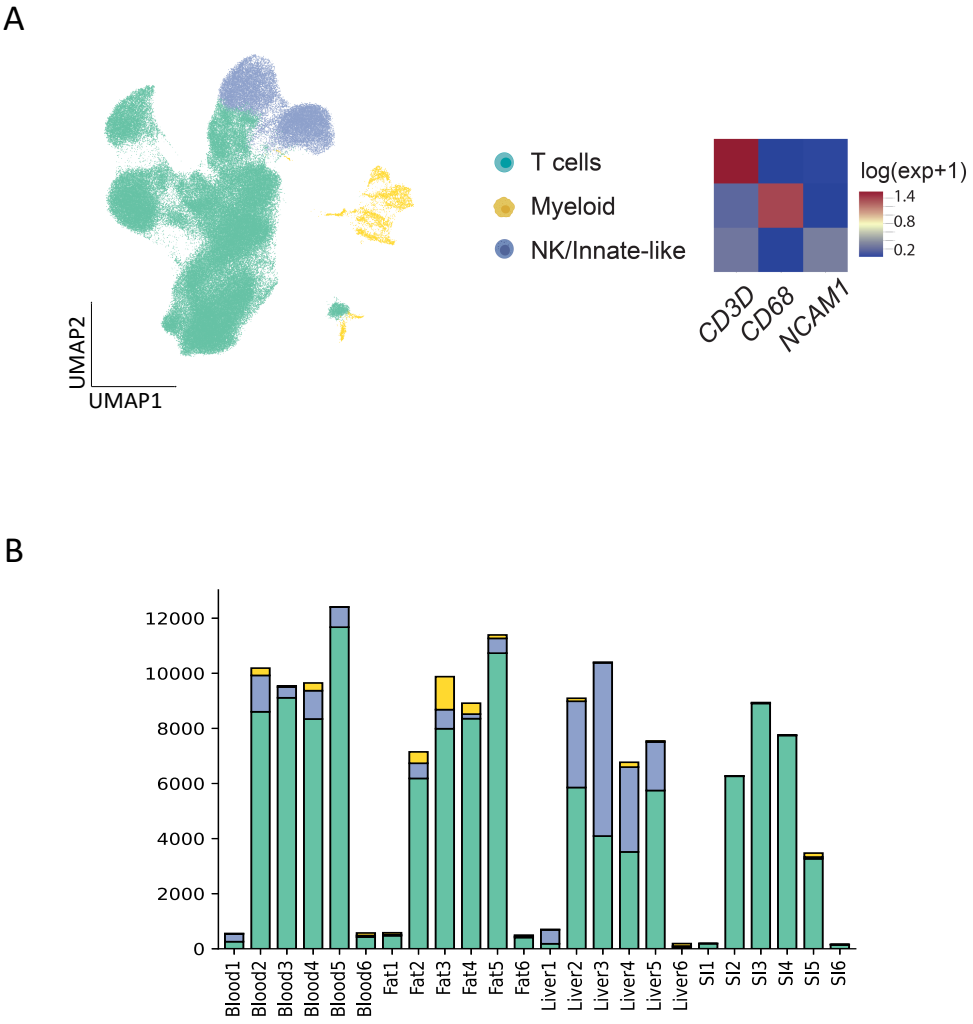



**Figure 3: Overview of global cell lineages.** A: UMAP of initial clustering revealing three global cell lineages: T cells, Myeloid cells and NK/Innate cells and the Heatmap of the mean expression of cell type markers. B: Cell numbers of each global cell lineage in different samples

### 3.1.1 T cell subsets across tissues

To explore more details of T cell adaptation across tissues, sub-clustering was performed on all annotated T cells. By using a list of well-recognized markers, we were able to annotate these clusters with the corresponding T cell sub-types. As a result, the analysis identified 16 cell sub-types within the T cell group (114,660 cells, Figure 4), including CD4+ T cells (Naive CD4 T cells, Intermediate CD4 T cells, Memory CD4 T cells, CD4 Trm cells, Th17 cells and FOXP3+ Tregs), CD8+ T cells (Naive CD8 T cells, CD8 Temra/Teff cells, CD8 Tem cells, CD8 Trm cells, MAIT cells and CD8 NKT cells), three mixed (CD4+ CD8+) cell groups (MT+ T cells, HSP+ T cells, TNF+ T cells), and gamma delta T cells (gd T cells).

Naive CD4 T cells, along with Naive CD8 T cells, prominently expressed *CCR7*, *SELL*, and *KLF2*. Memory CD4 T cells mainly composed of central memory cells displayed markers of both naive and memory T cells, including *LTB* and *TXNIP*. Through the high expression of *CD69*, *CXCR4*, and *PRDM1*, CD4 resident T cells were identified. Interestingly, the analysis recognized intermediate CD4 T cells as the transitional cells between naive cells and resident memory cells as they exhibited markers related to both categories. Th17 cells were characterized by expressing chemokine ligand *CCL20* and chemokine receptor *CCR6* as well as resident markers of CD4 Trm cells. Additionally, FOXP3+ regulatory T cells were identified by the elevated expression level of *FOXP3* and *CTLA4*. Within the CD8+ cell clusters, CD8 effector memory cells were detected by expressing granzyme K (*GZMK*), granzyme H (*GZMH*), and chemokine ligand *CCL4*. CD8 terminally differentiated effector memory (CD8 Temra/Teff ) cells were annotated due to their high expression of cytotoxic or exhausted-related markers like *FGFBP2*, *CX3CR1*, and *PRF1*. CD8 resident memory cells were characterized based on the high expression of *CD69*, *CD160*, and *ITGAE*. CD8 Mucosal-associated invariant T cells (MAIT) cells were recognized, which were noted for their high expression of *RORA*, *TRAV1-2*, and *KLRB1*. The analysis also identified CD8 NKT cells, which expressed natural killer cell markers such as *NKG7* and *FCGR3A*. Besides, three mixed (CD4+ CD8+) cell clusters were revealed in the study, representing groups of cells that do not predominantly express just one of the lineage markers, CD4 or CD8. They included MT+ T cells, which were characterized by the markers of metallothionein (e.g. *MT2A* and *MT1X*), HSP+ T cells expressing markers of heat shock protein family (e.g. *HSPA1A* and *HSPA1B*) and TNF+ T cells marked by high expression of tumor necrosis factor (*TNF*). Finally, gamma delta T cells identified by the expression

of *TRDC* were cells with gamma delta T-cell receptors on their surface.

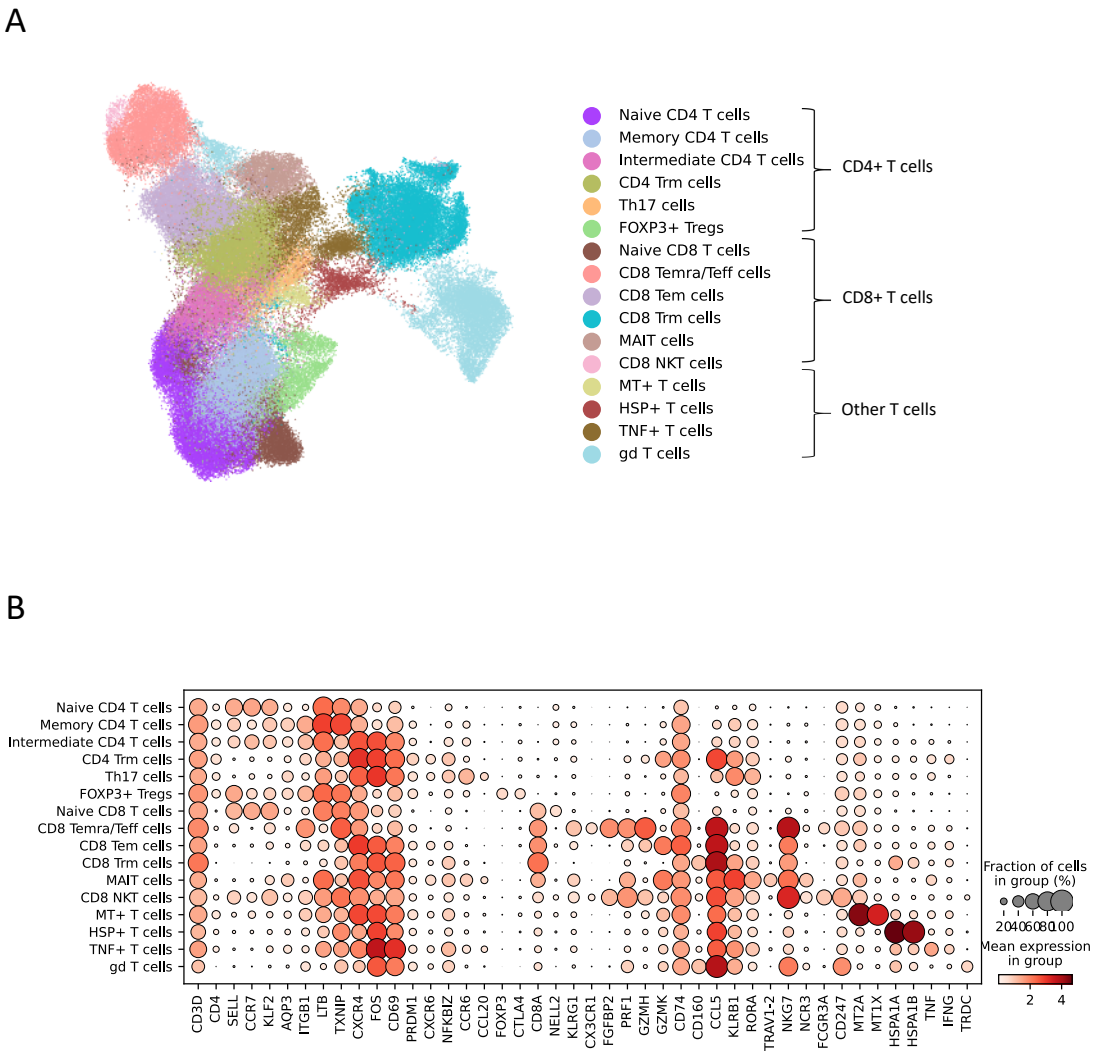



**Figure 4: Cluster annotation of T cell subsets.** A: UMAP showing 16 T cell sub-types including 6 CD4+ T cells, 6 CD8+ T cells, three mixed T cell groups, and gd T cells. B: Heatmap of the core markers for T cell sub-type annotation.

The cellular composition of T cell sub-types between tissues showed unequal distribution (Figure 5) and the patterns were similar for all individuals, despite their broad age and body weight ranges. In general, liver and VAT samples exhibited similar compositions of T cell types while blood and small intestine samples demonstrated distinct compositions from others. Specifically, the liver and VAT samples showed comparable cell type compositions in terms of various CD4 T cells, though the difference was observed in MAIT cells and TNF+ T cells with the preference for liver over

VAT. The majority of naive T cells including naive CD4 T cells and naive CD8 T cells were found in blood samples. Interestingly, a small proportion of these naive T cells also proved their presence in the tissues (liver & VAT) within at least three donors. Memory CD4 T cells were predominantly identified in blood as well, as they were probably central memory T cells that constantly circulate through the bloodstream. The T cell populations within the small intestine showed a distinct cellular composition. It contained very few CD4+ T cells but was enriched in CD8 Trm cells and gamma delta T cells, which was reasonably due to the intestine's frequent exposure to antigens. Overall, these results demonstrated the adaptation of T cells by tissue-specific compositions of T cell sub-types, indicating that different T cell sub-types may fulfill different roles within their respective resident tissues.

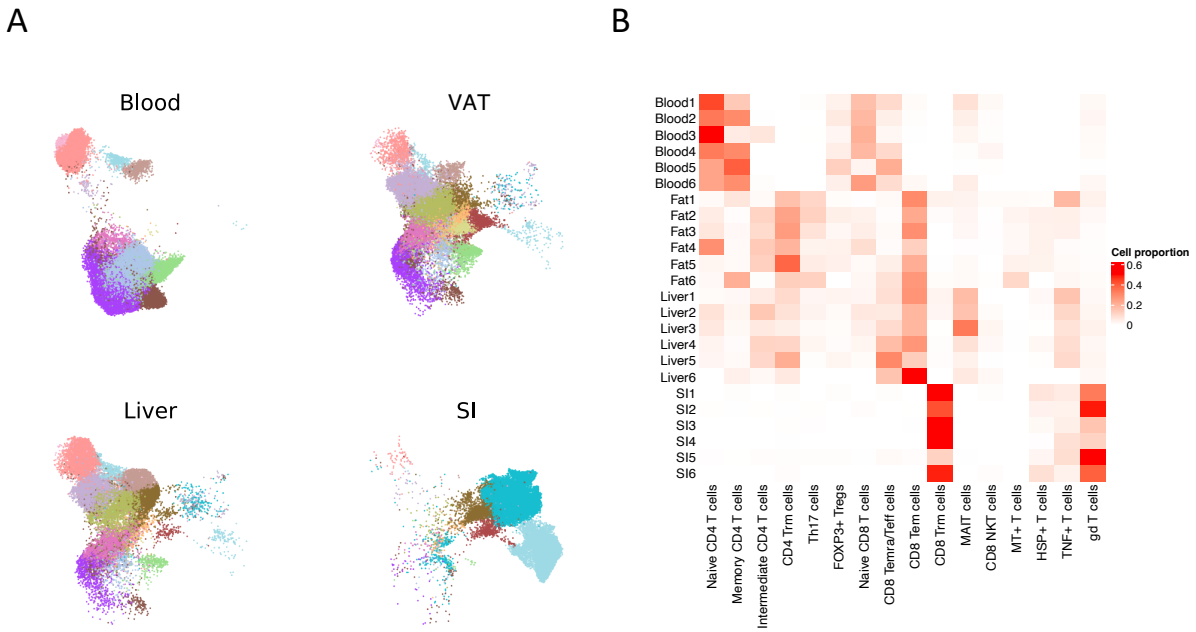



**Figure 5: Distribution of T cell sub-types in blood, VAT, liver, and SI.** A: UMAPs showing T cell cluster annotations in each tissue: blood, VAT, liver, and SI. B: Heatmap of cell type proportions in different samples.

### 3.1.2 T cell repertoires across tissues

T cell receptor (TCR) repertoire analysis was conducted using the paired single-cell TCR sequencing data. As expected, more than 80 percent of all T cells recovered from scRNA-seq data have at least one chain of alpha-beta TCR sequenced, including all

the CD4+ and CD8+ T cell clusters as well as the three mixed (CD4+ CD8+) T cell groups (Figure 6). Notably, within the gamma delta T cell population, a small percentage of cells were found to contain the alpha-beta TCR, which probably resulted from the co-clustering of CD8 cytotoxic T cells with gd T cells.

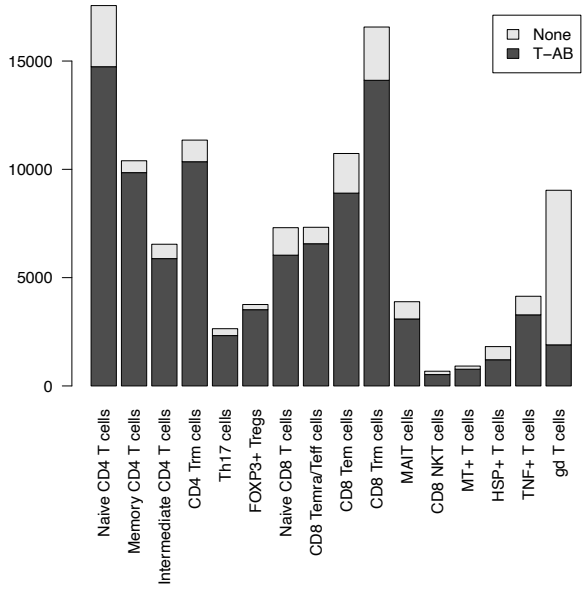



**Figure 6: The presence of alpha-beta TCR in each T cell group**

The analysis of clonal expansion among various T cell subsets revealed differences across tissues, as shown in Figure 7. It's noticed that most of the clonally expanded T cells were memory T cells, especially the CD8+ memory cells such as CD8 Temra/Teff cells, CD8 Tem cells, and CD8 Trm cells, and the others were CD4 Trm cells. Compared to blood where only the CD8 Temra/Teff cells displayed high expansion, cells in the other three tissues were overall clonally more expanded. This likely reflected localized inflammatory responses at these tissue sites. Despite the large differences in clonal expansion observed between different T cell subsets, it's difficult to conclude comparative results across tissues, since this could mainly be ascribed to the varied distribution of cell types.

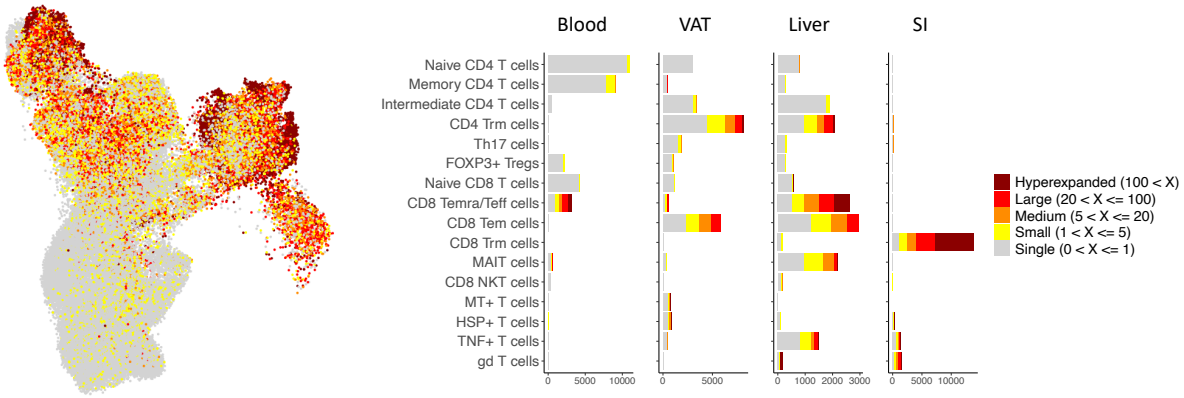



**Figure 7: TCR clonal expansion.** The clonal expansion of T cell receptors was classified into different level of groups: Hyperexpanded (100 < X), Large (20 < X <= 100), Medium (5 < X <= 20), Small (1 < X <= 5), and Single (0 < X <= 1). UMAP with the clonal expansion groups was shown (left) and the stacked barplot (right) presented the number of cells for each tissue in every T cell sub-type.

Next, the clonal overlap between tissues was examined for clonally expanded CD4+ and CD8+ T cells respectively (Figure 8). Interestingly, significant overlaps were revealed between blood and liver or VAT as well as between liver and VAT, for both expanded CD4+ and CD8+ T cell clonotypes. Moreover, CD8+ T cells showed slightly more overlap than CD4+ T cells, indicating that CD4+ T cell clones are more localized and site-specific compared to CD8+ T cells. By contrast, T cells from the small intestine showed negligible clonal overlap with other tissues, even for its large amount of CD8+ T cells, which suggested its highest tissue specificity.

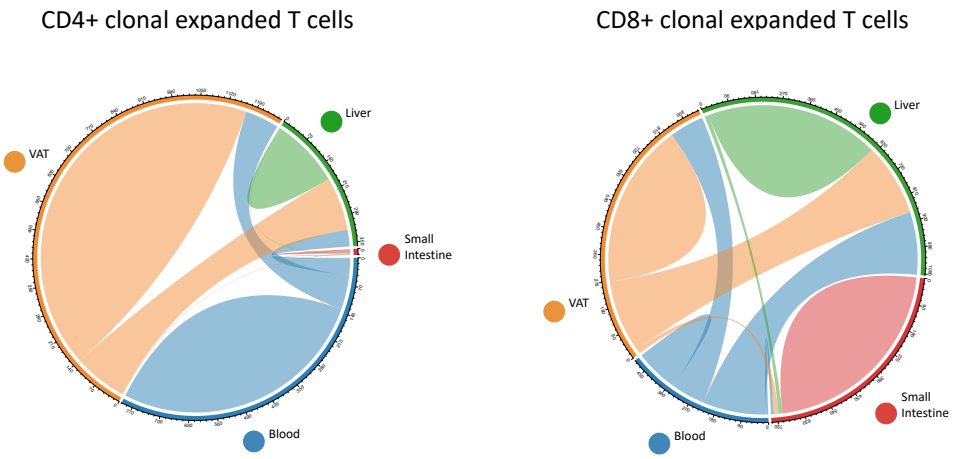



**Figure 8: TCR overlap between tissues.** The clonotypes from clonally expanded CD4+ T cells (left) and CD8+ T cells (right) were compared for overlapping between tissues. The connected path within the circle plot represents shared clonotypes and the width of the connection indicates the number of shared clones.

## 3.2 Tissue adaptation of CD4+ T cells

### 3.2.1 Tissue adaptation signatures of CD4+ T cells

The discovery of T cell clonal overlap between tissues led to the idea of comparing the gene expression profiles between these overlapping T cells. These cells, sharing the same clonotype, are believed to stem from a common precursor. Upon encountering antigens, they settle in specific tissues, where they adjust their functional behavior and gene expression profiles to align with the local tissue environment. Therefore, by analyzing these changes in overlapping T cells, we could uncover tissue adaptation at the transcription level, regardless of their T cell subset information.

Differential gene expression analysis was performed using these clonal overlapping CD4+ T cells (Figure 9.A). VAT/liver tissue adaptation signatures were defined by the significant upregulated genes comparing the gene expression profiles of cells from VAT/liver to Blood, which share the same clonotypes, respectively. The common adaptation signatures for both tissues were then determined through the intersection of the VAT and liver adaptation signatures, as shown in Table 3. To discern tissue-specific adaptation signatures, DE was conducted on clonal overlapping cells between liver and VAT. Liver-specific adaptation signatures were then delineated by

intersecting the upregulated genes identified in the comparison of liver to VAT with the liver adaptation signatures derived from the comparison of liver to blood. Similarly, VAT-specific adaptation signatures were obtained (Figure 9.B and Table 4). As a result, this analysis revealed 306 VAT tissue adaptation signatures and 205 liver tissue adaptation signatures, suggesting a more pronounced adaptation of CD4+ T cells in VAT than liver. Remarkably, around 170 genes were common, showing a substantial overlap in resident signatures shared by liver and VAT. Further, pathway analysis of these common adaptation signatures highlighted the activation of pathways related to interleukins, specifically IL4 and IL13, as well as Toll-like receptor-related pathways (Figure 9.C).

In detail, CD4+ T cells from the liver and VAT exhibited downregulated expression of genes linked to naive T cells and quiescence, such as *SELL*, *KLF2*, and *MAL*, when compared to blood. Conversely, *FOS* and *FOSB* were upregulated, suggesting enhanced cell proliferation and differentiation in tissue sites. The high expression of *DUSP1* in both tissue sites was noted which is related to cellular response to environmental stress. The chemokine receptor *CXCR4*, associated with tissue homing, was also identified as a common tissue adaptation signature in CD4+ T cells. Genes implicated in cytotoxicity (*SRGN*, *IFNG*) were more highly expressed than those in Blood. Among them, *SRGN* expression was higher in VAT than in liver, while *IFNG* showed the opposite pattern. Furthermore, *SRGN*, previously recognized as a signature for skin Trm cells [33], underscores its relevance in tissue-specific contexts. These findings unveiled tissue-common signatures associated with cell differentiation, residency, and localization of CD4+ T cells in both liver and VAT.

Based on the comparison between liver and VAT in CD4+ T cells, tissue-specific adaptation genes were identified and the tissue-common signatures could be further compared. CD4+ T cells from VAT highly expressed the transcription factor *BATF* (Basic Leucine Zipper ATF-Like Transcription Factor), which is required for the development of FOXP3+ regulatory T cells [49]. Given that *BATF* has been demonstrated to help reduce lipid accumulation caused by the high-fat diet in hepatocytes [191], it might play a role in lipid metabolism in adipose tissue. The cytotoxic marker *PRDM1*, encoding transcription factor Blimp1, was expressed highly in VAT CD4+ T cells, associated with tissue residency. Genes encoding members from the *ARID* protein family (*ARID5A*, *ARID5B*), critical in development, tissue-specific gene expression, and cell growth regulation [192, 193], were found to be expressed at higher

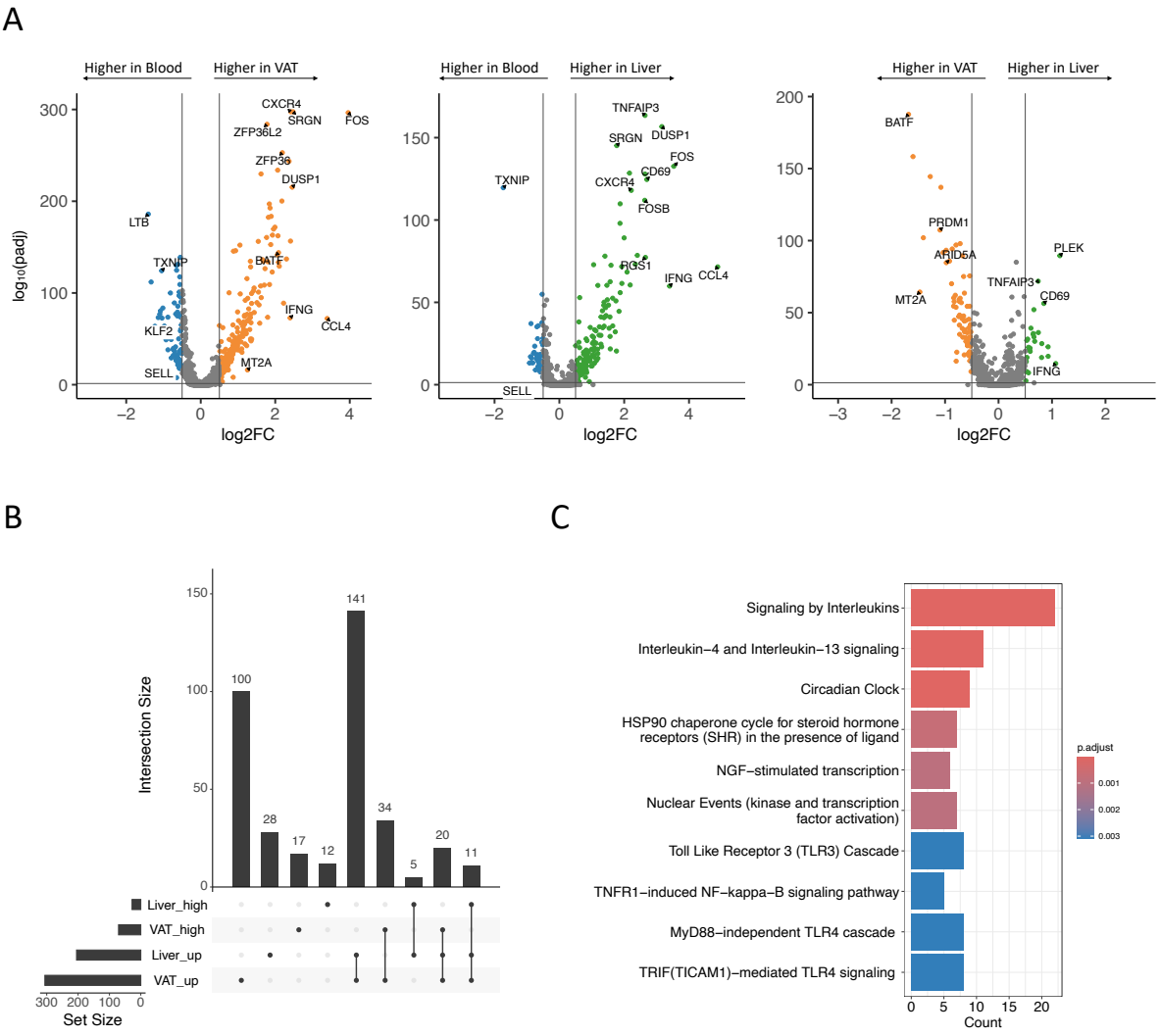



**Figure 9: Adaptation of CD4+ T cells by shared clonotypes.** All the comparisons here were based on the cells with shared clonotypes. A: Volcano plots of the DGE results for VAT vs blood (left), liver vs blood (middle), and liver vs VAT (right). B: Upset plot showing the intersection of different DEG sets: "Liver_high" represents the upregulated genes comparing liver to VAT, "Liver_up" represents the upregulated genes comparing liver to blood, and similar for "VAT_high" and "VAT_up". C: Top 10 enriched pathways of the tissue-common adaptation genes using the Reactome database.

**Table 3:** Tissue common adaptation signatures of CD4+ T cells

| Common between VAT and Liver | | | | VAT higher | Liver higher |
|---|---|---|---|---|---|
| CXCR4 | FAM177A1 | KDM6B | TSPYL2 | SRGN | DUSP1 |
| FOS | IER2 | BAZ1A | P2RY10 | ZFP36 | NR4A2 |
| ZFP36L2 | NAMPT | SLA | RAB8B | CREM | CD69 |
| SLC2A3 | HSP90AA1 | RBM8A | ARF4 | CDKN1A | TNFAIP3 |
| JUNB | EIF1 | ARAP2 | BCL3 | SOCS1 | RGS1 |
| KLF6 | SH2D2A | PTPN22 | CKS2 | LMNA | DUSP2 |
| TSC22D3 | SELENOK | ZNF831 | CHD1 | ALOX5AP | IFNG |
| BTG2 | DUSP4 | SRSF2 | TRAF1 | DDIT4 | CCL4 |
| MCL1 | REL | HIF1A | BHLHE40 | ANXA1 | NR4A1 |
| CST7 | MAP3K8 | KLRB1 | IFRD1 | PHLDA1 | JUN |
| CYTIP | ADGRE5 | HSPA1B | TAGAP | RNF19A | CD40LG |
| PTGER4 | PIK3R1 | STAT4 | CD83 | EZR | |
| BTG1 | SYTL3 | CD2 | GZMK | XCL1 | |
| UBC | MYADM | BTG3 | LINC01871 | BIRC3 | |
| ZNF331 | GZMH | SYAP1 | BCAS2 | AREG | |
| SAMSN1 | EIF4A1 | CDKN1B | IVNS1ABP | CRYBG1 | |
| TENT5C | PDCD1 | IDS | BCL2A1 | SDCBP | |
| PPP1R15A | DDX3X | VPS37B | TIPARP | SGK1 | |
| NFKBIA | RHOH | RAB11FIP1 | SLAMF1 | GPR183 | |
| FOSB | ID2 | PER1 | NFKB2 | CBLB | |
| EML4 | CCNH | EIF4A3 | HNRNPLL | | |
| FRMD4B | JUND | KPNA2 | | | |
| RGCC | YPEL5 | NAA50 | | | |
| CSRNP1 | CYCS | SLC38A2 | | | |
| PDE4B | XCL2 | WIPF1 | | | |
| IRF1 | PTPRC | PPP1R10 | | | |
| FTH1 | ICOS | CD96 | | | |
| SAT1 | GNG2 | STK17B | | | |
| SRSF7 | CLEC2B | MIR155HG | | | |
| CCL5 | NR3C1 | ZFP36L1 | | | |
| BRD2 | USP36 | AC020916.1 | | | |
| CXCR6 | PARP8 | NR4A3 | | | |
| DNAJB1 | GADD45B | CD53 | | | |
| SERTAD1 | TUBB4B | CD28 | | | |
| IFNGR1 | TUBA4A | EVI2A | | | |
| GZMA | ZC3H12A | IL18RAP | | | |
| ELF1 | PPP1CB | GPR65 | | | |
| DNAJA1 | RORA | ETV3 | | | |
| TNF | CITED2 | AMD1 | | | |
| NFKBIZ | NFE2L2 | DYNLT1 | | | |

**Table 4:** Tissue-specific adaptation signatures of CD4+ T cells

| VAT specific | Liver specific |
|---|---|
| H3F3B | NKG7 |
| BATF | PLEK |
| PRDM1 | CTSC |
| LEPROTL1 | FCRL6 |
| PFKFB3 | P2RX5 |
| ARID5B | |
| MTHFD2 | |
| ARID5A | |
| PLIN2 | |
| CD44 | |
| TPM4 | |
| FOSL2 | |
| GLIPR1 | |
| CLU | |
| CNOT6L | |
| TGIF1 | |
| CD99 | |
| ZNF267 | |
| GPCPD1 | |
| LMNB1 | |
| STAT3 | |
| SNX9 | |
| NIBAN1 | |
| FKBP5 | |
| RSBN1 | |
| ATP2B1 | |
| BACH2 | |
| LGALS1 | |
| MT2A | |
| GLUD1 | |
| CAPG | |
| SOCS3 | |
| PRKX | |
| MT1X | |

levels in VAT, hinting at their connection to adipose tissue homeostasis. VAT-specific signature *MT2A* was shown positively correlated with obesity [58], though further research is needed to fully unravel its functions and roles within the adipose tissue environment. Other tissue resident genes of CD4+ T cells such as *TNFAIP3*, *CD69*, and *IFNG* were expressed in both liver and VAT, but exhibited slightly higher expression in liver. Among the liver-specific signatures, only five were identified, including pleckstrin (*PLEK*). *PLEK* was implicated in the regulation of fatty acid and cholesterol metabolism [194], implying its integral role in the liver's metabolic processes and its potential impact on overall hepatic function.

Additionally, the top significant DE genes from the cells with shared clones were also explored in total CD4+ T cells (Figure 10). Their expression patterns between tissues were consistent with the results from those shared clones, indicating the broad application and high representation of these identified tissue adaptation signatures in the total CD4+ T cell population.

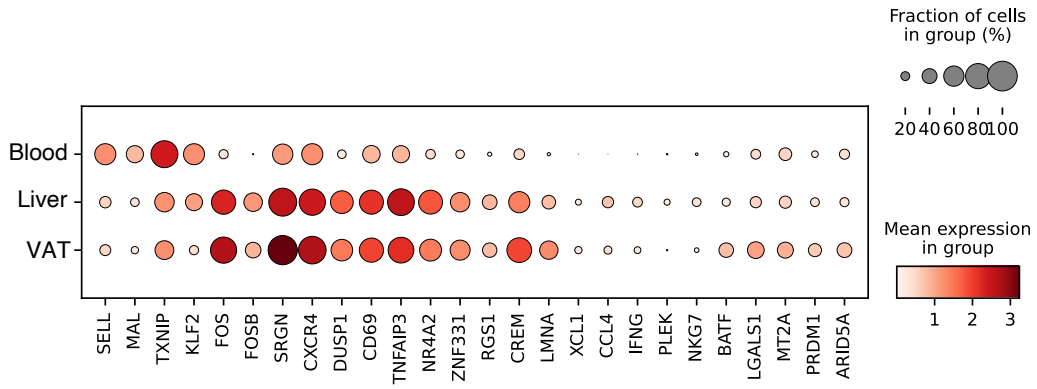



**Figure 10: Tissue adaptation signatures in all CD4+ T cells.**

Tissue-resident memory T cells, known for their non-recirculating nature and long-term residency in tissues, are considered well-adapted to their local environment. To delve deeper into the tissue-specific signatures of CD4+ Trm cells, differential gene expression (DGE) analysis was conducted on these cells in liver and VAT compared to all other CD4+ T cell subsets, respectively. This analysis identified a greater number of upregulated differentially expressed genes in VAT Trm cells (308) compared to liver Trm cells (190), with 120 gene signatures shared between the two (Figure 11.A-C, Table 5 and 6), including chemokine ligand *CCL4* and *CCL5*, granzyme *GZMA* and granzyme *GZMK*. *IFNG* and *CD69*, consistent with previous findings from the shared clone analysis, were highly expressed in these cells, particularly in the liver. As for the tissue-specific Trm signatures, most of the previously mentioned genes were reaffirmed in this context. To our interests, the transcriptional factor *GATA3* was identified as liver-specific, known for its diverse roles in immune regulation. However, its specific function in liver CD4+ T cells remains less clearly defined. Overall, the lower number of shared tissue signatures and the higher number of tissue-specific signatures in CD4+ Trm cells demonstrated the strong tissue residence of these cells. These observations also corroborated the findings from analysis on shared clones.

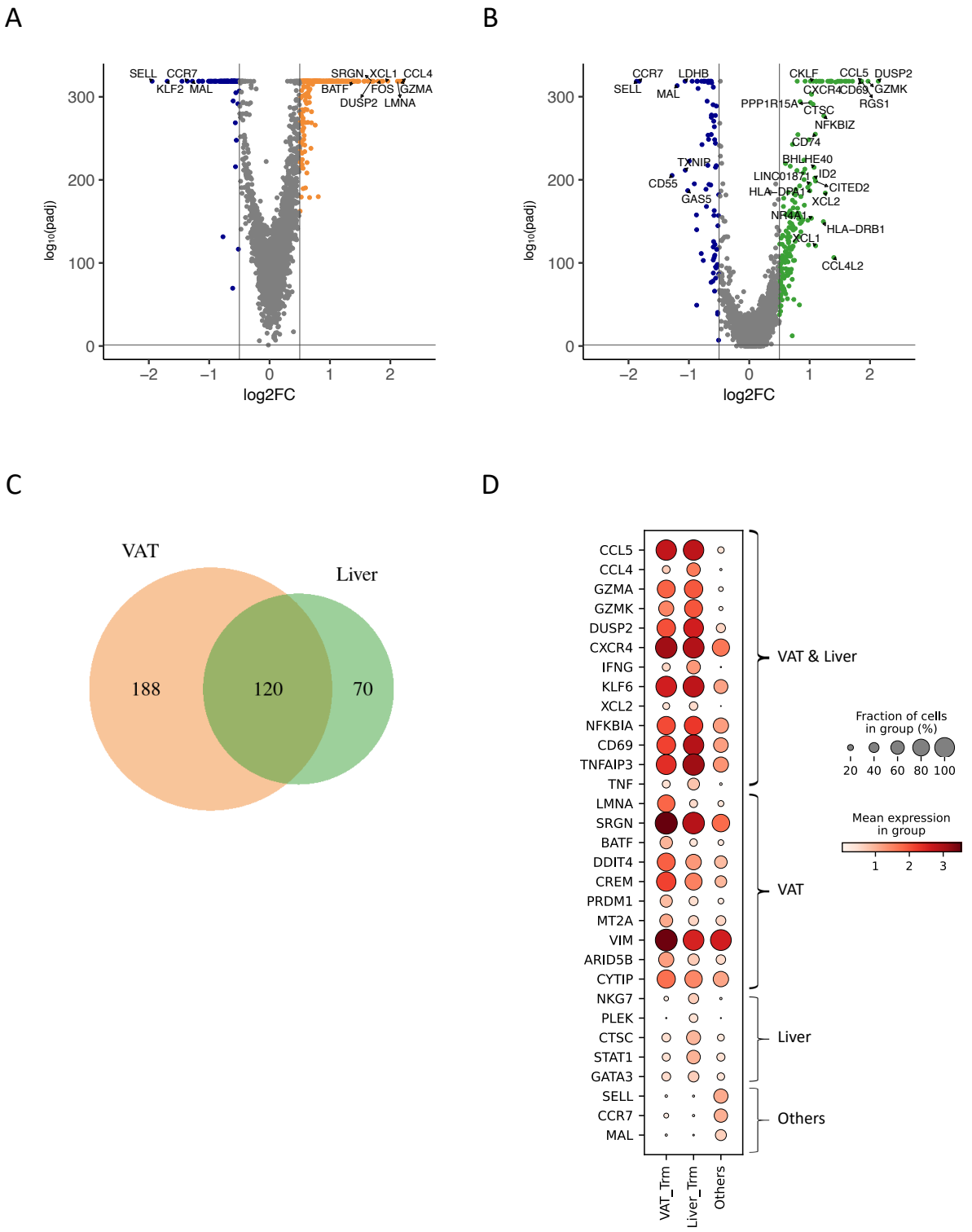



**Figure 11: Tissue signatures of CD4 Trm cells.** A-B: Volcano plots showing the DGE results of CD4+ Trm cells in VAT (A) and liver (B) compared with all other CD4+ T cells in the data. C: Venn diagram of the VAT resident genes and liver resident genes. D: Dotplot of key markers of tissue-resident signatures

**Table 5:** Common tissue resident signatures of CD4+ Trm cells

| Common between VAT and Liver | | |
|---|---|---|
| ACTN4 | HOPX | SLC2A3 |
| ADGRE5 | ID2 | SRSF7 |
| ALOX5AP | IDS | STOM |
| ARAP2 | IER2 | SYTL3 |
| BHLHE40 | IFNG | TENT5C |
| BTG2 | IFNGR1 | TMSB4X |
| CCL4 | IL32 | TNF |
| CCL5 | IRF1 | TNFAIP3 |
| CCNH | ITGB2 | TUBA4A |
| CD2 | JUN | USP36 |
| CD53 | JUNB | WIPF1 |
| CD69 | JUND | XCL1 |
| CD96 | KLF6 | XCL2 |
| CLEC2B | KLRB1 | ZC3H12A |
| CLIC1 | LCP1 | ZFP36 |
| CSRNP1 | LINC01871 | ZFP36L2 |
| CST7 | MAP3K8 | ZNF331 |
| CTSW | MCL1 | CD74 |
| CXCR4 | MYADM | RORA |
| CXCR6 | NAA50 | CKLF |
| CYTOR | NFKBIA | ATP1B3 |
| DNAJA1 | NFKBIZ | OASL |
| DNAJB1 | NR4A2 | NEDD9 |
| DUSP1 | PARP8 | APMAP |
| DUSP2 | PDE4B | CITED2 |
| DUSP4 | PHLDA1 | SH2D1A |
| EML4 | PLAAT4 | ZNF831 |
| F2R | PPP1R15A | CXCR3 |
| FOS | PRF1 | IKZF3 |
| FOSB | PTGER4 | ERN1 |
| FRMD4B | PTPN22 | NR4A1 |
| GNG2 | PTPRC | HNRNPLL |
| GZMA | REL | TAGAP |
| GZMH | RGCC | RAB8B |
| GZMK | RGS1 | ICOS |
| GZMM | RHOC | GBP5 |
| HCST | SAMSN1 | HLA-DRB1 |
| HLA-A | SAT1 | PNP |
| HLA-B | SERTAD1 | ATF3 |
| HLA-C | SH2D2A | CD40LG |

**Table 6:** Tissue-specific resident signatures of CD4+ Trm cells

| VAT specific | | | | Liver specific | |
|---|---|---|---|---|---|
| ACTB | GAPDH | PLIN2 | TUBB4B | NKG7 | TAPBP |
| ADAM19 | GLIPR1 | PLP2 | TXN | PLEK | CORO1A |
| ADSS | GLUD1 | PPP1CB | UBB | CTSC | MIR4435-2HG |
| AHNAK | GPCPD1 | PPP1R16B | UBC | FCRL6 | ARHGAP18 |
| AKAP13 | H3F3B | PRDM1 | VIM | HLA-DPA1 | RASAL3 |
| ANKRD28 | HERPUD1 | PRKX | YWHAZ | P2RX5 | DUSP5 |
| ANXA1 | HMGB2 | PRNP | ZNF267 | RASGEF1B | CD6 |
| ANXA2 | HNRNPUL1 | RAB11FIP1 | EIF4A3 | PSME2 | MYBL1 |
| ARHGDIB | HSP90AA1 | RGS2 | HSPA1B | BCL2A1 | CTSD |
| ARID5A | HSPA5 | RNF149 | TUBA1C | SLAMF7 | SIT1 |
| ARID5B | HSPD1 | RNF19A | KIF21A | RNF213 | AC004687.1 |
| ATP1B1 | HSPE1 | RSBN1 | JPT1 | IDH2 | RAB1B |
| ATP2B1 | ICAM1 | RUNX3 | ELL2 | PSMB9 | TAP2 |
| BATF | IFI16 | S100A10 | AMD1 | GPR65 | PHTF2 |
| BIRC3 | IFITM2 | S100A11 | GPR183 | RBL2 | CAP1 |
| BRD2 | IL2RB | S100A4 | ATP2B4 | APOBEC3G | GATA3 |
| BTG1 | IL7R | S100A6 | CAPG | LINC00892 | CD83 |
| BTG3 | IQGAP1 | SDCBP | LUZP1 | GYG1 | PYHIN1 |
| CALM2 | ITM2A | SELENOK | MATK | RBPJ | CALR |
| CBLB | JMJD1C | SGK1 | VPS37B | SLAMF1 | PMAIP1 |
| CD44 | KDM6B | SH3BGRL3 | IVNS1ABP | IL2RG | FERMT3 |
| CD63 | KMT2E | SLA | IFRD1 | HLA-F | OPTN |
| CD99 | LDHA | SLC38A2 | AKNA | LCK | TRBV5-1 |
| CDKN1A | LGALS1 | SLC7A5 | CYCS | THEMIS | |
| CEMIP2 | LITAF | SMIM3 | CDK17 | PSME1 | |
| CFL1 | LMNA | SOCS1 | GPRIN3 | TAP1 | |
| CLU | LMNB1 | SQSTM1 | LDLR | TANK | |
| CREM | LRRFIP1 | SRGN | AC007384.1 | LBH | |
| CRIP1 | LSP1 | SRSF2 | NEU1 | CHST12 | |
| CRYBG1 | MAF | STAT4 | FBXO34 | STAT1 | |
| CYBA | MAP1LC3B | STX11 | INSIG1 | IFNG-AS1 | |
| CYTIP | MSN | SYAP1 | DNAJC1 | ETV3 | |
| DDIT4 | MT2A | SYTL2 | GABARAPL1 | CCDC107 | |
| DDX3X | MTHFD2 | TAGLN2 | KPNA2 | ITGAL | |
| DENND4A | MYH9 | TC2N | MIR155HG | SASH3 | |
| DYNLT3 | MYL6 | TERF2IP | AHR | PPP1R18 | |
| EIF4A1 | NAMPT | TGFB1 | MPZL3 | CCL4L2 | |
| EIF5 | NIBAN1 | TGIF1 | TSPYL2 | GPR174 | |
| ELF1 | NR3C1 | TIMP1 | LAG3 | LCP2 | |
| EMB | OAZ1 | TIPARP | SLC16A3 | HLA-DPB1 | |
| EZR | PBX4 | TMX4 | AC020916.1 | SLC9A3R1 | |
| FAM177A1 | PDCD1 | TNFRSF1B | TNFRSF18 | TBX21 | |
| FKBP5 | PDCL3 | TPM4 | DYNLL1 | SLF1 | |
| FOSL2 | PERP | TSC22D3 | PPP1R10 | IL12RB2 | |
| FTH1 | PFKFB3 | TUBA1A | FAS | PPP2R2B | |
| FYN | PHLDA2 | TUBA1B | HSPA1A | RAC2 | |
| GADD45B | PIK3R1 | TUBB2A | AREG | ACAA2 | |

### 3.2.2 Tissue adaptation model of CD4+ T cells

To discover when and how different sub-types of CD4+ T cells adapted to tissues (liver and VAT), sub-clustering of all CD4+ T cells was performed (Figure 12.A), and subsequent velocity and trajectory analysis was applied. Three naive CD4+ T cell clusters were identified: Naive(1) and Naive (2) cells were predominantly in blood and Naive (3) were found mainly in tissue sites. The existence of naive CD4+ T cells in tissues was supported by the experimental validation gated on CD95 and CD62L (Figure 13, performed by Jonas Wagner). Despite universal naive cell markers (SELL, CCR7, etc), several tissue-specific naive markers were identified (Figure 12.B) containing LYPD3, known for roles in cell signaling and adhesion, and GPR18, involved in the regulation of the immune system. Regarding CD4 Trm cells, Trm(1) cells were more observed in VAT while Trm(2) cells were found more in the liver. Notably, Th17 cells exhibited a higher specificity in VAT compared to the liver, which may suggest a greater likelihood of naive cells developing into Th17-like cells in VAT. FOXP3+ Regulatory CD4 T cells also showed in two clusters according to their tissue of origin: one cluster of blood (Treg (1)) and the other tissue cluster (Treg (2)).

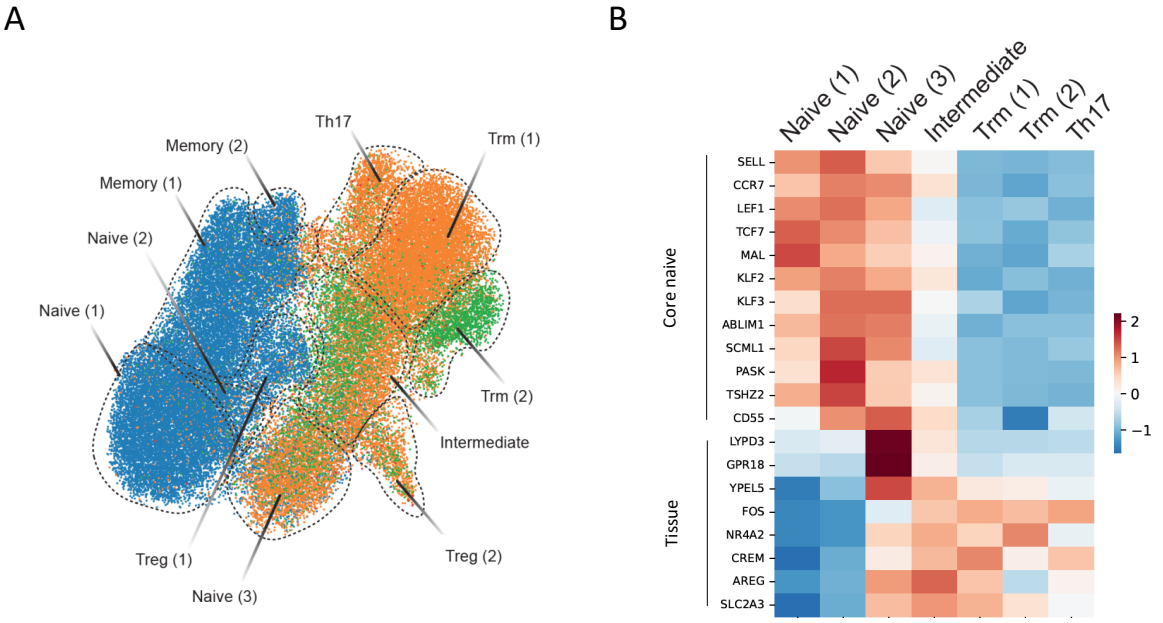



**Figure 12: Sub-clustering of CD4+ T cells.** A: UMAP plot of CD4+ T cells sub-clustering annotation. B: Heamap showing the expression of core naive and resident markers of CD4+ T cells.

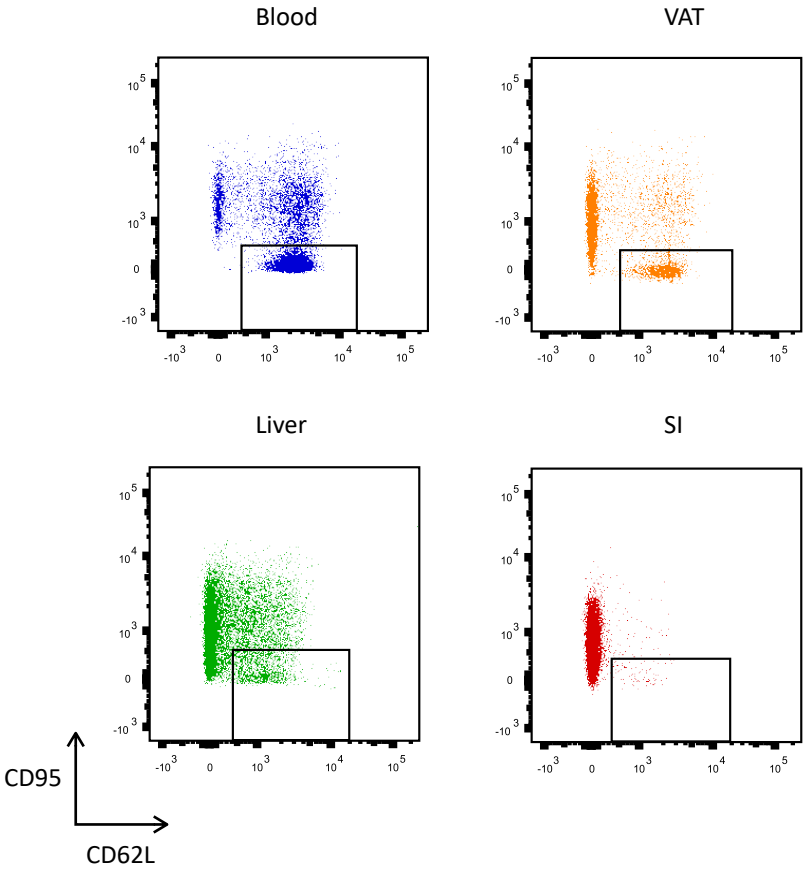



**Figure 13: CD95 and CD62L gating of naive CD4+ T cells.**

The obtained velocity vectors were projected onto UMAP (Figure 14.A), representing the direction and speed of movement of the individual cells. Additionally, the PAGA graph (Figure 14.B) was used for trajectory inference, revealing high-confidence developmental paths between groups. Consequently, we found that CD4+ T cells in blood and other tissues followed distinct developmental trajectories. In blood samples, naive CD4+ T cells from Naive (1) and Naive (2) developed into either memory cells or regulatory cells upon antigen recognition. In tissue samples, CD4+ T cells evolved from naive stages (Naive (2) and Naive (3)), potentially to intermediate cells, and ultimately differentiated into different resident memory cells or tissue regulatory cells.

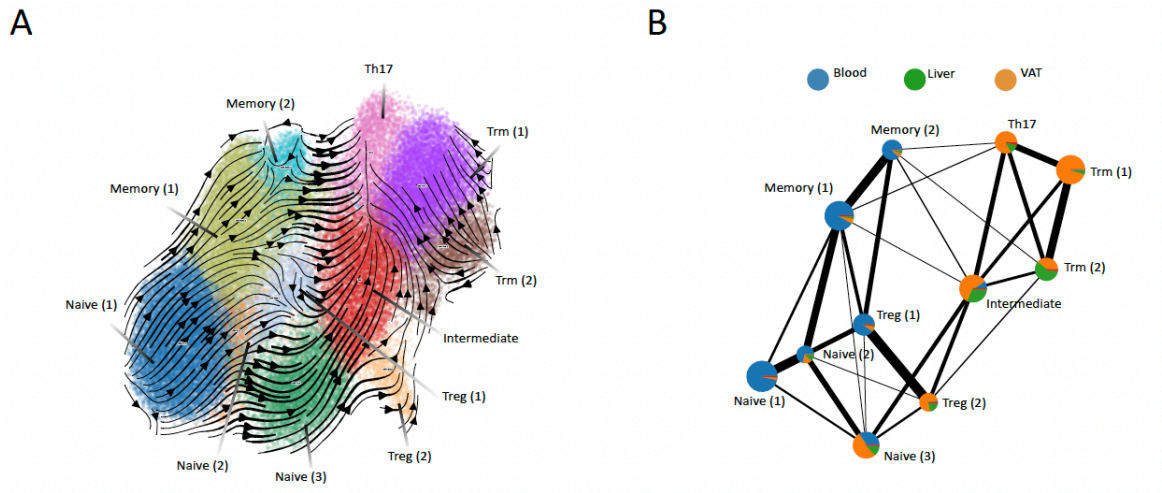



**Figure 14: Trajectory inference of CD4+ T cells.** A: UMAP plot of velocity projection. B: PAGA graph connecting different T cell groups. The weights of the edges represented the confidence of the connection between the two groups.

To delve deeper into the development of CD4+ T cells residency in the liver and VAT, their tissue-enriched clusters including Naive (3), Intermediate, Trm (1), Trm(2) and Th17 cell groups were extracted and analyzed. Velocity and trajectory analyses were then reapplied to these cells. The latent time, derived from the dynamic model of scVelo [158], combined with PAGA, facilitated the construction of a directed confidence path between these clusters (Figure 15.A and B). Cluster-specific driving genes were ordered by the latent time shown in Figure 15.C. As expected, naive marker

genes were shown to be highly expressed in the early time of the trajectory while the activation/cytotoxic markers like *CCL5* and *GZMA* exhibited increasing expression along the constructed trajectory. Next, to determine if different gene expression patterns existed along the trajectory between the liver and VAT, I compared the gene expression between these two tissues along the latent time. Figure 16 highlighted selected genes that exhibited significant differential expression between liver and VAT. It's revealed that these tissue-specific adaptation signatures began to express from the beginning of the trajectory, specifically, during the naive stage. For example, *BATF* and *SRGN*, which were identified as the tissue resident or tissue adaptation markers for VAT in this study, showed higher expression compared to the liver even before transitioning into resident memory T cells.

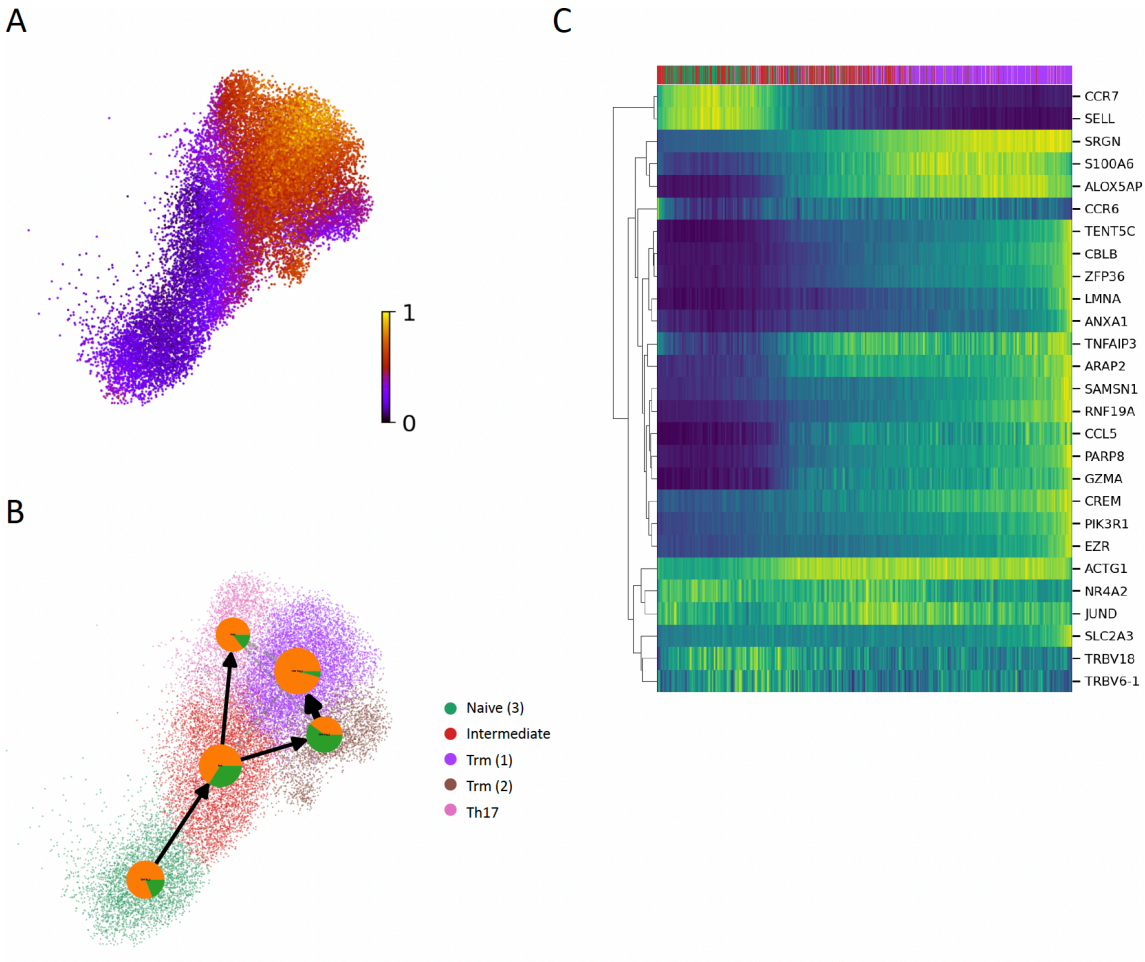



**Figure 15: Trajectory inference of tissue CD4+ T cell clusters.** A: UMAP of latent time in tissue-enriched clusters (Naive (3), Intermediate, Trm (1), Trm(2), and Th17 cells). B: Directed PAGA graph for the tissue trajectory. C: Heatmap showing top-ranked driving genes in every tissue-enriched cluster.

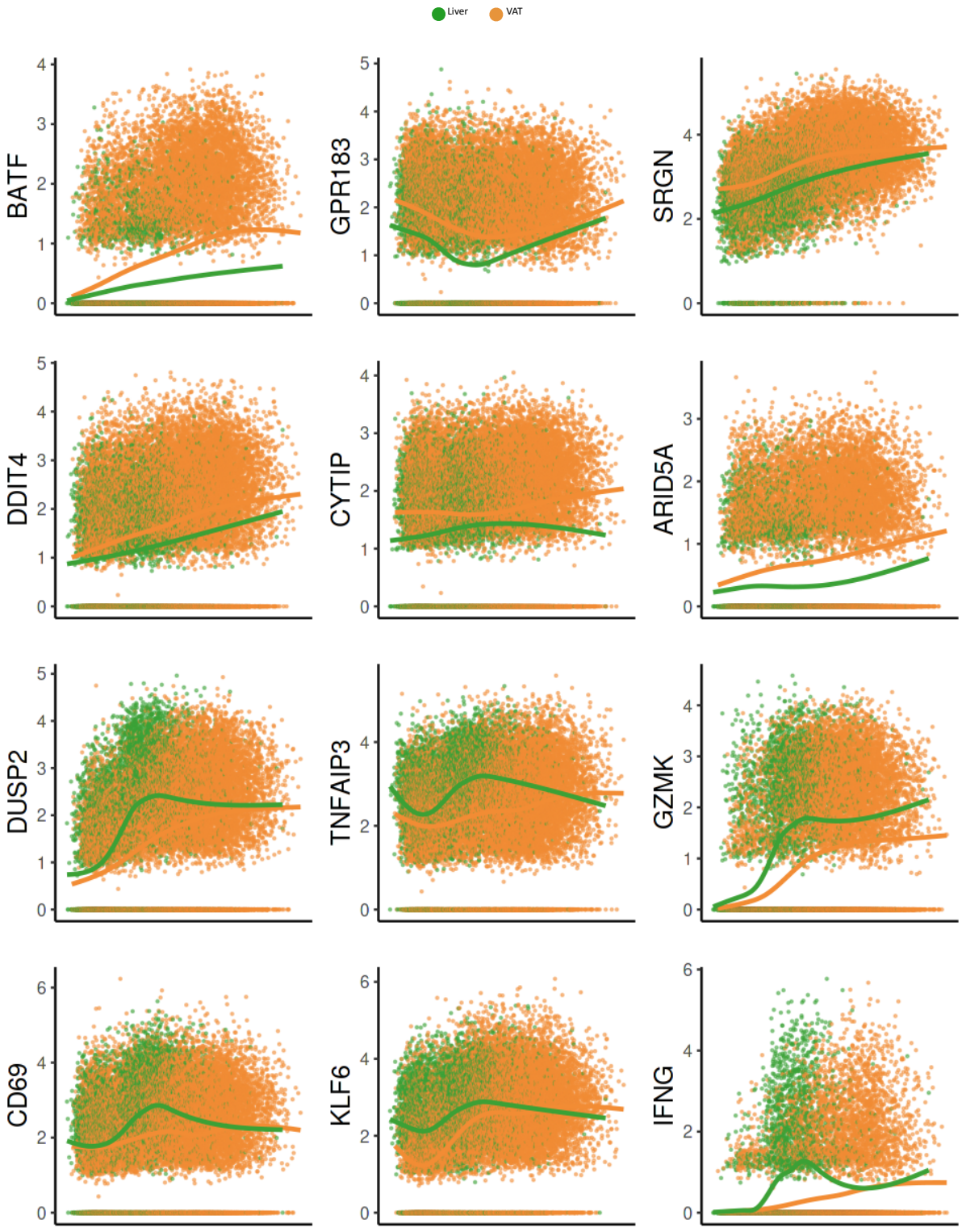



**Figure 16: Scatter plots of selected DE genes along latent time.**: The first two rows represented genes that expressed more in VAT and the others expressed more in liver.

Together with these findings, we proposed a tissue adaptation model (Figure 17). This model assumes that naive CD4+ T cells residing in tissues, begin to adapt to the tissue environment, gradually acquiring essential residency signatures, and eventually differentiate into tissue-resident memory cells, as a stable population of tissue-resident T cells. When naive CD4+ T cells migrate to tissues and encounter antigens, alongside receiving various signals from the local tissue environment, their adaptation to tissues strengthens and intensifies, represented by their elevated gene expression profiles of the corresponding tissue signatures.

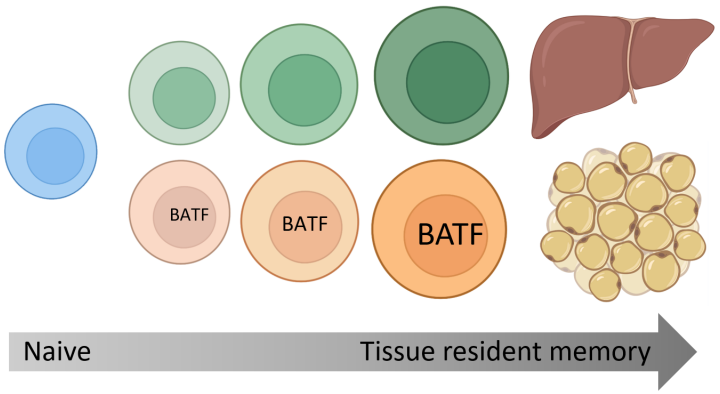



**Figure 17: Tissue adaptation model.**

Trajectory analysis has indicated that naive CD4+ T cells undergo adaptation to specific tissue sites. Moreover, the adaptation of these naive CD4+ T cells can be demonstrated through differential gene expression analysis (Figure 18 and Table 7). To ensure the focus on pure naive cells, clonally expanded cells were removed to minimize the inclusion of potential memory cells. It's found that numerous genes were significantly upregulated in tissues (liver and VAT), aligning with signatures identified in previous analyses and thereby validating the accuracy of the tissue adaptation model. Naive CD4+ T cells were traditionally perceived as homogenous by many researchers, this study sheds new light on naive CD4+ T cells, presenting them as dynamic entities capable of distinct adaptation within tissue environments.

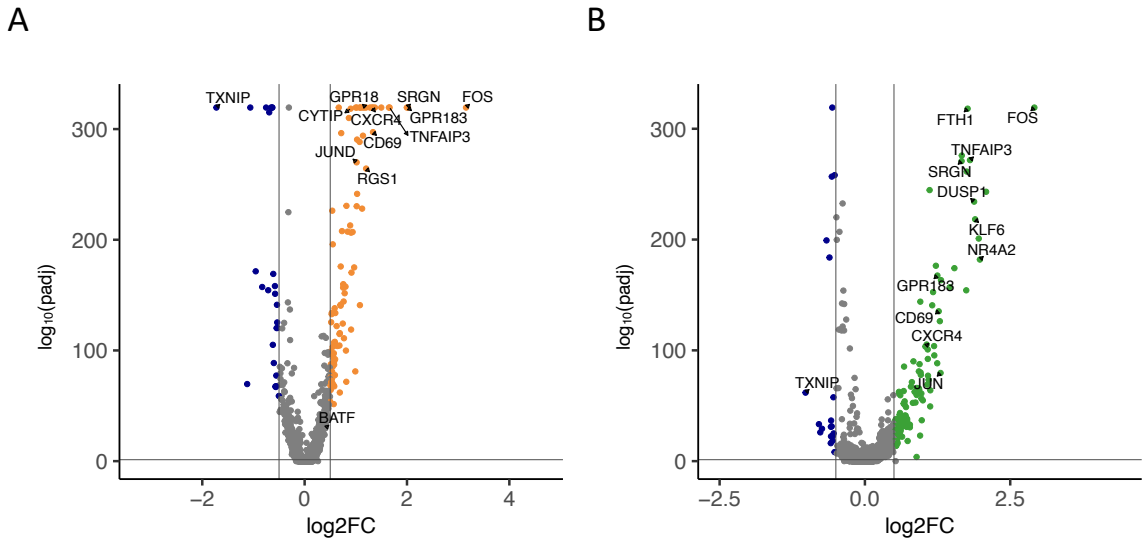



**Figure 18: Volcano plots of DGE results in naive CD4+ T cells.** A-B: Volcano plots showing the DGE results of non-expanded CD4+ naive T cells in VAT (A) and liver (B) compared with cells from blood.

**Table 7:** Upregulated DEGs in VAT naive CD4 T cells

| Gene | Log2FC | Gene | Log2FC | Gene | Log2FC |
|---|---|---|---|---|---|
| FOS | 3.15 | FAM177A1 | 0.76 | UBC | 0.53 |
| GPR183 | 2.02 | SYAP1 | 0.76 | GPCPD1 | 0.52 |
| SRGN | 2.00 | SC5D | 0.75 | CKAP2 | 0.52 |
| TNFAIP3 | 1.66 | DCXR | 0.75 | ARL4A | 0.52 |
| ZNF331 | 1.65 | ALG13 | 0.75 | MYL6 | 0.52 |
| YPEL5 | 1.50 | ATP5IF1 | 0.73 | VPS37B | 0.51 |
| LYPD3 | 1.37 | CCL4 | 0.72 | CD28 | 0.51 |
| SLC2A3 | 1.34 | ELF1 | 0.71 | P2RY10 | 0.51 |
| CD69 | 1.34 | CLEC2D | 0.70 | CCR7 | 0.51 |
| CXCR4 | 1.31 | ARL4C | 0.70 | HERPUD1 | 0.51 |
| NR4A2 | 1.28 | MT2A | 0.69 | CD2 | 0.50 |
| RGS1 | 1.20 | MYADM | 0.69 | GABARAPL1 | 0.50 |
| GPR18 | 1.20 | REL | 0.68 | | |
| CREM | 1.16 | TPM4 | 0.68 | | |
| EZR | 1.14 | BTG1 | 0.67 | | |
| FOSB | 1.13 | SYTL3 | 0.64 | | |
| GADD45B | 1.12 | TSC22D3 | 0.63 | | |
| SELENOK | 1.09 | SERTAD1 | 0.60 | | |
| AREG | 1.08 | CLEC2B | 0.60 | | |
| FTH1 | 1.08 | BIRC3 | 0.59 | | |
| PPP1R15A | 1.07 | SDCBP | 0.59 | | |
| CYCS | 1.03 | GTF2B | 0.59 | | |
| BCAS2 | 1.02 | CCL5 | 0.59 | | |
| DNAJB1 | 1.02 | GPR65 | 0.58 | | |
| RGCC | 1.02 | ICOS | 0.58 | | |
| JUND | 1.02 | HSPE1 | 0.58 | | |
| JUNB | 1.00 | SRSF3 | 0.58 | | |
| SOX4 | 0.99 | BTG2 | 0.58 | | |
| DUSP1 | 0.97 | NANS | 0.57 | | |
| PTGER4 | 0.94 | SRSF7 | 0.57 | | |
| SAT1 | 0.92 | TAGAP | 0.57 | | |
| ANXA1 | 0.91 | C1orf162 | 0.57 | | |
| PDE4B | 0.91 | SNHG15 | 0.55 | | |
| CYTIP | 0.90 | ACTB | 0.55 | | |
| EVI2A | 0.89 | RBM8A | 0.55 | | |
| RPS26 | 0.86 | SNHG12 | 0.55 | | |
| DDIT3 | 0.84 | DUSP2 | 0.54 | | |
| RHOH | 0.82 | CHMP1B | 0.54 | | |
| KLF6 | 0.81 | ARHGDIB | 0.54 | | |
| JUN | 0.81 | HSPD1 | 0.54 | | |
| HSPA5 | 0.80 | ITM2A | 0.54 | | |
| SAMSN1 | 0.78 | CTH | 0.53 | | |
| DDIT4 | 0.77 | TOB1 | 0.53 | | |

**Table 8:** Upregulated DEGs in Liver naive CD4 T cells

| Gene | Log2FC | Gene | Log2FC | Gene | Log2FC | Gene | Log2FC |
|---|---|---|---|---|---|---|---|
| FOS | 2.92 | PTGER4 | 0.94 | ARL4C | 0.64 | DDIT4 | 0.53 |
| FOSB | 2.08 | PDE4B | 0.94 | LDHA | 0.64 | CALR | 0.53 |
| NR4A2 | 1.98 | SELENOK | 0.93 | SERTAD1 | 0.64 | EMP3 | 0.53 |
| ANXA1 | 1.96 | ZFP36L2 | 0.93 | PIM2 | 0.63 | GPBP1 | 0.52 |
| KLF6 | 1.90 | TAGAP | 0.89 | ATP6V0C | 0.62 | BIRC3 | 0.52 |
| DUSP1 | 1.88 | ATP1B3 | 0.89 | BRD2 | 0.62 | NDUFV2 | 0.52 |
| TNFAIP3 | 1.81 | RPS4Y1 | 0.89 | SYTL3 | 0.61 | DYNLL1 | 0.52 |
| FTH1 | 1.77 | ARL4A | 0.87 | CXorf40A | 0.61 | TOB1 | 0.52 |
| EZR | 1.75 | ZFP36 | 0.86 | USP36 | 0.61 | PNP | 0.51 |
| ZNF331 | 1.74 | IL7R | 0.84 | IFRD1 | 0.61 | YWHAQ | 0.51 |
| DNAJB1 | 1.67 | IER2 | 0.83 | HSPD1 | 0.61 | UBE2A | 0.51 |
| SRGN | 1.66 | JUND | 0.81 | MT-ATP8 | 0.61 | KPNA2 | 0.50 |
| PPP1R15A | 1.54 | RHOH | 0.79 | TUBA1A | 0.61 | | |
| YPEL5 | 1.46 | CREM | 0.77 | H2AFZ | 0.60 | | |
| CYCS | 1.31 | GABARAPL1 | 0.77 | ALG13 | 0.60 | | |
| JUN | 1.30 | RGCC | 0.75 | HSPE1 | 0.60 | | |
| SLC2A3 | 1.29 | CHMP1B | 0.74 | IRF1 | 0.60 | | |
| CD69 | 1.27 | KDM6B | 0.73 | PDE4D | 0.60 | | |
| GPR183 | 1.25 | PPP1CB | 0.73 | PHF20 | 0.60 | | |
| GADD45B | 1.24 | EVI2A | 0.73 | RAB11A | 0.60 | | |
| TSC22D3 | 1.22 | TUBA4A | 0.72 | GATA3 | 0.59 | | |
| NR4A1 | 1.19 | SRSF2 | 0.71 | SNHG15 | 0.59 | | |
| MYADM | 1.19 | LYPD3 | 0.70 | EML4 | 0.59 | | |
| HSP90AA1 | 1.17 | SRSF3 | 0.70 | CYTIP | 0.59 | | |
| JUNB | 1.16 | ADGRE5 | 0.70 | DDX5 | 0.58 | | |
| AREG | 1.13 | TENT5C | 0.69 | EIF4A3 | 0.58 | | |
| LMNA | 1.12 | IDS | 0.69 | TPI1 | 0.58 | | |
| EIF1 | 1.11 | CD28 | 0.69 | NAMPT | 0.57 | | |
| DNAJA1 | 1.10 | SGK1 | 0.69 | CCND2 | 0.57 | | |
| DUSP2 | 1.09 | ELF1 | 0.68 | SDCBP | 0.57 | | |
| CRYBG1 | 1.08 | ICOS | 0.68 | GPR65 | 0.57 | | |
| SAT1 | 1.08 | ARF4 | 0.68 | TIPARP | 0.57 | | |
| REL | 1.07 | ICAM2 | 0.68 | FAM177A1 | 0.56 | | |
| BCAS2 | 1.06 | SYAP1 | 0.68 | BORCS5 | 0.56 | | |
| CXCR4 | 1.04 | CCL4 | 0.67 | CLIC1 | 0.56 | | |
| KLF2 | 1.00 | YWHAZ | 0.67 | HNRNPA0 | 0.55 | | |
| HSPA1A | 0.98 | SRSF7 | 0.67 | TPM4 | 0.54 | | |
| MCL1 | 0.97 | NFKB2 | 0.66 | GNG2 | 0.54 | | |
| TUBB4B | 0.97 | TSPYL2 | 0.66 | ID2 | 0.54 | | |
| CITED2 | 0.96 | SH2D2A | 0.66 | MAP1LC3B | 0.54 | | |
| CSRNP1 | 0.96 | SAMSN1 | 0.64 | CD99 | 0.53 | | |
| HSPA8 | 0.95 | CST7 | 0.64 | EIF4A1 | 0.53 | | |
| CCL5 | 0.95 | TAGLN2 | 0.64 | TRBC1 | 0.53 | | |

### 3.2.3 IL6 signaling pathway in VAT CD4+ T cells

To understand dynamic cell interactions between CD4+ T cells and other immune cells (NK cells and Myeloid cells in this study), Cell-cell communication analysis was applied using CellChat [172]. As a result, the most abundant interactions have occurred within CD4+ T cells for both VAT and liver (Figure 19.A). And CD4+ T cells in VAT exhibited a higher number of interactions with NK cells and Myeloid cells compared to their counterparts in the liver. The information flow or interaction strength of each signaling pathway was compared between VAT and liver (Figure 19.B). Interestingly, the interactions of interleukin IL10, IL1, and IL6 signaling pathways, were all more pronounced in VAT than in the liver.

In terms of the IL6 signaling pathway, different cellular communications were observed between VAT and the liver (Figure 20). In VAT, CD4+ T cells—including naive CD4 T cells, intermediate CD4 T cells, CD4 Trm cells, and FOXP3+ Tregs—demonstrated more interactions with myeloid cells. While in the Liver, the IL6 signaling interactions were mainly activated among different CD4+ T cells, such as between naive CD4 T cells and FOXP3+ Tregs. The enrichment of IL6 signaling in VAT was verified experimentally as well (Figure 21.A, performed by Jonas Wagner). To our notice, the IL6 signaling pathway can activate STAT3, which in turn leads to the high expression of BATF, a VAT-specific marker that we have discovered in the previous section in the CD4+ T cell population. To validate this hypothesis, BATF expression was examined following time with or without IL6 stimulation (Figure 21.B, performed by Jonas Wagner), and the higher expression of BATF after IL6 stimulation was observed. In sum, cell-cell communication of the IL6 signaling pathway between various CD4+ T cells and other innate immune cells was more enriched in VAT, which subsequently contributed to the increased expression of the transcription factor BATF in VAT.

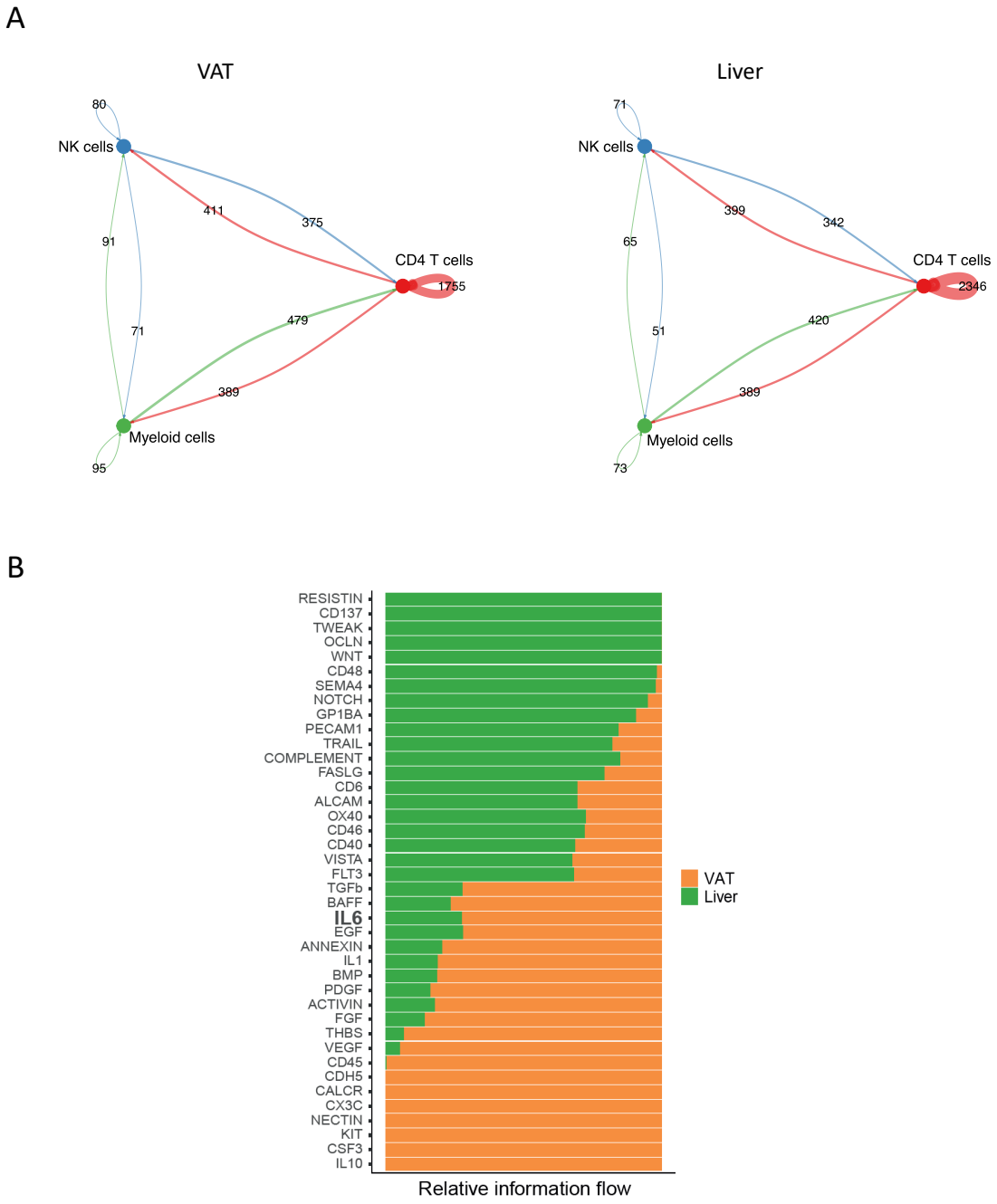



**Figure 19: Overview of cell-cell communication between CD4+ T cells and other non-T cells** A: number of interactions between CD4+ T cell, NK cells and Myeloid cells in VAT (left) and liver (right) respectively. B: The information flow was defined by the sum of communication probability among all pairs of cell groups in the inferred network. The relative information flow in liver and VAT was shown in the stacked barplot.

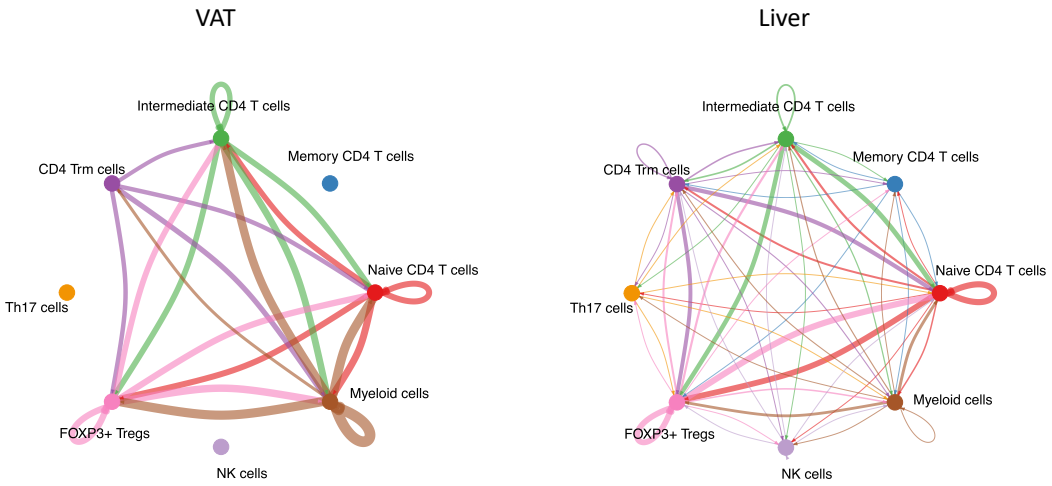



**Figure 20: IL6 signaling network**

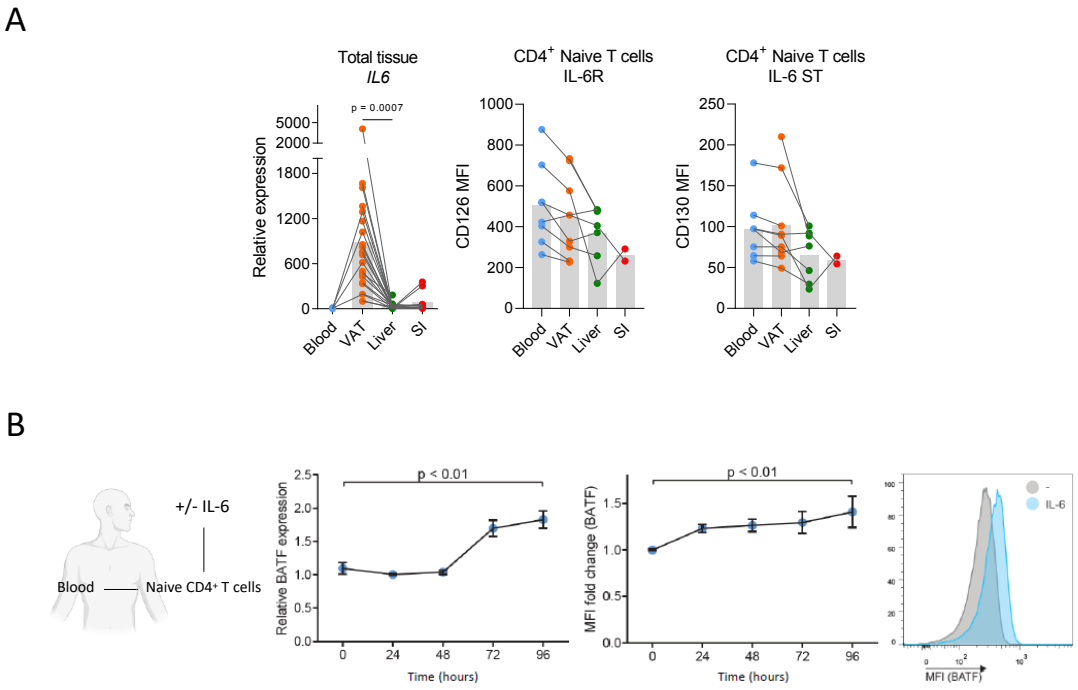



**Figure 21: IL6 signaling experimental validation.** A: qPCR of the receptor and ligands of IL6 signaling, including IL6, IL6R, and IL6ST. B: Flow cytometry analysis of BATF expression in naive CD4 T cells along time with or without IL6 stimulation. The fold change is measured compared to time zero.

## 3.3 Tissue adaptation of CD8+ T cells

### 3.3.1 Tissue adaptation signatures of CD8+ T cells

Applying the same strategy, tissue adaptation signatures of CD8+ T cells were obtained by comparing the cells with shared clonotypes in CD8+ T cells between tissues. Similarly, three comparisons were implemented (Figure 22.A), leading to the identification of common and tissue-specific adaptation signatures for liver and VAT (Table 9). In line with the findings in CD4+ T cells, VAT CD8+ T cells also showed more adaptive compared to the liver CD8+ T cells based on the higher number of adaptation signatures observed (VAT: 275 and Liver: 156). A significant overlap between the two sets of tissue signatures was found in CD8+ T cells as well, with a total of 141 shared gene signatures. Reactome pathway analysis by the common tissue signatures highlighted the activation of interleukins, pathways related to the higher expression of Heat Shock protein-coding genes (*HSP90AA1/HSPA1B/HSP90AB1*), and HSFP1-dependent transactivation and ESR-mediated signaling. For VAT-specific adaptation signatures, many previously identified tissue-specific signatures in CD4+ T cells were also found in CD8+ T cells, including *MT2A*, *ARID5A/B*, *CD44*, *ZFP267*, and *LGALS1* (Table 10). While for *BATF*, although it's still highly expressed in VAT CD8+ T cells, its expression was not significantly higher in VAT than in the liver. As for liver-specific signatures in CD8+ T cells, several distinct genes were uncovered, including chemokine ligand *CCL3* and the receptor *CD160*. Taken together, the substantial overlap of tissue signatures between CD8+ T cells and CD4+ T cells indicated a shared mechanism of T cell adaptation to their surrounding environment. Meanwhile, the unique tissue signatures specific to each T cell lineage may highlight their distinct roles and functions in the tissue site.

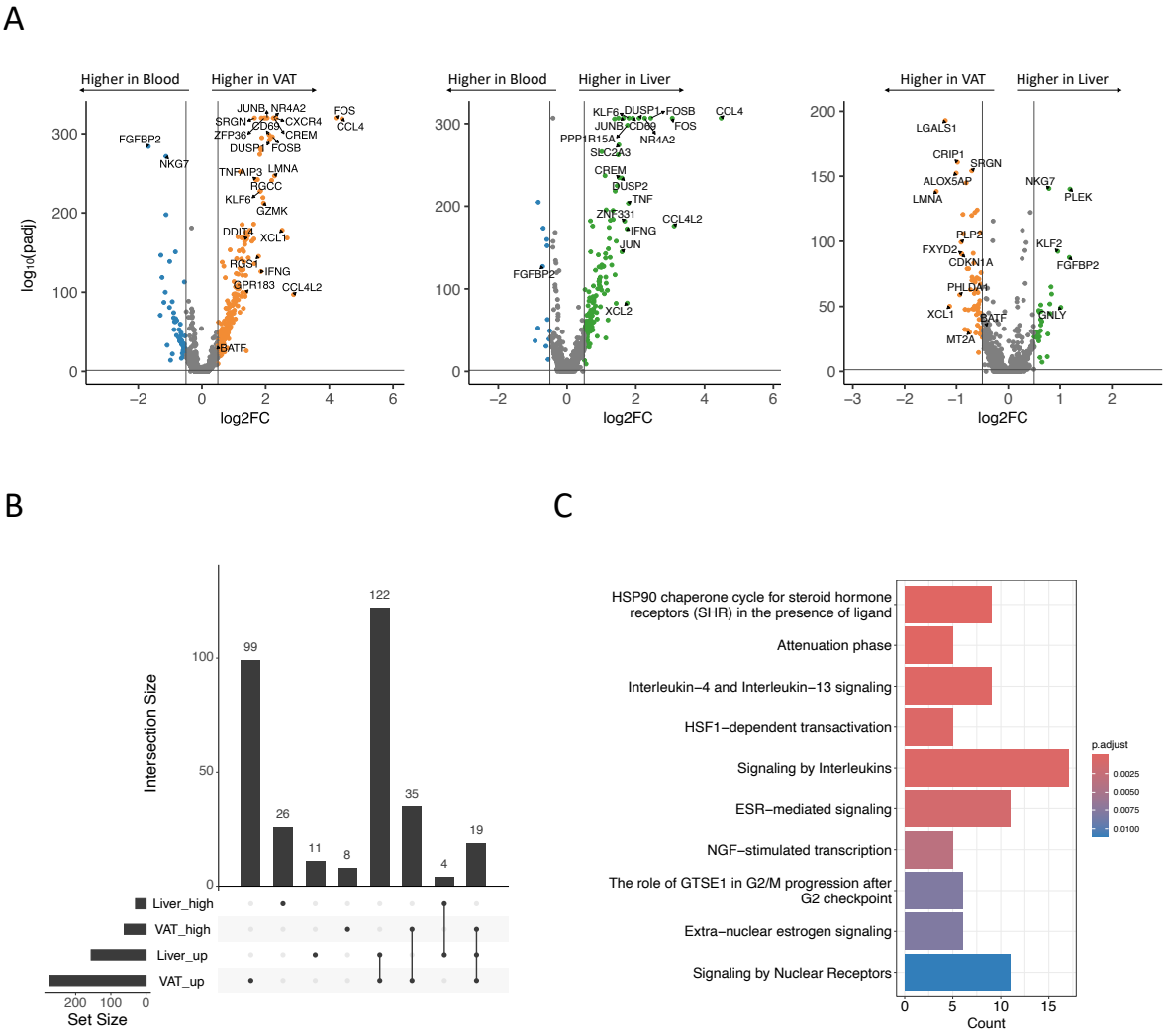



**Figure 22: Adaptation of CD8+ T cells with shared clonotypes.** All the comparisons here were based on the cells with shared clonotypes. A: Volcano plots of the DGE results for VAT vs blood (left), liver vs blood (middle), and liver vs VAT (right). B: Upset plot showing the intersection of different DEG sets: "Liver_high" represents the upregulated genes comparing liver to VAT, "Liver_up" represents the upregulated genes comparing liver to blood, and similar for "VAT_high" and "VAT_up". C: Top 10 enriched pathways of the tissue-common adaptation genes using the Reactome database.

**Table 9:** Tissue common adaptation signatures of CD8+ T cells

| Common between VAT and Liver | | | VAT higher | Liver higher |
|---|---|---|---|---|
| JUNB | REL | HSP90AB1 | CREM | DUSP1 |
| NR4A2 | DDX3X | VPS37B | CXCR4 | NR4A2 |
| TSC22D3 | CCL4L2 | RASGEF1B | FOS | CD69 |
| CCL4 | FAM177A1 | GABARAPL1 | SRGN | TNFAIP3 |
| CD69 | JUND | PTPN22 | ZFP36 | RGS1 |
| FOSB | NFKBIA | TRAF1 | ZFP36L2 | DUSP2 |
| BTG2 | NFKBIZ | EIF5 | LMNA | IFNG |
| DUSP1 | ZFP36L1 | H2AFZ | CDKN1A | CCL4 |
| SLC2A3 | SDCBP | HSPD1 | RGCC | NR4A1 |
| DUSP2 | YPEL5 | CD28 | TUBA4A | JUN |
| MCL1 | EIF4A1 | IVNS1ABP | XCL1 | CD40LG |
| BTG1 | SLA | TUBA1A | EZR | |
| TNFAIP3 | SAT1 | CRTAM | ANXA1 | |
| PPP1R15A | ID2 | EIF4A3 | XCL2 | |
| KLF6 | ELF1 | CLEC2B | PHLDA1 | |
| ZNF331 | TUBB4B | GNG2 | DDIT4 | |
| GZMK | CKS2 | SH2D2A | EML4 | |
| CYTIP | CRYBG1 | AC020916.1 | BIRC3 | |
| TENT5C | NFE2L2 | EVI2A | GPR183 | |
| FTH1 | CDKN1B | SFPQ | | |
| BRD2 | CITED2 | ARL4A | | |
| MYADM | HSP90AA1 | NFKB2 | | |
| DNAJB1 | HSPA1B | ARF4 | | |
| DNAJA1 | IER2 | RORA | | |
| RGS1 | IFNGR1 | ELL2 | | |
| SRSF7 | AREG | IDS | | |
| PIK3R1 | SYTL3 | CD83 | | |
| SAMSN1 | SRSF2 | PPP1R10 | | |
| EIF1 | ISG20 | IFRD1 | | |
| JUN | SERTAD1 | ENC1 | | |
| RHOH | NR4A3 | IL7R | | |
| PTGER4 | KDM6B | HSPA1A | | |
| PDE4B | SLC38A2 | GPR65 | | |
| IFNG | TIPARP | STAT4 | | |
| CSRNP1 | HSPE1 | RAB8B | | |
| TAGAP | USP36 | PMAIP1 | | |
| TNF | TSPYL2 | BAZ1A | | |
| NR4A1 | ZC3H12A | SMAP2 | | |
| ADGRE5 | PABPC1 | PDCD4 | | |
| CXCR6 | KPNA2 | BCL2A1 | | |
| GADD45B | CYCS | | | |

**Table 10:** Tissue-specific adaptation signatures of CD8+ T cells

| VAT specific | Liver specific |
|---|---|
| ALOX5AP | CCL3 |
| LEPROTL1 | KLF2 |
| SARAF | CD160 |
| SOCS1 | CMC1 |
| UBC | |
| VIM | |
| CD44 | |
| DUSP4 | |
| RNF19A | |
| HERPUD1 | |
| CAPG | |
| NAMPT | |
| FRMD4B | |
| CBLB | |
| FXYD2 | |
| TPM4 | |
| PLP2 | |
| PDCD1 | |
| CD74 | |
| HLA-DQB1 | |
| ARID5B | |
| CLU | |
| ARID5A | |
| ATP2B1 | |
| CRIP1 | |
| ITM2A | |
| TAGLN2 | |
| LGALS3 | |
| COTL1 | |
| TXN | |
| LGALS1 | |
| HLA-DRA | |
| ZNF267 | |
| MT2A | |
| HLA-DRB5 | |

### 3.3.2 CD8 Trm signatures of the small intestine

Sub-clustering of all CD8+ T cells was performed (Figure 23). To our interests, there were two naive CD8+ T cell clusters recovered: Naive (1) was enriched in Blood, and Naive (2) was mostly present in the liver and VAT. CD8 TEM/CTL cells were grouped into two clusters as well: TEM/CTL (1) were derived from VAT and liver while TEM/CTL (2) appeared to be exclusive in VAT. Besides, TEMRA/TEFF cells, MAIT cells, and NKT cells each formed their distinct clusters. For CD8 Trm cells, they were almost all from the small intestine and their gene markers showed exclusively high expression of *GPR34*, *CCR9*, *LDLRAD4*, *ITGA1*, and *SPRY1*, highlighting a unique gene expression profile of CD8 Trm cells in the small intestine.

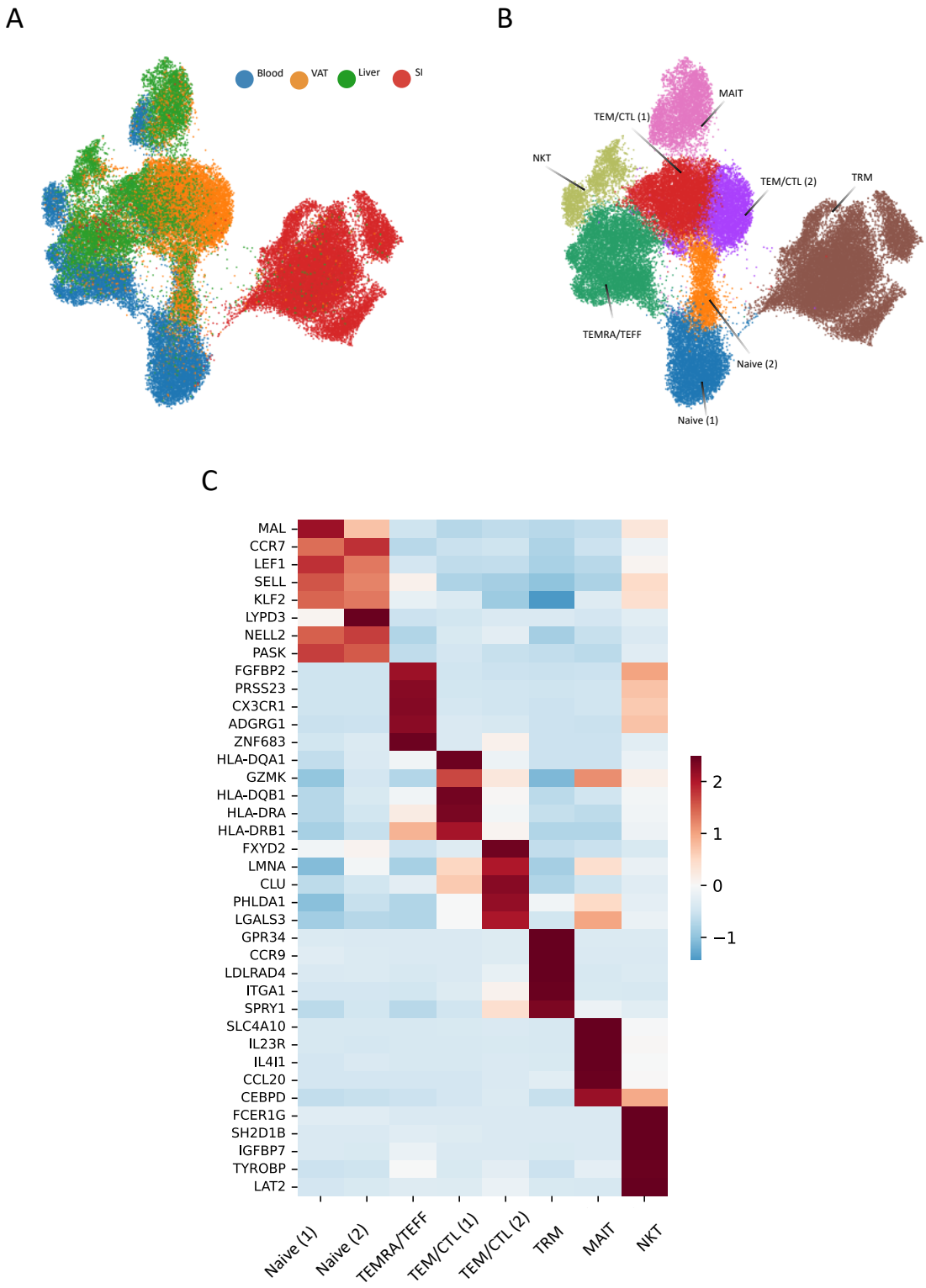



**Figure 23: Sub-clustering of CD8+ T cells.** A-B: UMAP of all CD8+ T cells colored by tissue (A) and cluster annotations (B). C: Heatmap showing the top 5 markers of each sub-type of CD8+ T cells.

In order to uncover tissue-resident signatures of CD8+ T cells for the small intestine, differential expression analysis comparing SI CD8 Trm cells to the other CD8 T cells was performed (Figure 24). A large number of upregulated and downregulated DEGs were revealed (Table 11). Among the upregulated genes, Chemokine receptor *CCR9* was known to play a role in the homing, migration, and retention of CD8+ T cells in the small intestine [195]. *CCR9* interacts with its ligand *CCL25*, which helps to ensure intestinal immunity and maintenance of mucosal homeostasis [196]. G-protein coupled receptor *GPR34*, implicated in the regulation of inflammatory responses and immune cell migration [197], was highly expressed in the small intestine, indicating the potential involvement in regulating T cell function in the small intestine, especially concerning tissue repair and the response to inflammation. The higher expression of *SPRY1* (sprouty RTK Signaling Antagonist 1), which was found to enhance the survival and early differentiation of effector and memory CD8+ T cells [198], suggested its function in maintaining the survival of T cells in the small intestine. Additionally, the transcription factor *IRF1* (interferon regulatory factor 1) was upregulated in the small intestine, although not among the top resident markers. The importance of *IRF1* in the intestine has been reported within many immune cells, for example, natural killer cells and intestinal intraepithelial T cells [199]. Overall, the significant and unique changes in gene expression profiles of SI CD8+ Trm cells suggested the strong and specialized tissue adaptation in response to the local environment.

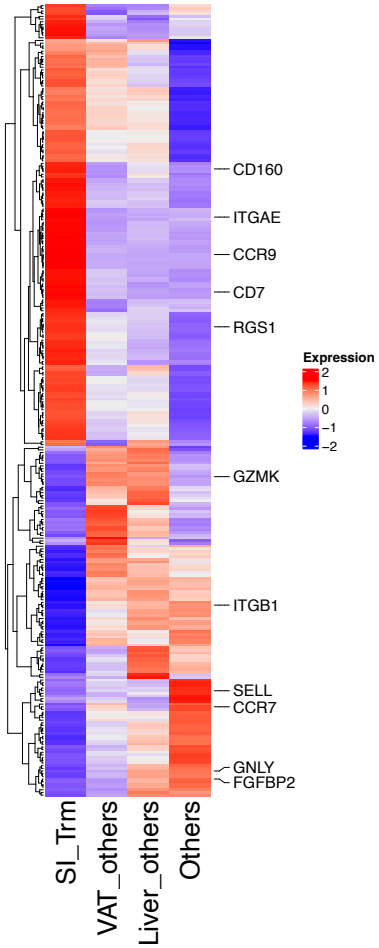



**Figure 24: DEGs of SI CD8 Trm cells.**

**Table 11:** DEGs in small intestine CD8+ Trm cells

| Up-regulated | | | |
|---|---|---|---|
| XCL2 | GFPT2 | ABLIM1 | HNRNPDL |
| LDLRAD4 | DNAJA1 | TRAV26-2 | KLRC1 |
| XCL1 | NR4A1 | FTH1 | IDS |
| HSPA1A | ANKRD28 | TIGIT | RGL4 |
| SPRY1 | TSC22D3 | DNAJB1 | LIMS1 |
| RGS1 | REL | TPI1 | ATP1B3 |
| TRBV6-1 | HSPH1 | MYC | IFNG |
| CD160 | MAP3K8 | GABARAPL1 | SIRPG |
| CCL4L2 | FDFT1 | SQLE | RELB |
| TMIGD2 | PGK1 | ZEB2 | SDCBP |
| ITGAE | APOA1 | SCML4 | PPP1CB |
| TNFAIP3 | GPR65 | TRAV19 | ABRACL |
| HMGCS1 | OXNAD1 | ENSA | MSMO1 |
| TSPYL2 | CYCS | CYP51A1 | FOS |
| CAPG | SPINK2 | RANBP2 | CD3E |
| CD7 | INSIG1 | NFKBIZ | RGS2 |
| PDE4D | CD69 | CD3G | TNIP3 |
| ITM2C | NR4A2 | BCL2A1 | CD8B |
| CKLF | P2RY10 | CCL3L1 | SELENOK |
| HSPA1B | LINC01871 | RPLP1 | MLF1 |
| TRBV6-5 | OSTF1 | EVL | TANK |
| EVI2A | HSP90AB1 | JUN | CHI3L2 |
| SMIM3 | CCL5 | HSPD1 | UBC |
| CCR9 | AUTS2 | CD8A | SAP18 |
| FOSB | PPP1R15A | CYTIP | GNG2 |
| KLF6 | RASGEF1B | CD248 | MCL1 |
| PTPN22 | H2AFZ | SLA2 | YME1L1 |
| CD96 | LDLR | ELF1 | HNRNPL |
| IL7R | PFKFB3 | H1FX | NR3C1 |
| GPR34 | CSRNP1 | LRRN3 | CD101 |
| ITGA1 | JUND | JMY | TRGV2 |
| PTGER4 | IL2 | PER1 | SFPQ |
| SYTL3 | UBE2S | NEU1 | EIF4A3 |
| ATF3 | ATF7IP | ZNF331 | MAP4 |
| YPEL5 | ARHGEF7 | ABCB1 | LINC00963 |
| TRAV17 | TRBV11-2 | NSMCE3 | FAM122C |
| HSP90AA1 | SAMSN1 | NFKBIA | PDCD4 |
| MYADM | CD3D | HSPA5 | PTGES3 |
| HSPE1 | NR4A3 | CYTOR | KLRD1 |
| CD55 | ARL4A | CRTAM | |
| HECW2-AS1 | GPR171 | TMEM14C | |
| CD9 | LZTFL1 | AREG | |

| Down-regulated | | | |
|---|---|---|---|
| GNLY | FLNA | PCED1B | CD27 |
| GZMK | TSPO | CST7 | RPL26 |
| GZMH | TXNIP | CD5 | ADD3 |
| AHNAK | RASA3 | GZMA | C12orf75 |
| ITGB2 | NOSIP | HLA-DQA1 | PRSS23 |
| CTSW | SAMD3 | LEF1 | RPL6 |
| KLF2 | S1PR1 | SORL1 | INPP4B |
| GZMB | TC2N | TIMP1 | AQP3 |
| FGFBP2 | TLE5 | SPOCK2 | TPST2 |
| HLA-DRB1 | ZNF683 | CNN2 | SYTL1 |
| ITGB1 | S100A11 | GLUL | EFHD2 |
| LTB | NCR3 | S1PR4 | STK38 |
| LMNA | CD44 | BHLHE40 | PIM1 |
| PRF1 | GIMAP4 | DDIT4 | RPS4Y1 |
| MT2A | SH2D1A | TGFBR3 | XBP1 |
| NKG7 | LGALS1 | S1PR5 | TCF7 |
| SELL | APOBEC3G | HLA-DRB5 | APMAP |
| CCR7 | HLA-DRA | PLAC8 | MATK |
| KLRG1 | ARID5B | IL10RA | LDHB |
| PCED1B-AS1 | CD74 | ADGRG1 | BIN2 |
| PLEK | PRMT2 | TMSB10 | DUSP2 |
| TUBA4A | FCGR3A | RGCC | ADRB2 |
| P2RY8 | LIME1 | TGFB1 | LYST |
| TYROBP | HLA-DPB1 | CD28 | SLC38A1 |
| GIMAP7 | CMC1 | HLA-DPA1 | NDFIP1 |
| LINC00861 | IFNGR1 | ATM | CYTH1 |
| LITAF | GBP5 | CDC25B | FCMR |
| FXYD5 | ENC1 | NELL2 | C1orf162 |
| RIPOR2 | EMP3 | OASL | PXN |
| S100A4 | HLA-DQB1 | GAS5 | RPS27 |
| KLF3 | HLA-C | IL16 | CCDC167 |
| IER2 | SYNE1 | CX3CR1 | RAB29 |
| S100A6 | LIMD2 | FCRL6 | |

## 3.4 Tissue adaptation of myeloid cells and NK cells

### 3.4.1 Tissue adaptation signatures of NK cells

Differential gene expression analysis was performed on natural killer or innate-like cells annotated from the global clustering. Since fewer than 100 cells were recovered in the small intestine, it was removed for further analysis in NK cells. The tissue adaptation gene signatures for VAT and liver were displayed (Figure 25.A and Table 12). The intersection between liver and VAT adaptation signatures was examined, revealing minimal to no overlap, which underscored the tissue specificity of NK cells in terms of their adaptation gene signatures (Figure 25.B). We also observed some common tissue signatures of NK cells with those of T cells, especially CD8+ T cells. For liver-specific signatures, genes such as *CCL3*, *CCL4*, *DUSP1*, *CD160*, and *CD69* were highly expressed in both NK cells and CD8+ T cells. And for VAT-specific signatures, genes including *LMNA*, *VIM*, *MT2A*, and *CXCR4* showed high expression levels in both NK cells and CD8+ T cells. These results could be explained by the shared characteristics between NK cells and CD8+ T cells, such as cytotoxicity, which remains functional across different tissues.

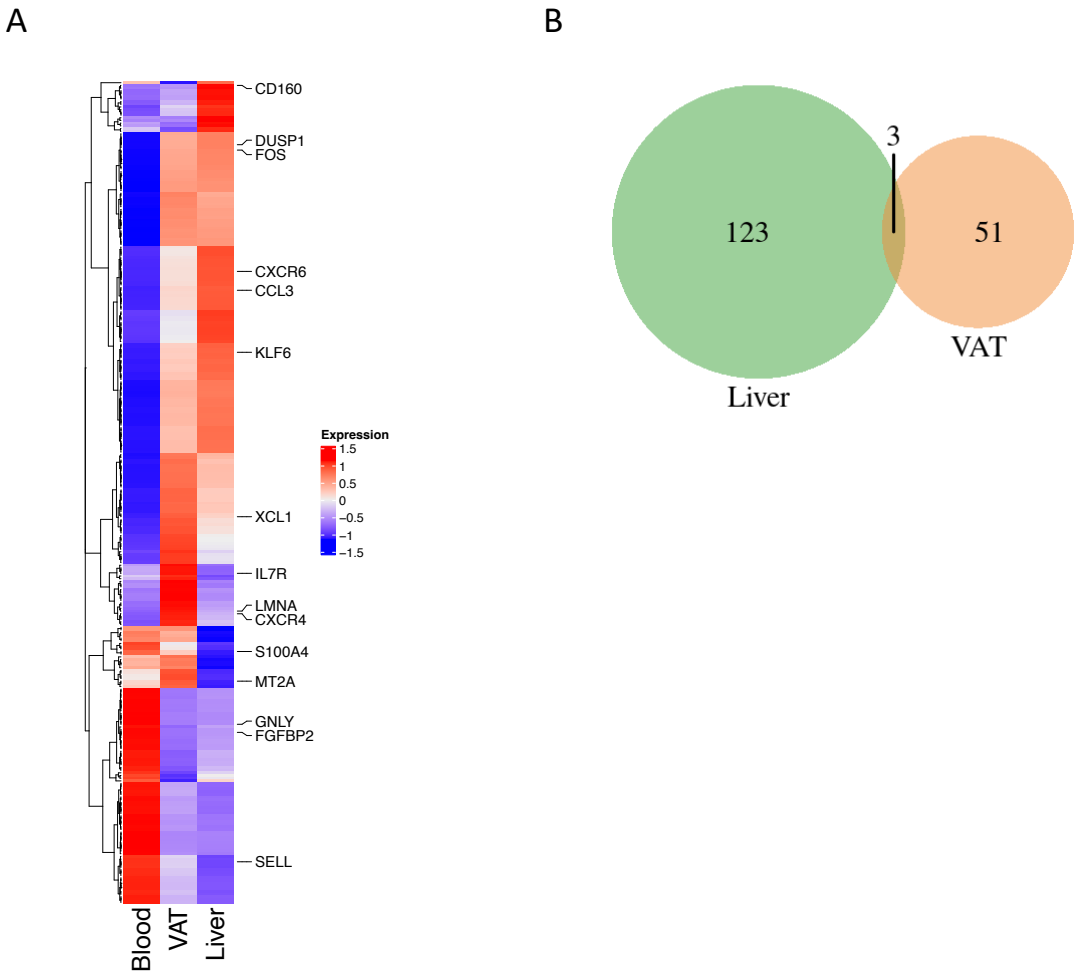



**Figure 25: Tissue adaptation of NK cells.** A: Heatmap showing DEGs comparing each tissue to others. B: Venn diagram of the intersection of tissue DEG lists.

**Table 12:** Tissue adaptation signatures of NK cells

| Liver | | | VAT | |
|---|---|---|---|---|
| CCL3 | KDM6B | FKBP5 | VIM | NR4A1 |
| CCL4 | USP36 | SRGN | LMNA | CD55 |
| CCL4L2 | GADD45B | GABARAPL1 | CSF2 | MIR155HG |
| KLF6 | NEAT1 | SLA | XCL1 | ICAM1 |
| FOSB | GYG1 | HSPH1 | CRIP1 | HSPA5 |
| NR4A2 | FAM177A1 | HNRNPAB | LGALS1 | ANXA5 |
| FOS | ZBTB16 | JUND | CAPG | NIBAN1 |
| CCL3L1 | MCL1 | SLC38A1 | CXCR4 | FTH1 |
| ATP1B3 | DUSP2 | SERTAD1 | CD44 | NINJ1 |
| YPEL5 | CRYBG1 | CD38 | MT2A | ANXA1 |
| ZNF331 | PABPC1 | PDE7A | CD83 | FXYD5 |
| FYN | FASLG | IFRD1 | XCL2 | PHLDA1 |
| SYTL3 | GZMK | GLUD1 | HSPA1B | |
| CEBPD | KLF2 | NFKBIZ | IL7R | |
| DUSP1 | UBE2S | RORA | AHNAK | |
| CEMIP2 | JUN | SH2D1A | DUSP4 | |
| CD160 | CITED2 | ADGRG3 | HSPA1A | |
| CXCR6 | PDCD4 | MYBL1 | S100A10 | |
| TAGAP | DNAJA1 | BRD2 | EZR | |
| PPP1R15A | EIF4A3 | REL | S100A11 | |
| RANBP2 | IFNG | EOMES | HLA-DRB1 | |
| TSC22D3 | AL135905.2 | SQSTM1 | S100A6 | |
| TSPYL2 | ELF1 | EIF4G2 | PLP2 | |
| SMAP2 | GFOD1 | RALGAPA1 | GPR183 | |
| ZFP36 | CXorf40A | HSP90AB1 | CRTAM | |
| ADGRE5 | PIK3R1 | IER5 | TNFRSF18 | |
| CMC1 | LZTFL1 | NR4A3 | BCL2A1 | |
| HSPD1 | OFD1 | BTG2 | TAGLN2 | |
| GRAMD2B | SDCBP | SRSF7 | TUBA4A | |
| PTGER4 | MYADM | HSP90AA1 | HLA-DRA | |
| AREG | ARHGAP9 | SH2D2A | RGCC | |
| CD69 | TIPARP | EIF5 | CDKN1A | |
| CSRNP1 | PTPN22 | IVNS1ABP | NFKBID | |
| ELL2 | NR4A1 | YES1 | OASL | |
| NFKBIA | HSPA8 | LBH | SAMSN1 | |
| TUBB4B | TOX2 | VPS37B | TRAF1 | |
| ZFP36L1 | ISG20 | EML4 | ZC3H12A | |
| PDE4B | PNRC1 | ATP6V0C | MT1X | |
| RAB11FIP1 | ICAM1 | H2AFZ | NFKBIZ | |
| TLE1 | TNFSF14 | SOCS1 | CXCR3 | |
| SFPQ | RELB | DENND2D | S100A4 | |
| DNAJB1 | TIGIT | WSB1 | BIRC3 | |

### 3.4.2 Tissue adaptation signatures of myeloid cells

In terms of myeloid cells, although only a few cells recovered, they were present in all tissue samples in our study. Thus DE analysis was performed on all four tissues, which revealed a large number of significant DEGs (Figure 26 and Table 13) . The small intestine exhibited the highest tissue specificity, characterized by the greatest number of upregulated genes, followed by VAT and then the liver. The overlapping tissue signatures were found between the liver and SI, as well as between VAT and SI. However, no overlap was detected between VAT and liver (Figure 26.B). Similarly, it was also observed that certain tissue signatures of myeloid cells were shared with those identified in other cell types (Figure 26.A). In sum, the DGE results on myeloid cells demonstrated their adaptive or resident nature, with specificity to each tissue site. The identified tissue signatures for both NK cells and myeloid cells overlapped with T cells, suggesting shared characteristics and potential common pathways of tissue adaptation across different immune cell populations.

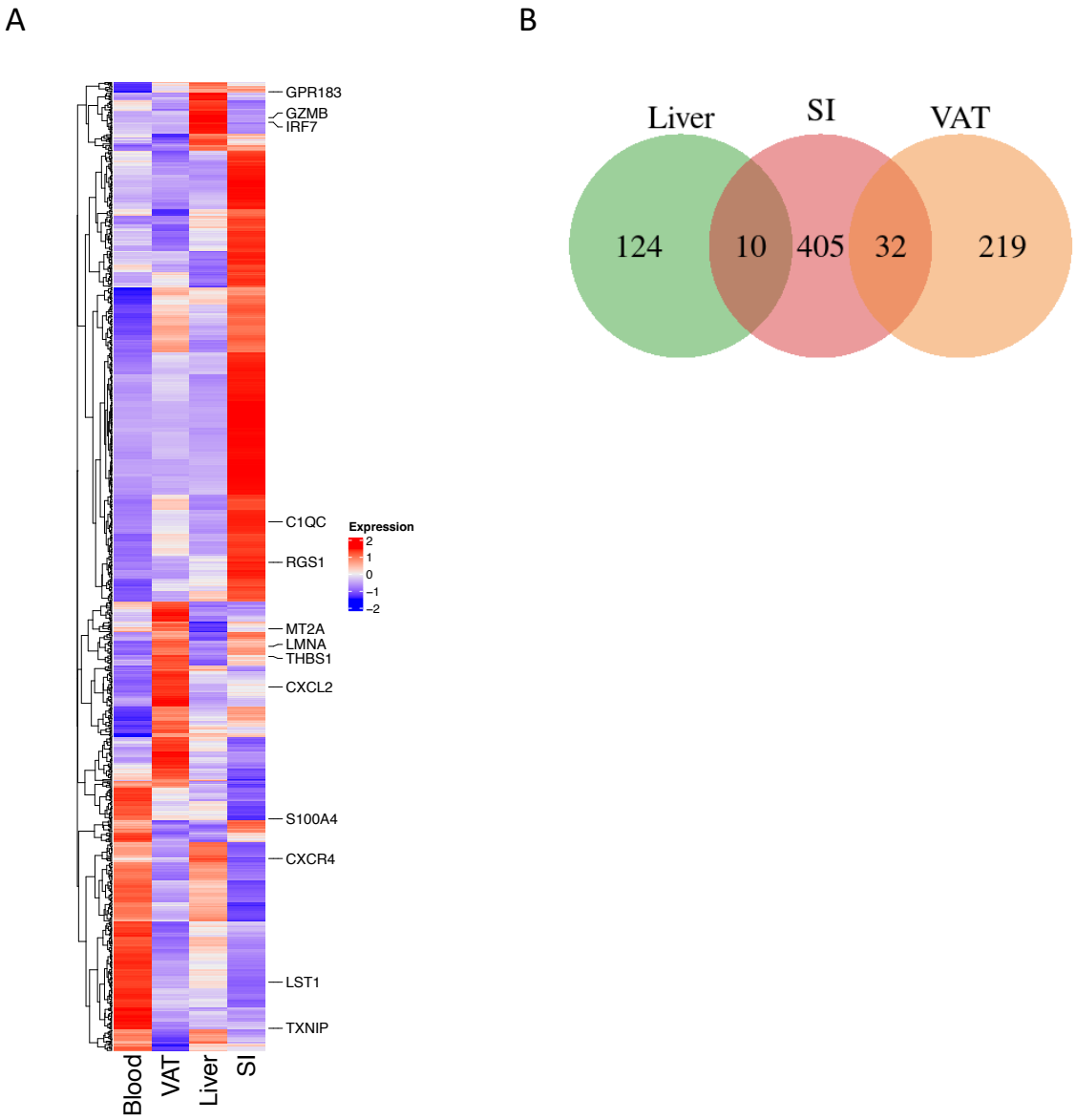



**Figure 26: Tissue adaptation of Myeloid cells.** A: Heatmap showing DEG comparing each tissue to others. B: Venn diagram of the intersection of tissue DEG lists.

**Table 13:** Tissue adaptation signatures of myeloid cells

| Liver | | | |
|---|---|---|---|
| IGKC | RAC2 | PSMB8 | MED30 |
| GZMB | CXCR4 | STMN1 | PIM2 |
| JCHAIN | SPN | PSMB9 | CORO1B |
| IRF7 | RASD1 | SERPINA1 | FDFT1 |
| PLAC8 | MAP1A | PSME2 | EEF1A1 |
| LILRA4 | CPNE3 | RPSA | SEPTIN9 |
| LTB | SELPLG | APBB1IP | RBBP4 |
| TPM2 | JUN | LILRB2 | C12orf57 |
| CCDC50 | PTPRCAP | NUCB2 | GNLY |
| ITM2C | CORO1A | PMAIP1 | IFIT2 |
| PLD4 | MYL12A | OAS1 | ABI3 |
| IL3RA | PTCRA | DERL3 | CYFIP2 |
| C12orf75 | SPCS1 | ARL4C | BCL11A |
| SOX4 | LIME1 | LILRB4 | NACA |
| MZB1 | TUBA1A | PAXX | RPL41 |
| FCGR3A | LSP1 | CXorf21 | ITGA4 |
| IRF4 | NKG7 | KLF4 | HIST1H2BK |
| CLIC3 | POLB | ABRACL | SRSF5 |
| EGLN3 | NR4A3 | CDKN1C | TNFSF10 |
| IRF8 | TSPAN13 | PSME1 | KPNA2 |
| UGCG | LST1 | LSM6 | SRSF7 |
| SMIM25 | RAB11FIP1 | PTPN6 | CD37 |
| RHOC | PHB | SPINT2 | LRRC26 |
| SPIB | HIST1H1C | PPIA | RNF5 |
| GPR183 | CYB561A3 | HVCN1 | RPS3A |
| TCF4 | PARK7 | IGFLR1 | CCNI |
| CLEC4C | LY9 | CXCR3 | RABGAP1L |
| SEC61B | GPR65 | CHCHD2 | HMGN1 |
| APP | PPIB | PFN1 | MT-ATP8 |
| NR4A1 | IL2RG | NAP1L1 | SH3BP1 |
| NIBAN3 | UCP2 | MYL12B | UNC119 |
| MYBL2 | LIMD2 | POLD4 | SEPTIN6 |
| SERPINF1 | HLA-DPB1 | PPP1R14B | |
| OFD1 | PECAM1 | CNPY3 | |

| VAT | | | | | |
|---|---|---|---|---|---|
| RNASE1 | ACSL1 | WTAP | SLC1A3 | CD55 | PELI1 |
| THBS1 | BCL2A1 | TRAF1 | GYPC | CD63 | PLIN2 |
| CXCL8 | ETS2 | MAF | CDKN1A | LTC4S | RAPH1 |
| CXCL2 | F13A1 | PHLDA1 | AC007384.1 | CEBPB | SRGAP1 |
| CXCL3 | OLR1 | SAT1 | FKBP5 | ITGAX | TFRC |
| LYVE1 | FILIP1L | TNFAIP8 | HBEGF | GADD45B | TLNRD1 |
| HMOX1 | LUCAT1 | DDIT4 | NEAT1 | CCR7 | RGL1 |
| IL1B | NAMPT | RNF144B | PRNP | KYNU | HSPA8 |
| SOD2 | HSPH1 | RGCC | TREM1 | SERTAD1 | JARID2 |
| AREG | MMP19 | AKAP13 | EIF4E | PHACTR1 | SDF2L1 |
| CCL2 | CCL4 | PNPLA8 | TOP1 | IL6ST | ARFGAP3 |
| G0S2 | SYAP1 | ATF3 | XBP1 | SPHK1 | CREB5 |
| EREG | SERPINB9 | MAN1A1 | ZFP36 | NABP1 | MCEMP1 |
| LMNA | NFE2L2 | HSPB1 | ICAM1 | ANKRD28 | P2RY14 |
| NFKBIA | MS4A4A | ZC3H12A | PTPN1 | ATP13A3 | VAPA |
| MRC1 | PPP1R15A | MAT2A | SNX9 | BCL3 | TRIB1 |
| EMP1 | THBD | GCLM | SPAG9 | ANXA1 | FOSL1 |
| IER3 | ATP1B3 | GLUL | MSR1 | CD14 | B3GNT5 |
| CD163 | CCL3L1 | SOCS3 | MAP2K1 | VEGFA | ZBTB16 |
| MT2A | TLR2 | SERPINB1 | CPM | PID1 | MAFF |
| HSPA1A | PNRC1 | HSP90AB1 | FCGR2A | ATP2B1 | IL4R |
| SERPINB2 | CTSL | SAMSN1 | CXCL16 | STK17A | ANXA5 |
| IL1R2 | DAB2 | ABL2 | ARL6IP1 | ADM | LITAF |
| FOLR2 | PLTP | DNAAF1 | METRNL | CDC42EP2 | ARHGAP6 |
| DNAJB1 | CCL4L2 | CD83 | ANPEP | HSPA5 | ARHGAP18 |
| TIMP1 | HSP90AA1 | BATF | SLC2A3 | MIR155HG | LHFPL2 |
| NFKB1 | MIR3945HG | PDE4DIP | PDE4B | INSIG1 | REL |
| HSPA1B | CLEC4E | NFKB2 | TUBA1C | ZEB2 | VSIG4 |
| IL1RN | CEBPD | DUSP2 | PFKFB3 | BRD2 | CCRL2 |
| TNFAIP6 | DNAJA1 | CD44 | PTX3 | ZNF267 | EPB41L3 |
| PPIF | BIRC3 | HIF1A | HSPD1 | ELF1 | EIF1B |
| SRGN | C5AR1 | EHD1 | FOXO3 | BEST1 | JMJD1C |
| ELL2 | KDM6B | PNP | TUBB4B | SFPQ | HIVEP2 |
| CREM | SH3BP5 | DSE | FPR1 | HSPE1 | PPP1CB |
| CCL3 | NRP2 | NFKBIZ | MARCKS | EZR | TP53BP2 |
| BASP1 | C15orf48 | CCL20 | MCL1 | GPBP1 | IL7R |
| GK | AQP9 | ATP2B1-AS1 | TXN | ACSL3 | SMAP2 |
| BTG1 | FTH1 | TGIF1 | IL6R | RSRP1 | CCNH |
| SQSTM1 | YBX3 | IVNS1ABP | PLSCR1 | RBPJ | ZFYVE16 |
| NLRP3 | SDCBP | FCGR2B | GPAT3 | RILPL2 | FCAR |
| PLAUR | IRAK2 | ITSN1 | DNTTIP2 | KLF6 | PHF20 |
| TNFAIP3 | ISG20 | CFLAR | MAP2K3 | H3F3B | |

| SI | | | | | |
|---|---|---|---|---|---|
| C1QC | PAPSS2 | PRMT9 | MERTK | SLC4A7 | GM2A |
| RGS1 | PLAU | FCHO2 | RAB20 | CREG1 | MT-ND6 |
| DNASE1L3 | LGALS3BP | ATP6V0D2 | FOSB | SIGLEC10 | FRMD4B |
| C1QB | GPR137B | AKR1B1 | CTSZ | SH3PXD2B | RBM47 |
| C1QA | IGF1 | GZMA | MMP12 | EPB41L2 | CTSL |
| SELENOP | HLA-DQA1 | JUN | AKR1A1 | HES1 | ERO1B |
| HLA-DRB5 | HLA-DRB1 | NGLY1 | NPC2 | ZNF10 | CTNND1 |
| ACP5 | ADAP2 | FCGRT | RANBP2 | RGS10 | SLC37A2 |
| SDS | PRDM1 | SLC19A2 | PTMS | HIST1H2AC | NFKBIE |
| LIPA | C2 | ZNF141 | KLF6 | HLA-DOA | CTSA |
| ZNF331 | CD74 | GNL3 | CKS2 | EIF4A3 | ELMO1 |
| CTSC | RGS2 | SRSF7 | HCAR2 | CLEC11A | PDE4B |
| CD163L1 | HLA-DQB1 | GNPDA1 | IL18BP | LPAR6 | TNFAIP3 |
| ID3 | LACC1 | PRDX1 | AC004817.3 | FPR3 | CCL3L1 |
| FUCA1 | HLA-DPB1 | C1orf54 | HSP90B1 | CNOT2 | RALGAPA1 |
| CTSD | HSPA5 | ENG | SLC20A1 | PDK3 | TCF12 |
| APOE | TIPARP | ANKH | PELI1 | MMP19 | TSC22D1 |
| PLD3 | CCL4L2 | SASH1 | GADD45B | STAB1 | CTTNBP2NL |
| SGK1 | HLA-DMA | A2M | SGPL1 | DOCK10 | INSIG1 |
| APOC1 | DOCK4 | OLFML3 | FOS | ARHGAP10 | NR1H3 |
| APOC2 | FRMD4A | GCLC | SRGAP3 | YWHAH | ZFP36 |
| CCL24 | ITGB7 | VSIG4 | PDIA4 | TMED9 | TCN2 |
| CHMP1B | ATF3 | CD83 | GPCPD1 | PRCP | CMKLR1 |
| ABCA1 | GAL3ST4 | CD300A | MS4A7 | IL6 | CCDC107 |
| NRP1 | CALR | JAML | RASGEF1B | CTTNBP2 | NUPR1 |
| LGMN | YPEL5 | GATM | PDIA6 | TTC3 | DNAJB11 |
| TMEM176B | CCL4 | NR4A1 | MS4A6A | SMPDL3A | ACADVL |
| NR4A2 | CD80 | MANBA | SLC8A1 | RGS12 | SERPING1 |
| GLA | SLAMF7 | CSF1R | GLUD1 | SLCO2B1 | LAMP2 |
| SLC40A1 | CPVL | EPSTI1 | PIK3R3 | MMD | IDH1 |
| HEXA | ENPP2 | DNPH1 | NAIP | TPD52 | RAB7B |
| HLA-DPA1 | MT1X | STX4 | CDH17 | PPA1 | LNCAROD |
| TMEM176A | DAPK1 | QPRT | CD68 | VAMP8 | RAB9A |
| CD209 | IFRD1 | UCP2 | ICAM1 | MMP14 | CLEC4F |
| GPNMB | AP1B1 | MT1G | TAGAP | SIK2 | SYTL3 |
| ABHD5 | BORCS5 | PTGS2 | TOR3A | HEXB | CBLB |
| GEM | HLA-DRA | PDK4 | SKIL | MACF1 | KCNQ1OT1 |
| CTSB | ITPR2 | ENTPD1 | BAIAP2 | IGSF6 | SLC8B1 |
| LILRB5 | SERPINF1 | ABI3 | JPT1 | ABCC4 | IFI30 |
| PLA2G7 | HLA-DMB | ATP1B1 | CD86 | HPS5 | EBI3 |
| RCAN1 | HERPUD1 | CD40 | ARHGAP18 | CAPZB | ITM2B |
| ADAMDEC1 | RNASET2 | ADAM28 | CSF2RA | CALHM6 | CD59 |
| LINC00996 | MT-ATP8 | SNHG12 | AXL | LAIR1 | SERINC5 |
| DST | PTGER4 | SELENOK | LAG3 | FGL2 | FNDC3B |
| CXCL12 | GRN | NCF4 | ABHD12 | CHD2 | MPEG1 |

| | | | | | |
|---|---|---|---|---|---|
| C3AR1 | WIPF1 | SEPHS2 | GLMP | TMEM150B | HECTD2 |
| CMSS1 | IL2RA | SAT1 | MTR | BHLHE40 | IPMK |
| ATP6V1B2 | NCOA7 | SPIDR | PTPN22 | AC090617.10 | GALM |
| CYFIP1 | PMP22 | MGAT4A | TMEM26 | TMEM51 | XIST |
| ARPIN | SPATS2L | HSPA1B | GPR34 | TSPYL2 | GAA |
| NAF1 | OTULINL | REL | BST2 | ARHGAP17 | LIMS1 |
| ODF2L | DNAJB4 | TMBIM6 | ABCG1 | CLK1 | CLIC2 |
| DDB1 | LITAF | P4HB | FUS | IFNGR1 | CDKN1B |
| JUNB | CORO7 | ASAH1 | CNDP2 | AHI1 | ZFYVE16 |
| DMXL2 | TRIM14 | CCDC152 | LRMDA | ALOX5 | HERC1 |
| TFRC | UBAC2 | GADD45A | CHD9 | ATP8B4 | CD63 |
| MFSD1 | NR4A3 | EML4 | ASMTL | TGFBI | SEL1L |
| ATP2A2 | PRDX2 | GZF1 | USP37 | HERC2 | CLASP2 |
| SDSL | PALLD | TNIK | PSAP | HSPD1 | CREBRF |
| MKNK1 | WDR45B | TSC22D2 | RNF149 | TBXAS1 | TMEM37 |
| PDIA3 | RPN2 | EPAS1 | ZNF330 | USP36 | RB1 |
| B3GNT5 | GLUL | ARL5B | AP3B1 | EGR2 | FNBP4 |
| OSBPL1A | ALDH9A1 | NPL | TMEM97 | AHR | FAM53C |
| NFKBID | MAT2A | SLC7A8 | CISD2 | CLTA | HLA-E |
| LCORL | MAN2B1 | TENT2 | P4HA1 | MAPRE2 | RUNX3 |
| IL18 | SESN2 | PLBD2 | ABCC5 | BEX4 | SPNS1 |
| RALB | SAMSN1 | SLAMF8 | CD38 | ATP2C1 | SEMA6B |
| GPRIN3 | IRF8 | GIMAP4 | IL1RAP | ARHGAP25 | STT3A |
| TBC1D22A | RBP1 | ITGAV | CCNL1 | FN3K | SND1 |
| NFATC2 | P2RY6 | TBC1D4 | ZZEF1 | RENBP | NABP1 |
| CTSH | MS4A4E | HPS3 | ETNK1 | METTL15 | ZBTB1 |
| VASH1 | WDR91 | DNMT1 | MGST2 | GFRA2 | MCUR1 |
| TNFRSF14 | CDK5RAP3 | RHOH | WSB1 | CCL3 | |
| SCARB2 | DDIT3 | SEC23B | PAK1 | ATP6V0A1 | |
| CD36 | FNDC3A | NEU1 | MTAP | ME1 | |

# 4 Discussion

## 4.1 Tissue T cell adaptation

In this study, four human tissues were involved including three organs: visceral adipose tissue, liver, and small intestine, and the periphery blood. Single-cell RNA sequencing and TCR sequencing provided valuable high-quality data for the exploration of gene expression profiles across these tissues on the single-cell level. Particularly, there were a large number of T cells identified in the dataset, which allowed us to perform various data analyses, revealing the heterogeneity of T cells across tissues and how they adapt to their tissue environment.

### 4.1.1 Tissue T cell adaptation signatures revealed by DGE analysis

Differential gene expression analysis was applied to multiple cell populations across tissues in this study. Firstly, to uncover tissue adaptation signatures for CD4+ and CD8+ T cell lineages, we compared the gene expression profiles of cells from each lineage with overlapping T cell receptor clones. The obtained VAT and liver tissue adaptation signatures were compared, allowing the identification of both common and tissue-specific adaptation signatures. The DGE analysis was also conducted within subpopulations of T cells, including CD4 Trm cells and Naive CD4 T cells. These results revealed numerous overlapping tissue signatures across T cell subsets for both VAT and liver, demonstrating the reliability of these tissue signatures in capturing the adaptive responses of T cells within specific tissue environments.

**Liver and VAT shared common tissue adaptation signatures**
It's interesting to observe that the liver and VAT, as two distinct human organs, shared a large number of tissue signatures of T cells. Among these common adaptation

genes are the widely recognized tissue resident marker *CD69*, the chemokine receptor *CXCR4*, and the cytotoxic T cell markers *SRGN* and *IFNG*, highlighting the shared mechanism of T cell activation in diverse tissue environments. *CD69* is widely recognized as an early activation marker for T cells and has also been identified as a canonical marker of tissue residency across various tissues. Its high expression in tissue-resident memory T cells has been reported in many tissue sites, including skin, lung, liver, and intestine [200]. In this study, *CD69* was identified as a tissue-common adaptation/residence signature for both VAT and the liver in CD4+ T cells as well as CD8+ T cells. Additionally, the expression of *CD69* was more upregulated in the liver than VAT for CD4+ T cells but not for CD8+ T cells. The potential explanation could be the existence of CD4+CD69- Trm cells in the adipose tissue. As studies have shown that *CD69* is not a definitive marker for T-cell residence, evidenced by the identification of CD69- Trm cells in skin [23]. Additional data and validation are required to further confirm the expression of *CD69* in VAT.

***BATF* as a novel tissue residence marker for VAT**

*BATF*, the AP-1 family transcription factor, was identified as a significant VAT resident marker for CD4+ T cells. It was highly expressed in VAT CD4 Trm cells, with its expression initiating from naive states as indicated by pseudotime analysis. Although *BATF*'s expression in CD8+ T cells was not as significant as CD4+ T cells, it was still higher compared to blood. As a well-known transcription factor in T cell activation, *BATF* has often been studied especially in regulatory T cells and Th17 cells due to its critical function to maintain the development of these cells cells. However, there was limited research about the role of *BATF* associated with tissue adaptation of T cells. Very recently, a cross-tissue paper about CD8+ T cells found *BATF*'s higher expression in skin and adipose tissue than in blood [59], which further validated our finding that *BATF* could be a novel VAT-specific signature of T cells. By cell-cell communication analysis, we further came up with a hypothesis explaining the high expression of *BATF*, which might due to the activation from STAT3 activated by IL6. Future research should go deeper in the role of *BATF* as a tissue-resident marker, especially in adipose tissue, unveiling the specific mechanism in promoting T cell residency within tissues.

### 4.1.2 Naive T cells adapt to tissues

The naive T cell population (both CD4+ and CD8+ naive T cells) was surprisingly distributed also in the tissue sites (liver and VAT), in addition to blood. The existence of naive T cells in tissues has added evidence for the heterogeneity of the naive T cell population. Moreover, by DGE analysis, a lot of genes were found to be significantly upregulated in VAT and the liver than in blood for CD4+ naive T cells, including *CD69*. Although the expression of these adaptation signatures was not as high as what we have seen in CD4 Trm cells, their increasing expressions were demonstrated as naive CD4+ T cells migrated into tissues and became more resident. Trajectory inference has revealed distinct differentiation paths for the tissue-enriched naive CD4 T cells and circulating naive CD4 T cells: tissue-enriched naive T cells tended to stay and evolve into tissue-resident cells, while circulating T cells were more likely to develop into central memory cells. However, these results require further validation through additional data and experimental verification.

In short, these findings have shed light on the adaptive potential of naive CD4+ T cells within tissues, revealing a new layer of naive T cell heterogeneity that had been previously overlooked. However, it still remains unclear how these naive T cells in tissues differ functionally from their circulating counterparts and what specific environmental niches support their residency in tissues. Consequently, future studies regarding naive T cells are suggested to extend to tissue sites outside circulation, considering their presence and roles within various tissue environments.

## 4.2 Challenges of scRNA-seq data analysis

Single-cell RNA sequencing offers many advantages that have revolutionized our understanding of cellular biology, particularly in complex tissues and heterogeneous cell populations. scRNA-seq provides a high-resolution view of gene expression patterns within a tissue, allowing for the identification of distinct cell types, states, and subpopulations that are masked in bulk RNA sequencing. It's well-suited for studying dynamic biological processes, such as immune responses, where the cellular compositions and states can change rapidly over time. By constructing dynamic models on the single-cell level, we can trace the lineage relationships and developmental trajectories of cells in gene expression over time. Moreover, its ability to identify disease

mechanisms and potential therapeutic targets makes it invaluable in the realm of personalized medicine. As such, single-cell RNA sequencing has emerged as a key tool in modern biological and medical research. However, there are several common obstacles when analyzing and interpreting scRNA-seq data [201, 202].

**Dropout Events**

scRNA-seq typically experiences dropout events, where a given gene in a given cell has zero observed expression, leading to the sparsity of scRNA-seq data. Dropout events can come from two different sources: technical noise, where a gene is expressed but not detected by the sequencing technology, and true biological absence of gene expression. Beyond the true biological dropouts, numerous technical limitations can contribute to the data sparsity such as the cell isolation process and the sequencing platform used. For example, droplet-based approaches like 10X Genomics Chromium displayed more severe dropout problems with fewer genes detected than plate-based method Smart-seq2 [203]. The sparsity of scRNA-seq data complicates the analysis and interpretation of gene expression patterns. In order to solve this challenge, imputation methods have been applied to the single-cell RNA-seq data. These methods generally bring some benefits in data recovery, especially those novel algorithms using deep learning techniques such as DISCERN [204], which uses a novel deep generative network to reconstruct missing single-cell gene expression with a reference dataset. However, there are still limitations that remain unsolved and caution is suggested in using imputation techniques as they may artificially enhance the correlations and introduce false positives in downstream analyses, leading to misleading conclusions [205].

**Batch effects**

Batch effects can emerge from various technical variations between different sequencing runs, sample replicates, and so on. This technical noise can strongly distort gene expression profiles, affecting downstream analysis to reveal true biological meanings. Moreover, the high-dimensional nature of single-cell RNA-seq data together with the increasing number of cells sequenced has added additional challenges for the goal of bach removal. Tools like Harmony and Seurat v3 consistently perform well in addressing simple batch correction tasks, while for more complex challenges, these linear embedding models are less effective than the more recent and sophisticated deep learning methods. Among deep learning integration techniques, scVI [206] has become one of the most well-used methods, and it has been shown to perform well

for more complex data integration tasks as well as several other deep learning-based methods [133]. Although many integration methods have demonstrated success in recovering biological differences and removing batch effects, selecting a universally optimal method that fits every scenario is challenging. Each method has its strengths and limitations, making it crucial to consider the specific requirements and characteristics of the dataset at hand. Besides, most batch effect removal methods, including Harmony, primarily generate an integrated low-dimensional embedding without offering a corrected gene expression matrix, which largely restricts the scope of potential downstream applications that can be performed with the integrated data.

**Cell type identification**

The identification of cell types or states from scRNA-seq data often contains two steps: cell clustering and cluster annotation. In addition to the computational challenges of performing clustering on the high-dimensional sparse scRNA-seq data, determining the optimal number of clusters ($K$) poses another challenge. For example, k-means clustering requires the direct specification of ($K$) and other graph-based methods use several parameters which in turn influence the number of clusters. The selection of $K$ usually depends on the researcher's subjective judgment, when screening the cluster presence from the dimensionality-reduced space. Additionally, an automated approach to determine $K$ in scRNA-seq data was proposed [207], by repeatedly perturbing the original data through subsampling. However, such approaches are computationally intensive and may overlook some smaller, rare cell populations.

After cell clustering, annotating these clusters brings additional challenges. Typically, researchers manually identify cell types by focusing on the most highly expressed gene signatures within each cluster. However, cell identities often encompass a spectrum, including various subpopulations, making it challenging to precisely determine their type or state using only transcriptomic data. Moreover, manual annotation produces limited reproducibility, especially in dynamic fields like T cell research where new subsets and states are continually identified. Besides, innovative approaches for cell annotation do not need predefined cell group information. Instead, they utilize reference datasets to train classifiers for automated cell annotation. An example is CellTypist [148], an automated annotation method that leverages classifiers trained on cross-dataset atlases to give robust and high-quality annotations. However, these classifiers are usually not transparent, and they are unsuitable for identifying novel cell populations.

**Trajectory inference**

Trajectory inference aims to reveal the cellular dynamics from the static snapshots scRNA-seq data by ordering cells in a sequential order along the developmental trajectory. To accomplish this, it's necessary to make assumptions about the continuity and directionality of the developmental process, though these assumptions might not always reflect actual biological conditions or be easily validated. Although several trajectory inference methods have been demonstrated the best performance consistently across all evaluation aspects [150], including PAGA and Slingshot, the results from them could still be misleading. Additionally, RNA velocity information enables more accurate reconstruction of the developmental trajectory. The dynamic model used in scVelo is more applicable compared to the steady-state model due to the more relaxed model assumptions. However, the estimated RNA velocity might still contradict prior biological knowledge. As described by Bergen et al. [208], when genes show multiple kinetic regimes across cell populations and lineages or when the observation of cells stay only in steady states, for example, peripheral blood mononuclear cells lacking any transitions between (mature) cell types, current velocity models would result in arbitrary results. Future improvements could include applying multivariate models that consider gene-gene regulation instead of treating each gene independently. Utilizing data from other modalities, such as proteomic data or spatial transcriptomic data, can provide additional information and increase the statistical power.

Beyond the challenges previously mentioned, numerous others persist in scRNA-seq data analysis, including conducting differential gene expression analysis across multiple conditions and analyzing cell-cell communication. Overall, the analysis of single-cell RNA sequencing data is a challenging and quickly evolving field of research, characterized by the rapid development of innovative algorithms. The systematic comparison of these methods is essential for each analytical problem, while choosing the superior method is also challenging, as it requires the selection of comprehensive evaluation metrics, which cover different analysis results with the ground truth.

## 4.3 Single-cell multimodal data integration

Single-cell RNA sequencing offers insights into gene expression profiles for individual cells, while this single modality is challenging to reveal the full picture of the cells' complex gene regulatory networks. The advent of single-cell multimodal techniques has made it possible for the simultaneous measurements of two or more data modalities, which allows a more comprehensive understanding of cellular and molecular biological networks in different tissues.

**Single-cell multimodal techniques**
The single-cell RNA sequencing data, together with the paired TCR sequencing data were analyzed in this study. With the clonotype defined from the T cell receptor sequence and the cell subsets identified from the transcriptome data, integrative clonal analyses were performed, including the clonal overlap analysis between tissues. These analyses generally consider the clonotype as the text and categorical data, revealing their dynamics between different conditions or subsets of T cells. Many widely used tools such as scRepertoire [178] apply a similar strategy for integrative analysis. However, these methods do not achieve complete integration of these two data modalities since they still require separate downstream analysis, thus limiting downstream visualization and analysis.

Besides the paired sequencing of transcriptome and immune cell receptors, emerging single-cell multi-omics technologies now enable the simultaneous profiling of many different modalities from the same cells, including genome, epigenome, transcriptome and proteome [209]. For example, CITE-seq (cellular indexing of transcriptomes and epitopes by sequencing) is used for the joint measurement of surface protein abundance and transcriptome profiles within individual cells. This technology is particularly valuable for identifying complex immune cell populations, such as various T cell subsets. The chromatin accessibility is another critical layer of gene regulation, which can be analyzed using single-cell ATAC-seq (assay for transposase-accessible chromatin using sequencing). Through technologies like the 10x Genomics multiome, the profiling of gene expression and open chromatin from the same cell is possible as well. Additionally, spatial technologies have added an extra dimension to single-cell data by incorporating spatial localization information, which can also be jointly assayed with other omics data, such as spatial ATAC–RNA-seq and spatial CUT&Tag–RNA-seq [210]. Using the spatial coordinates of cells can help resolve

tissue structure and complex cellular interactions. Besides, spatial multi-omics techniques are often combined with traditional histological staining methods, like H&E (Hematoxylin and Eosin) images, to provide extra morphological insights [211].

**Single-cell multimodal integration**

The rapid development of single-cell multimodal technologies is driving the advancement of computational methods for data integration and analysis. These integration methods characterize multiple functional measurements, revealing interactions and connections among different variables, therefore enhancing our understanding of cell phenotype and interactions. While the task of integrating multimodal single-cell data is challenging. One major problem is the missing data or data sparsity which is common in various single-cell sequencing experiments where not all biomolecules are measured in all samples. In addition, the two data modalities may not have a one-to-one feature correspondence as they represent measurements of distinct chemical constructs. On top of that, the two datasets can follow different distributions with different dimensionality. Moreover, datasets are often unpaired, meaning that different modalities are not profiled from the same cells but rather from the same sample or tissue. Consequently, a number of algorithms with different strategies are being developed continually, and each of them comes with specific advantages and limitations.

Many algorithms are designed to integrate the specific pair of data modalities, such as the integration of scRNA-seq & scTCR-seq in this study, and scRNA-seq & scATAC-seq, for which the majority of methods have been developed [212]. Meanwhile, there are also universal frameworks which support the integration across diverse data modalities. For example, the weighted-nearest neighbor (WNN) analysis implemented in Seurat v4 [114] has been demonstrated through the integration of CITE-seq and 10x multiome data. As a graph-based approach, weighted-nearest neighbor constructs neighbor graphs for each individual modality and then creates a common graph from the weighted combination of these modality graphs, which is used to generate an integrated representation for downstream analysis. Although WNN has shown good performance in integrating scATAC-seq and multiome data [213], it's limited by its scalability with very large datasets and is not applicable to image data.

**Multimodal deep learning methods**

Recently, there has been a significant increase in the development of multimodal deep learning (MDL) algorithms for integrating single-cell multimodal data. These MDL

methods have shown remarkable predictive capabilities and provided several advantages over traditional computational approaches, including the ability to model complex, non-linear relationships between modalities and enhanced scalability for handling large datasets [212]. After being trained, deep learning models can potentially be applied to new datasets that are similar to the training data, to predict absent modalities.

The key of these deep learning approaches is to map features from different modalities into a smaller, unified common latent space using multilayer neural network models. Yang, K.D. et.al [214] created a framework capable of integrating data modalities of very different structures by using autoencoders, to map them to a shared latent representation. Autoencoder (AE) is a neural network architecture comprised of an encoder, which maps data to a latent space, and a decoder, which maps it back to the original space, and it is widely used for representation learning. By incorporating autoencoder architectures customized for the specific modalities, this cross-modal autoencoder approach can translate between different data modalities at the single-cell level, such as scRNA-seq data and chromatin imaging data [214]. Similar to this methodology, a variety of deep learning models have been developed using autoencoders to integrate different modalities. GLUE [215] is able to perform triple-omics data integration on the unpaired datasets. This method applies VAE (variational AE) to model each data modality and merges modality-specific feature spaces by a knowledge-based graph using cross-modality regulatory interactions. This deep learning technique has also been used for the complete integration of scRNA-seq and scTCR-seq data in a recent study [216], which highlighted its ability to handle diverse data types including text, and categorical information from TCR-seq data in addition to numeric data type. Furthermore, it's now possible to obtain a tissue sample's single-cell gene expression profile from the histology images, achieved by an adversarial autoencoder-based model trained on the matched samples of both sc/snRNA-seq and H&E staining data [217]. Further studies can explore this framework in different types of tissues, and consider more complicated model structures, such as using cell segmentation in the H&E image data, to improve the model performance.

In summary, deep learning-based models have been widely applied in the integration of single-cell multimodal data integration containing diverse data types. The active development of these techniques is enhancing our understanding of single-cell data and potentially uncovering new biological insights. While despite their great poten-

tial, several challenges remain in these methods, such as the overfitting challenge during model training, particularly for high-dimensional single-cell data. Moreover, there is no consensus or pipeline for the choice of integration methods when analyzing single-cell multimodal data. Various downstream analyses can be employed with experiments fact to validate the power to these methods. Therefore, there is still a wide array of ongoing opportunities in this field for researchers to investigate.

## 4.4 Summary and outlook

**Summary**

The analyses of single-cell RNA-seq and TCR-seq data from four human tissues successfully uncovered their tissue adaptation of immune cells, especially T cells. A variety of data analysis has been performed, including data preprocessing, dimenstioanality reduction, cell clustering and annotation, differential gene expression analysis and cell-cell communication analysis from scRNA-seq data as well as clonal analysis from TCR-seq data. Through unsupervised clustering, all the global cell lineages including T cells, myeloid cells and NK cells were recovered, and then all subsets of T cells were annotated after sub-clustering, which allowed us to go deeper into the tissue adaptation of different subsets of T cells.

With the focus on T cells, particularly CD4+ T cells, tissue T cell adaptation in VAT and liver was mainly investigated in this study. Differential gene expression analysis on the shared clones of T cells revealed lists of genes which represented the tissue common and specific adaptation signatures. These common tissue adaptation signatures indicate the shared mechanism of T cell activation in local tissue environments. While the tissue-specific adaptation signatures are the representation of the underlying tissue specificity, which offer valuable insights into the diverse functions of T cells within different tissue environments. In addition, the adaptation observed in naive CD4+ T cells provides evidence supporting the heterogeneity within this cell population. By trajectory analysis, the tissue adaptation model was proposed, which summarized our findings that T cells gradually change their gene expression profiles in order to adapt to the tissue environment. Moreover, cell-cell communication analysis provided deep insights into the cellular interactions between different T cell subsets and other innate immune cells in tissues, uncovering potential signaling pathways contributing to tissue T cell adaptation. Besides, the small intestine showed unique

cell type compositions with mostly CD8+ cytotoxic T cells enriched. With differential gene expression analysis, these tissue-resident genes in SI CD8+ resident T cells were revealed. The same analysis was also applied on other innate immune cells recovered in the study. As a result, the tissue adaptation of myeloid cells and NK cells were proved, which was evidenced by lists of significantly differentially expressed genes for each tissue site.

In sum, our findings highlighted the adaptability and functional diversity of T cells. The study of tissue adaptation signatures is crucial for advancing our knowledge of the immune system's organization and function within distinct tissue environments, with the great potential of driving novel diagnostic, therapeutic strategies.

**Outlook**

Several limitations exist in this study related to the datasets and the bioinformatics analyses. Firstly, there were in total six donors sequenced among which two failed to provide enough cells for robust downstream analysis. The limited number of samples in the study thus makes it challenging to draw a solid conclusion. Meanwhile, these donors are not completely healthy and their biopsies were obtained during the bariatric surgery. The potential inflammation in their organs could be a confounding factor in the data analysis, influencing the accuracy of the revealed tissue adaptation signatures.

The single-cell RNA-seq data contain mainly T cells and other innate immune cells, with B cells intentionally filtered out during initial cell sorting. T cells as the most important adaptive immune cells, actively interact with other immune cells as well as the non-immune cells in the environment. Due to the data restrictions, the full picture of the T-cell interaction network in tissues cannot be displayed by cell-cell communication analysis in this paper. Therefore, future studies can include various cell types in addition to T cells in their research designs, such as stromal cells, adipocytes from visceral adipose tissue, as well as B cells. This extension of data would allow a comprehensive cell-cell communication analysis between different immune cells or/and with other non-immune cells, revealing how these cells are induced and maintained in the tissues.

In terms of data analyses, although multiple analysis steps were performed on the integrated single-cell RNA-seq data, the TCR-seq data analysis shown in the paper is relatively simple. Indeed, several more advanced analyses with TCR-seq data have

been conducted such as the the gene usage analysis and the TCR graph analysis, however, no novel biological insights can be concluded from these analyses. The novel integration method using deep learning [216] offers a new direction for the complete integration of these two data modalities, which may uncover previously overlooked perspectives.

Additionally, with the increasing availability of single-cell multimodal data and deep learning techniques, the missing modality of the single-cell data can be predicted using the novel deep learning model with a suitable reference. For example, many tissue signatures identified in the study are related to gene regulation, such as the VAT-specific T cell marker *BATF*. In order to further reveal the gene regulatory network behind these signatures, it's now possible to obtain the predicted chromosome accessibility information from the trained model on public scRNA-seq and scATAC-seq datasets. Moreover, multimodal deep-learning techniques for the integration of scRNA-seq data and spatial image data are also available [217]. By applying these advanced techniques to our single-cell RNA-seq data, we could expect to dissolve the spatial organization of the tissue environments.

Finally, there is a lack of additional validation using publicly available data. For instance, to validate tissue-specific adaptation signatures, it would be beneficial to use public datasets that include samples from the same tissue origin as an external benchmark, enhancing the credibility and reliability of these findings. Besides, the study is limited to the tissue of human blood, VAT, liver, and small intestine, the common tissue signatures should be extended to other human organs as well. With more and more cross-tissue research, it's promising to generate a solid list of adaptation signatures for each tissue or organ for human T cells in the future.

# Publications

Here the list includes publications that the author (Tianran Zhang) has been involved in during the doctoral research.

1. Jan Czogalla, Simon Schliffke, Shun Lu, Maria Schwerk, Helena Petereit, **Tianran Zhang**, Shuya Liu, Bernhard Dumoulin, Sydney Gies, Guochao Wu, Sonja Hänzelmann, Marlies Bode, Florian Grahammer, Markus Gödel, Minna Voigtländer, Linus Butt, Carsten Bokemeyer, Carsten Bergmann, Thomas Benzing, Thorsten Wiech, Victor G Puelles, Tobias B Huber. (2024) Ibrutinib-Associated Focal Segmental Glomerulosclerosis and the Impact of Podocin Mutations in Chronic Lymphocytic Leukemia. *Kidney International*, 105(4), 877–881..

2. Sydney E. Gies, Sonja Hänzelmann, Dominik Kylies, Simon Lagies, Moritz Lassé, Fabian Hausmann, Robin Khatri, Manuela Poets, **Tianran Zhang**, Shun Lu, Shuya Liu, Silvia Chilla, Ilka Edenhofer, Jan Czogalla, Fabian Braun, Bernd Kammerer, Markus M. Rinschen, Victor G. Puelles, Stefan Bonn, Maja T. Lindenmeyer, Tobias B. Huber. (2023). A universal preservation protocol for multi-omic and histological analysis of kidney tissue. *bioRxiv*, 2023-08.

3. Shuya Liu, Yu Zhao, Shun Lu, **Tianran Zhang**, Maja T. Lindenmeyer, Viji Nair, Sydney E. Gies, Guochao Wu, Robert G. Nelson, Jan Czogalla, Hande Aypek, Stephanie Zielinski, Zhouning Liao, Melanie Schaper, Damian Fermin, Clemens D. Cohen, Denis Delic, Christian F. Krebs, Florian Grahammer, Thorsten Wiech, Matthias Kretzler, Catherine Meyer-Schwesinger, Stefan Bonn & Tobias B. Huber. (2023). Single-cell transcriptomics reveals a mechanosensitive injury signaling pathway in early diabetic nephropathy. *Genome Medicine*, 15(1), 1-19.

4. Anna Woestemeier, Pasquale Scognamiglio, Yu Zhao, Jonas Wagner, Franziska Muscate, Christian Casar, Francesco Siracusa, Filippo Cortesi, Theodora Agalioti, Simone Müller, Adrian Sagebiel, Leonie Konczalla, Ramez Wahib, Karl-Frederick

Karstens, Anastasios D. Giannou, Anna Duprée, Stefan Wolter, Milagros N. Wong, Anne K. Mühlig, Agata A. Bielecka, Vikas Bansal, **Tianran Zhang**, Oliver Mann, Victor G. Puelles, Tobias B. Huber, Ansgar W. Lohse, Jakob R. Izbicki, Noah W. Palm, Stefan Bonn, Samuel Huber, and Nicola Gagliani. (2023). Multicytokine-producing CD4+ T cells characterize the livers of patients with NASH. *JCI insight*, 8(1), e153831.

# Bibliography


1. Abbas, A., Lichtman, A. & Pillai, S. *Cellular and molecular immunology E-book* (Elsevier Health Sciences, 2014).
2. Nicholson, L. B. The immune system. *Essays in biochemistry* **60,** 275–301 (2016).
3. Shields, A. M., Burns, S. O., Savic, S., Richter, A. G., Anantharachagan, A., Arumugakani, G., Baker, K., Bahal, S., Bermingham, W., Bhole, M., *et al.* COVID-19 in patients with primary and secondary immunodeficiency: the United Kingdom experience. *Journal of Allergy and Clinical Immunology* **147,** 870–875 (2021).
4. Julier, Z., Park, A. J., Briquez, P. S. & Martino, M. M. Promoting tissue regeneration by modulating the immune system. *Acta biomaterialia* **53,** 13–28 (2017).
5. Rosales, C., Uribe-Querol, E., *et al.* Phagocytosis: a fundamental process in immunity. *BioMed research international* **2017** (2017).
6. Cormican, S. & Griffin, M. D. Human monocyte subset distinctions and function: insights from gene expression analysis. *Frontiers in immunology* **11,** 529059 (2020).
7. Gordon, S. & Plüddemann, A. Tissue macrophages: heterogeneity and functions. *BMC biology* **15,** 1–18 (2017).
8. Collin, M. & Bigley, V. Human dendritic cell subsets: an update. *Immunology* **154,** 3–20 (2018).
9. Fu, B., Tian, Z. & Wei, H. Subsets of human natural killer cells and their regulatory effects. *Immunology* **141,** 483–489 (2014).
10. Gray, J. I. & Farber, D. L. Tissue-resident immune cells in humans. *Annual Review of Immunology* **40,** 195–220 (2022).
11. Le, T., Reeves, R. K. & McKinnon, L. R. The functional diversity of tissue-resident natural killer cells against infection. *Immunology* **167,** 28–39 (2022).

12. Hoffman, W., Lakkis, F. G. & Chalasani, G. B cells, antibodies, and more. *Clinical Journal of the American Society of Nephrology* **11,** 137–154 (2016).

13. Szabo, P. A., Miron, M. & Farber, D. L. Location, location, location: Tissue resident memory T cells in mice and humans. *Science immunology* **4,** eaas9673 (2019).

14. Donnelly, R. P., Young, H. A. & Rosenberg, A. S. An overview of cytokines and cytokine antagonists as therapeutic agents. *Annals of the New York Academy of Sciences* **1182,** 1–13 (2009).

15. Sant, P., Rippe, K. & Mallm, J.-P. Approaches for single-cell RNA sequencing across tissues and cell types. *Transcription* **14,** 127–145 (2023).

16. Qi, Q., Liu, Y., Cheng, Y., Glanville, J., Zhang, D., Lee, J.-Y., Olshen, R. A., Weyand, C. M., Boyd, S. D. & Goronzy, J. J. Diversity and clonal selection in the human T-cell repertoire. *Proceedings of the National Academy of Sciences* **111,** 13139–13144 (2014).

17. Van den Broek, T., Borghans, J. A. & Van Wijk, F. The full spectrum of human naive T cells. *Nature Reviews Immunology* **18,** 363–373 (2018).

18. Golubovskaya, V. & Wu, L. Different subsets of T cells, memory, effector functions, and CAR-T immunotherapy. *Cancers* **8,** 36 (2016).

19. Zhu, X. & Zhu, J. CD4 T helper cell subsets and related human immunological disorders. *International journal of molecular sciences* **21,** 8011 (2020).

20. Komai, T., Inoue, M., Okamura, T., Morita, K., Iwasaki, Y., Sumitomo, S., Shoda, H., Yamamoto, K. & Fujio, K. Transforming growth factor-$\beta$ and interleukin-10 synergistically regulate humoral immunity via modulating metabolic signals. *Frontiers in immunology* **9,** 1364 (2018).

21. Sallusto, F., Lenig, D., Förster, R., Lipp, M. & Lanzavecchia, A. Two subsets of memory T lymphocytes with distinct homing potentials and effector functions. *Nature* **401,** 708–712 (1999).

22. Szabo, P. A., Levitin, H. M., Miron, M., Snyder, M. E., Senda, T., Yuan, J., Cheng, Y. L., Bush, E. C., Dogra, P., Thapa, P., *et al.* Single-cell transcriptomics of human T cells reveals tissue and activation signatures in health and disease. *Nature communications* **10,** 4706 (2019).

23. Steinert, E. M., Schenkel, J. M., Fraser, K. A., Beura, L. K., Manlove, L. S., Igyártó, B. Z., Southern, P. J. & Masopust, D. Quantifying memory CD8 T cells reveals regionalization of immunosurveillance. *Cell* **161,** 737–749 (2015).

24. Klicznik, M. M., Morawski, P. A., Höllbacher, B., Varkhande, S. R., Motley, S. J., Kuri-Cervantes, L., Goodwin, E., Rosenblum, M. D., Long, S. A., Brachtl, G., *et al.* Human CD4+ CD103+ cutaneous resident memory T cells are found in the circulation of healthy individuals. *Science immunology* **4,** eaav8995 (2019).

25. Oja, A., Piet, B., Helbig, C., Stark, R., Van Der Zwan, D., Blaauwgeers, H., Remmerswaal, E., Amsen, D., Jonkers, R., Moerland, P., *et al.* Trigger-happy resident memory CD4+ T cells inhabit the human lungs. *Mucosal immunology* **11,** 654–667 (2018).

26. FitzPatrick, M. E., Provine, N. M., Garner, L. C., Powell, K., Amini, A., Irwin, S. L., Ferry, H., Ambrose, T., Friend, P., Vrakas, G., *et al.* Human intestinal tissue-resident memory T cells comprise transcriptionally and functionally distinct subsets. *Cell reports* **34** (2021).

27. Konjar, Š., Ficht, X., Iannacone, M. & Veldhoen, M. Heterogeneity of tissue resident memory T cells. *Immunology Letters* **245,** 1–7 (2022).

28. Thome, J. J., Bickham, K. L., Ohmura, Y., Kubota, M., Matsuoka, N., Gordon, C., Granot, T., Griesemer, A., Lerner, H., Kato, T., *et al.* Early-life compartmentalization of human T cell differentiation and regulatory function in mucosal and lymphoid tissues. *Nature medicine* **22,** 72–77 (2016).

29. Wong, M. T., Ong, D. E. H., Lim, F. S. H., Teng, K. W. W., McGovern, N., Narayanan, S., Ho, W. Q., Cerny, D., Tan, H. K. K., Anicete, R., *et al.* A high-dimensional atlas of human T cell diversity reveals tissue-specific trafficking and cytokine signatures. *Immunity* **45,** 442–456 (2016).

30. Thome, J. J., Grinshpun, B., Kumar, B. V., Kubota, M., Ohmura, Y., Lerner, H., Sempowski, G. D., Shen, Y. & Farber, D. L. Long-term maintenance of human naıve T cells through in situ homeostasis in lymphoid tissue sites. *Science immunology* **1,** eaah6506 (2016).

31. Masopust, D. & Soerens, A. G. Tissue-resident T cells and other resident leukocytes. *Annual review of immunology* **37,** 521–546 (2019).

32. Crowl, J. T., Heeg, M., Ferry, A., Milner, J. J., Omilusik, K. D., Toma, C., He, Z., Chang, J. T. & Goldrath, A. W. Tissue-resident memory CD8+ T cells possess unique transcriptional, epigenetic and functional adaptations to different tissue environments. *Nature immunology* **23,** 1121–1131 (2022).

33. Poon, M. M., Caron, D. P., Wang, Z., Wells, S. B., Chen, D., Meng, W., Szabo, P. A., Lam, N., Kubota, M., Matsumoto, R., *et al.* Tissue adaptation and clonal segregation of human memory T cells in barrier sites. *Nature Immunology* **24,** 309–319 (2023).

34. Kershaw, E. E. & Flier, J. S. Adipose tissue as an endocrine organ. *The Journal of Clinical Endocrinology & Metabolism* **89,** 2548–2556 (2004).

35. Luo, L. & Liu, M. Adipose tissue in control of metabolism. *Journal of endocrinology* **231,** R77–R99 (2016).

36. Cannon, B. & Nedergaard, J. Brown adipose tissue: function and physiological significance. *Physiological reviews* (2004).

37. Schoettl, T., Fischer, I. P. & Ussar, S. Heterogeneity of adipose tissue in development and metabolic function. *Journal of Experimental Biology* **221,** jeb162958 (2018).

38. Foster, M. T. & Pagliassotti, M. J. Metabolic alterations following visceral fat removal and expansion: Beyond anatomic location. *Adipocyte* **1,** 192 (2012).

39. West-Eberhard, M. J. Nutrition, the visceral immune system, and the evolutionary origins of pathogenic obesity. *Proceedings of the National Academy of Sciences* **116,** 723–731 (2019).

40. Porter, S. A., Massaro, J. M., Hoffmann, U., Vasan, R. S., O'Donnel, C. J. & Fox, C. S. Abdominal subcutaneous adipose tissue: a protective fat depot? *Diabetes care* **32,** 1068–1075 (2009).

41. Weinstock, A., Moura Silva, H., Moore, K. J., Schmidt, A. M. & Fisher, E. A. Leukocyte heterogeneity in adipose tissue, including in obesity. *Circulation research* **126,** 1590–1612 (2020).

42. Man, K., Kallies, A. & Vasanthakumar, A. Resident and migratory adipose immune cells control systemic metabolism and thermogenesis. *Cellular & Molecular Immunology* **19,** 421–431 (2022).

43. Ferrante Jr, A. W. The immune cells in adipose tissue. *Diabetes, Obesity and Metabolism* **15,** 34–38 (2013).

44. Hildreth, A. D., Ma, F., Wong, Y. Y., Sun, R., Pellegrini, M. & O'Sullivan, T. E. Single-cell sequencing of human white adipose tissue identifies new cell states in health and obesity. *Nature immunology* **22,** 639–653 (2021).

45. Wang, Q. & Wu, H. T cells in adipose tissue: critical players in immunometabolism. *Frontiers in immunology* **9,** 2509 (2018).

46. Zhou, H. & Liu, F. Regulation, communication, and functional roles of adipose tissue-resident CD4+ T cells in the control of metabolic homeostasis. *Frontiers in immunology* **9,** 1961 (2018).

47. Kintscher, U., Hartge, M., Hess, K., Foryst-Ludwig, A., Clemenz, M., Wabitsch, M., Fischer-Posovszky, P., Barth, T. F., Dragun, D., Skurk, T., *et al.* T-lymphocyte infiltration in visceral adipose tissue: a primary event in adipose tissue inflammation and the development of obesity-mediated insulin resistance. *Arteriosclerosis, thrombosis, and vascular biology* **28,** 1304–1310 (2008).

48. Han, S.-J., Zaretsky, A. G., Andrade-Oliveira, V., Collins, N., Dzutsev, A., Shaik, J., da Fonseca, D. M., Harrison, O. J., Tamoutounour, S., Byrd, A. L., *et al.* White adipose tissue is a reservoir for memory T cells and promotes protective memory responses to infection. *Immunity* **47,** 1154–1168 (2017).

49. Vasanthakumar, A., Moro, K., Xin, A., Liao, Y., Gloury, R., Kawamoto, S., Fagarasan, S., Mielke, L. A., Afshar-Sterle, S., Masters, S. L., *et al.* The transcriptional regulators IRF4, BATF and IL-33 orchestrate development and maintenance of adipose tissue–resident regulatory T cells. *Nature immunology* **16,** 276–285 (2015).

50. Kolodin, D., van Panhuys, N., Li, C., Magnuson, A. M., Cipolletta, D., Miller, C. M., Wagers, A., Germain, R. N., Benoist, C. & Mathis, D. Antigen-and cytokine-driven accumulation of regulatory T cells in visceral adipose tissue of lean mice. *Cell metabolism* **21,** 543–557 (2015).

51. Winer, S., Chan, Y., Paltser, G., Truong, D., Tsui, H., Bahrami, J., Dorfman, R., Wang, Y., Zielenski, J., Mastronardi, F., *et al.* Normalization of obesity-associated insulin resistance through immunotherapy. *Nature medicine* **15,** 921–929 (2009).

52. Fabbrini, E., Cella, M., Mccartney, S. A., Fuchs, A., Abumrad, N. A., Pietka, T. A., Chen, Z., Finck, B. N., Han, D. H., Magkos, F., *et al.* Association between specific adipose tissue CD4+ T-cell populations and insulin resistance in obese individuals. *Gastroenterology* **145,** 366–374 (2013).

53. Chehimi, M., Vidal, H. & Eljaafari, A. Pathogenic role of IL-17-producing immune cells in obesity, and related inflammatory diseases. *Journal of clinical medicine* **6,** 68 (2017).

54. McLaughlin, T., Liu, L.-F., Lamendola, C., Shen, L., Morton, J., Rivas, H., Winer, D., Tolentino, L., Choi, O., Zhang, H., *et al.* T-cell profile in adipose tissue is associated with insulin resistance and systemic inflammation in humans. *Arteriosclerosis, thrombosis, and vascular biology* **34,** 2637–2643 (2014).

55. Nishimura, S., Manabe, I., Nagasaki, M., Eto, K., Yamashita, H., Ohsugi, M., Otsu, M., Hara, K., Ueki, K., Sugiura, S., *et al.* CD8+ effector T cells contribute to macrophage recruitment and adipose tissue inflammation in obesity. *Nature medicine* **15,** 914–920 (2009).

56. Yang, H., Youm, Y.-H., Vandanmagsar, B., Ravussin, A., Gimble, J. M., Greenway, F., Stephens, J. M., Mynatt, R. L. & Dixit, V. D. Obesity increases the production of proinflammatory mediators from adipose tissue T cells and compromises TCR repertoire diversity: implications for systemic inflammation and insulin resistance. *The Journal of Immunology* **185,** 1836–1845 (2010).

57. Wensveen, F. M., Valentić, S., Šestan, M., Turk Wensveen, T. & Polić, B. The "Big Bang" in obese fat: Events initiating obesity-induced adipose tissue inflammation. *European journal of immunology* **45,** 2446–2456 (2015).

58. Vijay, J., Gauthier, M.-F., Biswell, R. L., Louiselle, D. A., Johnston, J. J., Cheung, W. A., Belden, B., Pramatarova, A., Biertho, L., Gibson, M., *et al.* Single-cell analysis of human adipose tissue identifies depot-and disease-specific cell types. *Nature metabolism* **2,** 97–109 (2020).

59. Delacher, M., Schmidleithner, L., Simon, M., Stüve, P., Sanderink, L., Hotz-Wagenblatt, A., Wuttke, M., Schambeck, K., Ruhland, B., Hofmann, V., *et al.* The effector program of human CD8 T cells supports tissue remodeling. *Journal of Experimental Medicine* **221** (2024).

60. Silva, H. M., Báfica, A., Rodrigues-Luiz, G. F., Chi, J., Santos, P. d. A., Reis, B. S., Hoytema van Konijnenburg, D. P., Crane, A., Arifa, R. D. N., Martin, P., *et al.* Vasculature-associated fat macrophages readily adapt to inflammatory and metabolic challenges. *Journal of Experimental Medicine* **216,** 786–806 (2019).

61. Mehta, P., Nuotio-Antar, A. M. & Smith, C. W. $\gamma\delta$ T cells promote inflammation and insulin resistance during high fat diet-induced obesity in mice. *Journal of Leucocyte Biology* **97,** 121–134 (2015).

62. Hu, B., Jin, C., Zeng, X., Resch, J. M., Jedrychowski, M. P., Yang, Z., Desai, B. N., Banks, A. S., Lowell, B. B., Mathis, D., *et al.* $\gamma\delta$ T cells and adipocyte IL-17RC control fat innervation and thermogenesis. *Nature* **578,** 610–614 (2020).

63. Lynch, L., O'Shea, D., Winter, D. C., Geoghegan, J., Doherty, D. G. & O'Farrelly, C. Invariant NKT cells and CD1d+ cells amass in human omentum and are depleted in patients with cancer and obesity. *European journal of immunology* **39,** 1893–1901 (2009).

64. Crosby, C. M. & Kronenberg, M. Tissue-specific functions of invariant natural killer T cells. *Nature Reviews Immunology* **18,** 559–574 (2018).

65. Koethe, J. R., McDonnell, W., Kennedy, A., Abana, C. O., Pilkinton, M., Setliff, I., Georgiev, I., Barnett, L., Hager, C. C., Smith, R., *et al.* Adipose tissue is enriched for activated and late-differentiated CD8+ T cells, and shows distinct CD8+ receptor usage, compared to blood in HIV-infected persons. *Journal of acquired immune deficiency syndromes (1999)* **77,** e14 (2018).

66. Weisberg, S. P., McCann, D., Desai, M., Rosenbaum, M., Leibel, R. L., Ferrante, A. W., *et al.* Obesity is associated with macrophage accumulation in adipose tissue. *The Journal of clinical investigation* **112,** 1796–1808 (2003).

67. Michailidou, Z., Gomez-Salazar, M. & Alexaki, V. I. Innate immune cells in the adipose tissue in health and metabolic disease. *Journal of innate immunity* **14,** 4–30 (2022).

68. Lumeng, C. N., DelProposto, J. B., Westcott, D. J. & Saltiel, A. R. Phenotypic switching of adipose tissue macrophages with obesity is generated by spatiotemporal differences in macrophage subtypes. *Diabetes* **57,** 3239–3246 (2008).

69. Rao, R. R., Long, J. Z., White, J. P., Svensson, K. J., Lou, J., Lokurkar, I., Jedrychowski, M. P., Ruas, J. L., Wrann, C. D., Lo, J. C., *et al.* Meteorin-like is a hormone that regulates immune-adipose interactions to increase beige fat thermogenesis. *Cell* **157,** 1279–1291 (2014).

70. Hill, D. A., Lim, H.-W., Kim, Y. H., Ho, W. Y., Foong, Y. H., Nelson, V. L., Nguyen, H. C., Chegireddy, K., Kim, J., Habertheuer, A., *et al.* Distinct macrophage populations direct inflammatory versus physiological changes in adipose tissue. *Proceedings of the National Academy of Sciences* **115,** E5096–E5105 (2018).

71. Li, C., Menoret, A., Farragher, C., Ouyang, Z., Bonin, C., Holvoet, P., Vella, A. T. & Zhou, B. Single-cell transcriptomics–based MacSpectrum reveals macrophage activation signatures in diseases. *JCI insight* **4** (2019).

72. Pamir, N., Liu, N.-C., Irwin, A., Becker, L., Peng, Y., Ronsein, G. E., Bornfeldt, K. E., Duffield, J. S. & Heinecke, J. W. Granulocyte/macrophage colony-stimulating factor-dependent dendritic cells restrain lean adipose tissue expansion. *Journal of Biological Chemistry* **290,** 14656–14667 (2015).

73. Weinstock, A., Brown, E. J., Garabedian, M. L., Pena, S., Sharma, M., Lafaille, J., Moore, K. J. & Fisher, E. A. Single-cell RNA sequencing of visceral adipose tissue leukocytes reveals that caloric restriction following obesity promotes the accumulation of a distinct macrophage population with features of phagocytic cells. *Immunometabolism* **1** (2019).

74. Wu, D., Molofsky, A. B., Liang, H.-E., Ricardo-Gonzalez, R. R., Jouihan, H. A., Bando, J. K., Chawla, A. & Locksley, R. M. Eosinophils sustain adipose alternatively activated macrophages associated with glucose homeostasis. *Science* **332,** 243–247 (2011).

75. Talukdar, S., Oh, D. Y., Bandyopadhyay, G., Li, D., Xu, J., McNelis, J., Lu, M., Li, P., Yan, Q., Zhu, Y., *et al.* Neutrophils mediate insulin resistance in mice fed a high-fat diet through secreted elastase. *Nature medicine* **18,** 1407–1412 (2012).

76. Liu, J., Divoux, A., Sun, J., Zhang, J., Clément, K., Glickman, J. N., Sukhova, G. K., Wolters, P. J., Du, J., Gorgun, C. Z., *et al.* Genetic deficiency and pharmacological stabilization of mast cells reduce diet-induced obesity and diabetes in mice. *Nature medicine* **15,** 940–945 (2009).

77. O'rourke, R., Metcalf, M., White, A., Madala, A., Winters, B., Maizlin, I., Jobe, B., Roberts, C., Slifka, M. & Marks, D. Depot-specific differences in inflammatory mediators and a role for NK cells and IFN-$\gamma$ in inflammation in human adipose tissue. *International Journal of Obesity* **33,** 978–990 (2009).

78. Haugstøyl, M. E., Cornillet, M., Strand, K., Stiglund, N., Sun, D., Lawrence-Archer, L., Hjellestad, I. D., Busch, C., Mellgren, G., Björkström, N. K., *et al.* Phenotypic diversity of human adipose tissue-resident NK cells in obesity. *Frontiers in Immunology* **14,** 1130370 (2023).

79. Molofsky, A. B., Nussbaum, J. C., Liang, H.-E., Van Dyken, S. J., Cheng, L. E., Mohapatra, A., Chawla, A. & Locksley, R. M. Innate lymphoid type 2 cells sustain visceral adipose tissue eosinophils and alternatively activated macrophages. *Journal of Experimental Medicine* **210,** 535–549 (2013).

80. Winer, D. A., Winer, S., Shen, L., Wadia, P. P., Yantha, J., Paltser, G., Tsui, H., Wu, P., Davidson, M. G., Alonso, M. N., *et al.* B cells promote insulin resistance through modulation of T cells and production of pathogenic IgG antibodies. *Nature medicine* **17,** 610–617 (2011).

81. Cheng, M. L., Nakib, D., Perciani, C. T. & MacParland, S. A. The immune niche of the liver. *Clinical Science* **135,** 2445–2466 (2021).

82. Kubes, P. & Jenne, C. Immune responses in the liver. *Annual review of immunology* **36,** 247–277 (2018).

83. Ficht, X. & Iannacone, M. Immune surveillance of the liver by T cells. *Science immunology* **5,** eaba2351 (2020).

84. Bertolino, P., Bowen, D. G., McCaughan, G. W. & Fazekas de St Groth, B. Antigen-specific primary activation of CD8+ T cells within the liver. *The Journal of Immunology* **166,** 5430–5438 (2001).

85. Bowen, D. G., Zen, M., Holz, L., Davis, T., McCaughan, G. W. & Bertolino, P. The site of primary T cell activation is a determinant of the balance between intrahepatic tolerance and immunity. *The Journal of Clinical Investigation* (Sept. 2004).

86. Pallett, L. J., Davies, J., Colbeck, E. J., Robertson, F., Hansi, N., Easom, N. J., Burton, A. R., Stegmann, K. A., Schurich, A., Swadling, L., *et al.* IL-2high tissue-resident T cells in the human liver: Sentinels for hepatotropic infection. *Journal of Experimental Medicine* **214,** 1567–1580 (2017).

87. Walsh, D. A., Borges da Silva, H., Beura, L. K., Peng, C., Hamilton, S. E., Masopust, D. & Jameson, S. C. The functional requirement for CD69 in establishment of resident memory CD8+ T cells varies with tissue location. *The Journal of Immunology* **203,** 946–955 (2019).

88. Bartsch, L. M., Damasio, M. P., Subudhi, S. & Drescher, H. K. Tissue-Resident memory T cells in the Liver—Unique characteristics of local specialists. *Cells* **9,** 2457 (2020).

89. Kim, J. H., Han, J. W., Choi, Y. J., Rha, M.-S., Koh, J. Y., Kim, K. H., Kim, C. G., Lee, Y. J., Kim, A. R., Park, J., *et al.* Functions of human liver CD69+ CD103- CD8+ T cells depend on HIF-2$\alpha$ activity in healthy and pathologic livers. *Journal of Hepatology* **72,** 1170–1181 (2020).

90. Tse, S.-W., Radtke, A. J., Espinosa, D. A., Cockburn, I. A. & Zavala, F. The chemokine receptor CXCR6 is required for the maintenance of liver memory CD8+ T cells specific for infectious pathogens. *The Journal of infectious diseases* **210,** 1508–1516 (2014).

91. Horst, A. K., Neumann, K., Diehl, L. & Tiegs, G. Modulation of liver tolerance by conventional and nonconventional antigen-presenting cells and regulatory immune cells. *Cellular & molecular immunology* **13,** 277–292 (2016).

92. Wiggins, B. G., Pallett, L. J., Li, X., Davies, S. P., Amin, O. E., Gill, U. S., Kucykowicz, S., Patel, A. M., Aliazis, K., Liu, Y. S., *et al.* The human liver microenvironment shapes the homing and function of CD4+ T-cell populations. *Gut* **71,** 1399–1411 (2022).

93. Muscate, F., Woestemeier, A. & Gagliani, N. *Functional heterogeneity of CD4+ T cells in liver inflammation* in *Seminars in Immunopathology* **43** (2021), 549–561.

94. Poch, T., Krause, J., Casar, C., Liwinski, T., Glau, L., Kaufmann, M., Ahrenstorf, A. E., Hess, L. U., Ziegler, A. E., Martrus, G., *et al.* Single-cell atlas of hepatic T cells reveals expansion of liver-resident naive-like CD4+ T cells in primary sclerosing cholangitis. *Journal of hepatology* **75,** 414–423 (2021).

95. Woestemeier, A., Scognamiglio, P., Zhao, Y., Wagner, J., Muscate, F., Casar, C., Siracusa, F., Cortesi, F., Agalioti, T., Müller, S. & et al. Multicytokine-producing CD4+ T cells characterize the livers of patients with Nash. *JCI Insight* **8** (Jan. 2023).

96. Zhou, Y., Zhang, H., Yao, Y., Zhang, X., Guan, Y. & Zheng, F. CD4+ T cell activation and inflammation in NASH-related fibrosis. *Frontiers in immunology* **13,** 967410 (2022).

97. Salou, M., Legoux, F., Gilet, J., Darbois, A., du Halgouet, A., Alonso, R., Richer, W., Goubet, A.-G., Daviaud, C., Menger, L., *et al.* A common transcriptomic program acquired in the thymus defines tissue residency of MAIT and NKT subsets. *Journal of Experimental Medicine* **216,** 133–151 (2019).

98. Bilzer, M., Roggel, F. & Gerbes, A. L. Role of Kupffer cells in host defense and liver disease. *Liver International* **26,** 1175–1186 (2006).

99. Huang, L.-R., Wohlleber, D., Reisinger, F., Jenne, C. N., Cheng, R.-L., Abdullah, Z., Schildberg, F. A., Odenthal, M., Dienes, H.-P., Van Rooijen, N., *et al.* Intrahepatic myeloid-cell aggregates enable local proliferation of CD8+ T cells and successful immunotherapy against chronic viral liver infection. *Nature immunology* **14,** 574–583 (2013).

100. MacParland, S. A., Liu, J. C., Ma, X.-Z., Innes, B. T., Bartczak, A. M., Gage, B. K., Manuel, J., Khuu, N., Echeverri, J., Linares, I., *et al.* Single cell RNA sequencing of human liver reveals distinct intrahepatic macrophage populations. *Nature communications* **9,** 4383 (2018).

101. Filipovic, I., Sönnerborg, I., Strunz, B., Friberg, D., Cornillet, M., Hertwig, L., Ivarsson, M. A. & Björkström, N. K. 29-color flow cytometry: unraveling human liver NK cell repertoire diversity. *Frontiers in immunology* **10,** 2692 (2019).

102. Lopes, N., Galluso, J., Escaliere, B., Carpentier, S., Kerdiles, Y. M. & Vivier, E. Tissue-specific transcriptional profiles and heterogeneity of natural killer cells and group 1 innate lymphoid cells. *Cell Reports Medicine* **3** (2022).

103. Méndez-Sánchez, N., Córdova-Gallardo, J., Barranco-Fragoso, B. & Eslam, M. Hepatic dendritic cells in the development and progression of metabolic steatohepatitis. *Frontiers in Immunology* **12,** 641240 (2021).

104. Mowat, A. M. & Agace, W. W. Regional specialization within the intestinal immune system. *Nature Reviews Immunology* **14,** 667–685 (2014).

105. Lin, Y. H., Duong, H. G., Limary, A. E., Kim, E. S., Hsu, P., Patel, S. A., Wong, W. H., Indralingam, C. S., Liu, Y. C., Yao, P., *et al.* Small intestine and colon tissue-resident memory CD8+ T cells exhibit molecular heterogeneity and differential dependence on Eomes. *Immunity* **56,** 207–223 (2023).

106. Wang, Z., Gerstein, M. & Snyder, M. RNA-Seq: a revolutionary tool for transcriptomics. *Nature reviews genetics* **10,** 57–63 (2009).

107. Tang, F., Barbacioru, C., Wang, Y., Nordman, E., Lee, C., Xu, N., Wang, X., Bodeau, J., Tuch, B. B., Siddiqui, A., *et al.* mRNA-Seq whole-transcriptome analysis of a single cell. *Nature methods* **6,** 377–382 (2009).

108. Method of the Year 2013. *Nature Methods* **11.** Published 30 December 2013, Issue Date January 2014, 1. `https://doi.org/10.1038/nmeth.2801` (2014).

109. Svensson, V., Vento-Tormo, R. & Teichmann, S. A. Exponential scaling of single-cell RNA-seq in the past decade. *Nature protocols* **13,** 599–604 (2018).

110. Jovic, D., Liang, X., Zeng, H., Lin, L., Xu, F. & Luo, Y. Single-cell RNA sequencing technologies and applications: A brief overview. *Clinical and Translational Medicine* **12,** e694 (2022).

111. Zappia, L., Phipson, B. & Oshlack, A. Exploring the single-cell RNA-seq analysis landscape with the scRNA-tools database. *PLoS computational biology* **14,** e1006245 (2018).

112. Luecken, M. D. & Theis, F. J. Current best practices in single-cell RNA-seq analysis: a tutorial. *Molecular systems biology* **15,** e8746 (2019).

113. Heumos, L., Schaar, A. C., Lance, C., Litinetskaya, A., Drost, F., Zappia, L., Lücken, M. D., Strobl, D. C., Henao, J., Curion, F., *et al.* Best practices for single-cell analysis across modalities. *Nature Reviews Genetics,* 1–23 (2023).

114. Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W. M., Zheng, S., Butler, A., Lee, M. J., Wilk, A. J., Darby, C., Zager, M., *et al.* Integrated analysis of multimodal single-cell data. *Cell* **184,** 3573–3587 (2021).

115. Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: large-scale single-cell gene expression data analysis. *Genome biology* **19,** 1–5 (2018).

116. Virshup, I., Bredikhin, D., Heumos, L., Palla, G., Sturm, G., Gayoso, A., Kats, I., Koutrouli, M., Berger, B., *et al.* The scverse project provides a computational ecosystem for single-cell omics data analysis. *Nature biotechnology* **41,** 604–606 (2023).

117. Lun, A. T., McCarthy, D. J. & Marioni, J. C. A step-by-step workflow for low-level analysis of single-cell RNA-seq data with Bioconductor. *F1000Research* **5** (2016).

118. Amezquita, R., Lun, A., Hicks, S., Gottardo, R. & O'Callaghan, A. *Orchestrating Single-Cell Analysis with Bioconductor* Modified: 2022-09-05. Compiled: 2023-10-25. Environment: R version 4.3.1 (2023-06-16), Bioconductor 3.18. `https://github.com/OSCA-source/OSCA` (Bioconductor, 2022).

119. Andrews, S. *et al. FastQC: a quality control tool for high throughput sequence data* 2010.

120. Parekh, S., Ziegenhain, C., Vieth, B., Enard, W. & Hellmann, I. zUMIs-A fast and flexible pipeline to process RNA sequencing data with UMIs. *Gigascience* **7,** giy059 (2018).

121. Kaminow, B., Yunusov, D. & Dobin, A. STARsolo: accurate, fast and versatile mapping/quantification of single-cell and single-nucleus RNA-seq data. *Biorxiv,* 2021–05 (2021).

122. Xi, N. M. & Li, J. J. Benchmarking computational doublet-detection methods for single-cell RNA sequencing data. *Cell systems* **12,** 176–194 (2021).

123. McGinnis, C. S., Murrow, L. M. & Gartner, Z. J. DoubletFinder: doublet detection in single-cell RNA sequencing data using artificial nearest neighbors. *Cell systems* **8,** 329–337 (2019).

124. Wolock, S. L., Lopez, R. & Klein, A. M. Scrublet: computational identification of cell doublets in single-cell transcriptomic data. *Cell systems* **8,** 281–291 (2019).

125. Ahlmann-Eltze, C. & Huber, W. Comparison of transformations for single-cell RNA-seq data. *Nature Methods,* 1–8 (2023).

126. Mayer, C., Hafemeister, C., Bandler, R. C., Machold, R., Batista Brito, R., Jaglin, X., Allaway, K., Butler, A., Fishell, G. & Satija, R. Developmental diversification of cortical inhibitory interneurons. *Nature* **555,** 457–462 (2018).

127. Büttner, M., Miao, Z., Wolf, F. A., Teichmann, S. A. & Theis, F. J. A test metric for assessing single-cell RNA-seq batch correction. *Nature methods* **16,** 43–49 (2019).

128. Johnson, W. E., Li, C. & Rabinovic, A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* **8,** 118–127 (2007).

129. Korsunsky, I., Millard, N., Fan, J., Slowikowski, K., Zhang, F., Wei, K., Baglaenko, Y., Brenner, M., Loh, P.-r. & Raychaudhuri, S. Fast, sensitive and accurate integration of single-cell data with Harmony. *Nature methods* **16,** 1289–1296 (2019).

130. Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W. M., Hao, Y., Stoeckius, M., Smibert, P. & Satija, R. Comprehensive integration of single-cell data. *Cell* **177,** 1888–1902 (2019).

131. Tran, H. T. N., Ang, K. S., Chevrier, M., Zhang, X., Lee, N. Y. S., Goh, M. & Chen, J. A benchmark of batch-effect correction methods for single-cell RNA sequencing data. *Genome biology* **21,** 1–32 (2020).

132. Chazarra-Gil, R., van Dongen, S., Kiselev, V. Y. & Hemberg, M. Flexible comparison of batch correction methods for single-cell RNA-seq using BatchBench. *Nucleic acids research* **49,** e42–e42 (2021).

133. Luecken, M. D., Büttner, M., Chaichoompu, K., Danese, A., Interlandi, M., Müller, M. F., Strobl, D. C., Zappia, L., Dugas, M., Colomé-Tatché, M., *et al.* Benchmarking atlas-level data integration in single-cell genomics. *Nature methods* **19,** 41–50 (2022).

134. Satija, R., Farrell, J. A., Gennert, D., Schier, A. F. & Regev, A. Spatial reconstruction of single-cell gene expression data. *Nature biotechnology* **33,** 495–502 (2015).

135. Zheng, G. X., Terry, J. M., Belgrader, P., Ryvkin, P., Bent, Z. W., Wilson, R., Ziraldo, S. B., Wheeler, T. D., McDermott, G. P., Zhu, J., *et al.* Massively parallel digital transcriptional profiling of single cells. *Nature communications* **8,** 14049 (2017).

136. Germain, P.-L., Sonrel, A. & Robinson, M. D. pipeComp, a general framework for the evaluation of computational pipelines, reveals performant single cell RNA-seq preprocessing tools. *Genome biology* **21,** 1–28 (2020).

137. Xiang, R., Wang, W., Yang, L., Wang, S., Xu, C. & Chen, X. A comparison for dimensionality reduction methods of single-cell RNA-seq data. *Frontiers in genetics* **12,** 646936 (2021).

138. Van der Maaten, L. & Hinton, G. Visualizing data using t-SNE. *Journal of machine learning research* **9** (2008).

139. McInnes, L., Healy, J. & Melville, J. Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv preprint arXiv:1802.03426* (2018).

140. Kim, T., Chen, I. R., Lin, Y., Wang, A. Y.-Y., Yang, J. Y. H. & Yang, P. Impact of similarity metrics on single-cell RNA-seq data clustering. *Briefings in bioinformatics* **20,** 2316–2326 (2019).

141. Duò, A., Robinson, M. D. & Soneson, C. A systematic performance evaluation of clustering methods for single-cell RNA-seq data. *F1000Research* **7** (2018).

142. Freytag, S., Tian, L., Lönnstedt, I., Ng, M. & Bahlo, M. Comparison of clustering tools in R for medium-sized 10x Genomics single-cell RNA-sequencing data. *F1000Research* **7,** 1297 (2018).

143. Traag, V. A., Waltman, L. & Van Eck, N. J. From Louvain to Leiden: guaranteeing well-connected communities. *Scientific reports* **9,** 5233 (2019).

144. Zhang, X., Lan, Y., Xu, J., Quan, F., Zhao, E., Deng, C., Luo, T., Xu, L., Liao, G., Yan, M., *et al.* CellMarker: a manually curated resource of cell markers in human and mouse. *Nucleic acids research* **47,** D721–D728 (2019).

145. Franzén, O., Gan, L.-M. & Björkegren, J. L. PanglaoDB: a web server for exploration of mouse and human single-cell RNA sequencing data. *Database* **2019,** baz046 (2019).

146. Pasquini, G., Arias, J. E. R., Schäfer, P. & Busskamp, V. Automated methods for cell type annotation on scRNA-seq data. *Computational and Structural Biotechnology Journal* **19,** 961–969 (2021).

147. Pliner, H. A., Shendure, J. & Trapnell, C. Supervised classification enables rapid annotation of cell atlases. *Nature methods* **16,** 983–986 (2019).

148. Domínguez Conde, C., Xu, C., Jarvis, L., Rainbow, D., Wells, S., Gomes, T., Howlett, S., Suchanek, O., Polanski, K., King, H., *et al.* Cross-tissue immune cell analysis reveals tissue-specific features in humans. *Science* **376,** eabl5197 (2022).

149. Lotfollahi, M., Naghipourfar, M., Luecken, M. D., Khajavi, M., Büttner, M., Wagenstetter, M., Avsec, Ž., Gayoso, A., Yosef, N., Interlandi, M., *et al.* Mapping single-cell data to reference atlases by transfer learning. *Nature biotechnology* **40,** 121–130 (2022).

150. Saelens, W., Cannoodt, R., Todorov, H. & Saeys, Y. A comparison of single-cell trajectory inference methods. *Nature biotechnology* **37,** 547–554 (2019).

151. Deconinck, L., Cannoodt, R., Saelens, W., Deplancke, B. & Saeys, Y. Recent advances in trajectory inference from single-cell omics data. *Current Opinion in Systems Biology* **27,** 100344 (2021).

152. Pettie, S. & Ramachandran, V. An optimal minimum spanning tree algorithm. *Journal of the ACM (JACM)* **49,** 16–34 (2002).

153. Wolf, F. A., Hamey, F. K., Plass, M., Solana, J., Dahlin, J. S., Göttgens, B., Rajewsky, N., Simon, L. & Theis, F. J. PAGA: graph abstraction reconciles clustering with trajectory inference through a topology preserving map of single cells. *Genome biology* **20,** 1–9 (2019).

154. Street, K., Risso, D., Fletcher, R. B., Das, D., Ngai, J., Yosef, N., Purdom, E. & Dudoit, S. Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. *BMC genomics* **19,** 1–16 (2018).

155. Perešíni, P., Kuźniar, M. & Kostić, D. *Monocle: Dynamic, fine-grained data plane monitoring* in *Proceedings of the 11th ACM Conference on Emerging Networking Experiments and Technologies* (2015), 1–13.

156. Van den Berge, K., Roux de Bézieux, H., Street, K., Saelens, W., Cannoodt, R., Saeys, Y., Dudoit, S. & Clement, L. Trajectory-based differential expression analysis for single-cell sequencing data. *Nature communications* **11,** 1201 (2020).

157. La Manno, G., Soldatov, R., Zeisel, A., Braun, E., Hochgerner, H., Petukhov, V., Lidschreiber, K., Kastriti, M. E., Lönnerberg, P., Furlan, A., *et al.* RNA velocity of single cells. *Nature* **560,** 494–498 (2018).

158. Bergen, V., Lange, M., Peidli, S., Wolf, F. A. & Theis, F. J. Generalizing RNA velocity to transient cell states through dynamical modeling. *Nature biotechnology* **38,** 1408–1414 (2020).

159. Pullin, J. M. & McCarthy, D. J. A comparison of marker gene selection methods for single-cell RNA sequencing data. *bioRxiv,* 2022–05 (2022).

160. Das, S., Rai, A., Merchant, M. L., Cave, M. C. & Rai, S. N. A comprehensive survey of statistical approaches for differential expression analysis in single-cell RNA sequencing studies. *Genes* **12,** 1947 (2021).

161. Soneson, C. & Robinson, M. D. Bias, robustness and scalability in single-cell differential expression analysis. *Nature methods* **15,** 255–261 (2018).

162. Finak, G., McDavid, A., Yajima, M., Deng, J., Gersuk, V., Shalek, A. K., Slichter, C. K., Miller, H. W., McElrath, M. J., Prlic, M., *et al.* MAST: a flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data. *Genome biology* **16,** 1–13 (2015).

163. Liberzon, A., Subramanian, A., Pinchback, R., Thorvaldsdóttir, H., Tamayo, P. & Mesirov, J. P. Molecular signatures database (MSigDB) 3.0. *Bioinformatics* **27,** 1739–1740 (2011).

164. Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., *et al.* Gene ontology: tool for the unification of biology. *Nature genetics* **25,** 25–29 (2000).

165. Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y. & Morishima, K. KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic acids research* **45,** D353–D361 (2017).

166. Fabregat, A., Jupe, S., Matthews, L., Sidiropoulos, K., Gillespie, M., Garapati, P., Haw, R., Jassal, B., Korninger, F., May, B., *et al.* The reactome pathway knowledgebase. *Nucleic acids research* **46,** D649–D655 (2018).

167. Gable, A. L., Szklarczyk, D., Lyon, D., Matias Rodrigues, J. F. & von Mering, C. Systematic assessment of pathway databases, based on a diverse collection of user-submitted experiments. *Briefings in Bioinformatics* **23,** bbac355 (2022).

168. Boyle, E. I., Weng, S., Gollub, J., Jin, H., Botstein, D., Cherry, J. M. & Sherlock, G. GO:: TermFinder—open source software for accessing Gene Ontology information and finding significantly enriched Gene Ontology terms associated with a list of genes. *Bioinformatics* **20,** 3710–3715 (2004).

169. 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.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences* **102,** 15545–15550 (2005).

170. Armingol, E., Officer, A., Harismendy, O. & Lewis, N. E. Deciphering cell–cell interactions and communication from gene expression. *Nature Reviews Genetics* **22,** 71–88 (2021).

171. Dimitrov, D., Türei, D., Garrido-Rodriguez, M., Burmedi, P. L., Nagai, J. S., Boys, C., Ramirez Flores, R. O., Kim, H., Szalai, B., Costa, I. G., *et al.* Comparison of methods and resources for cell-cell communication inference from single-cell RNA-Seq data. *Nature communications* **13,** 3224 (2022).

172. Jin, S., Guerrero-Juarez, C. F., Zhang, L., Chang, I., Ramos, R., Kuan, C.-H., Myung, P., Plikus, M. V. & Nie, Q. Inference and analysis of cell-cell communication using CellChat. *Nature communications* **12,** 1088 (2021).

173. Efremova, M., Vento-Tormo, M., Teichmann, S. A. & Vento-Tormo, R. CellPhoneDB: inferring cell–cell communication from combined expression of multi-subunit ligand–receptor complexes. *Nature protocols* **15,** 1484–1506 (2020).

174. Liu, Z., Sun, D. & Wang, C. Evaluation of cell-cell interaction methods by integrating single-cell RNA sequencing data with spatial information. *Genome Biology* **23,** 1–38 (2022).

175. Wilk, A. J., Shalek, A. K., Holmes, S. & Blish, C. A. Comparative analysis of cell–cell communication at single-cell resolution. *Nature Biotechnology,* 1–14 (2023).

176. De Simone, M., Rossetti, G. & Pagani, M. Single cell T cell receptor sequencing: techniques and future challenges. *Frontiers in Immunology* **9,** 1638 (2018).

177. Hou, D., Chen, C., Seely, E. J., Chen, S. & Song, Y. High-throughput sequencing-based immune repertoire study during infectious disease. *Frontiers in immunology* **7,** 336 (2016).

178. Borcherding, N., Bormann, N. L. & Kraus, G. scRepertoire: An R-based toolkit for single-cell immune receptor analysis. *F1000Research* **9** (2020).

179. Sturm, G., Szabo, T., Fotakis, G., Haider, M., Rieder, D., Trajanoski, Z. & Finotello, F. Scirpy: a Scanpy extension for analyzing single-cell T-cell receptor-sequencing data. *Bioinformatics* **36,** 4817–4818 (2020).

180. Schuldt, N. J. & Binstadt, B. A. Dual TCR T cells: identity crisis or multitaskers? *The Journal of Immunology* **202,** 637–644 (2019).

181. Dash, P., Fiore-Gartland, A. J., Hertz, T., Wang, G. C., Sharma, S., Souquette, A., Crawford, J. C., Clemens, E. B., Nguyen, T. H., Kedzierska, K., *et al.* Quantifiable predictive features define epitope-specific T cell receptor repertoires. *Nature* **547,** 89–93 (2017).

182. Sureshchandra, S., Lewis, S. A., Doratt, B. M., Jankeel, A., Ibraim, I. C. & Messaoudi, I. Single-cell profiling of T and B cell repertoires following SARS-CoV-2 mRNA vaccine. *JCI insight* **6** (2021).

183. Zhao, Y., Kilian, C., Turner, J.-E., Bosurgi, L., Roedl, K., Bartsch, P., Gnirck, A.-C., Cortesi, F., Schultheiß, C., Hellmig, M., *et al.* Clonal expansion and activation of tissue-resident memory-like TH17 cells expressing GM-CSF in the lungs of patients with severe COVID-19. *Science Immunology* **6,** eabf6692 (2021).

184. Xie, J., Jeon, H., Xin, G., Ma, Q. & Chung, D. LRT: Integrative analysis of scRNA-seq and scTCR-seq data to investigate clonal differentiation heterogeneity. *PLOS Computational Biology* **19,** e1011300 (2023).

185. Schattgen, S. A., Guion, K., Crawford, J. C., Souquette, A., Barrio, A. M., Stubbington, M. J., Thomas, P. G. & Bradley, P. Integrating T cell receptor sequences and transcriptional profiles by clonotype neighbor graph analysis (CoNGA). *Nature biotechnology* **40,** 54–63 (2022).

186. James, K. R., Elmentaite, R., Teichmann, S. A. & Hold, G. L. Redefining intestinal immunity with single-cell transcriptomics. *Mucosal Immunology* **15,** 531–541 (2022).

187. Han, X., Zhou, Z., Fei, L., Sun, H., Wang, R., Chen, Y., Chen, H., Wang, J., Tang, H., Ge, W., *et al.* Construction of a human cell landscape at single-cell level. *Nature* **581,** 303–309 (2020).

188. Mulder, K., Patel, A. A., Kong, W. T., Piot, C., Halitzki, E., Dunsmore, G., Khalilnezhad, S., Irac, S. E., Dubuisson, A., Chevrier, M., *et al.* Cross-tissue single-cell landscape of human monocytes and macrophages in health and disease. *Immunity* **54,** 1883–1900 (2021).

189. Wu, T., Hu, E., Xu, S., Chen, M., Guo, P., Dai, Z., Feng, T., Zhou, L., Tang, W., Zhan, L., *et al.* clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *The innovation* **2** (2021).

190. Jin, S., Plikus, M. V. & Nie, Q. CellChat for systematic analysis of cell-cell communication from single-cell and spatially resolved transcriptomics. *BioRxiv,* 2023–11 (2023).

191. Zhang, Z., Liao, Q., Pan, T., Yu, L., Luo, Z., Su, S., Liu, S., Hou, M., Li, Y., Damba, T., *et al.* BATF relieves hepatic steatosis by inhibiting PD1 and promoting energy metabolism. *Elife* **12,** RP88521 (2023).

192. Chalise, J. P., Hashimoto, S., Parajuli, G., Kang, S., Singh, S. K., Gemechu, Y., Metwally, H., Nyati, K. K., Dubey, P. K., Zaman, M. M.-U., *et al.* Feedback regulation of Arid5a and Ppar-$\gamma$2 maintains adipose tissue homeostasis. *Proceedings of the National Academy of Sciences* **116,** 15128–15133 (2019).

193. Nyati, K. K., Agarwal, R. G., Sharma, P. & Kishimoto, T. Arid5a regulation and the roles of Arid5a in the inflammatory response and disease. *Frontiers in immunology* **10,** 2790 (2019).

194. Desterke, C. & Chiappini, F. Lipid related genes altered in NASH connect inflammation in liver pathogenesis progression to HCC: a canonical pathway. *International journal of molecular sciences* **20,** 5594 (2019).

195. Uehara, S., Grinberg, A., Farber, J. M. & Love, P. E. A role for CCR9 in T lymphocyte development and migration. *The Journal of Immunology* **168,** 2811–2819 (2002).

196. Papadakis, K. A., Prehn, J., Nelson, V., Cheng, L., Binder, S. W., Ponath, P. D., Andrew, D. P. & Targan, S. R. The role of thymus-expressed chemokine and its receptor CCR9 on lymphocytes in the regional specialization of the mucosal immune system. *The Journal of Immunology* **165,** 5069–5076 (2000).

197. Ge, Y.-j., Liao, Q.-w., Xu, Y.-c., Zhao, Q., Wu, B.-l. & Ye, R. D. Anti-inflammatory signaling through G protein-coupled receptors. *Acta Pharmacologica Sinica* **41,** 1531–1538 (2020).

198. Shehata, H. M., Khan, S., Chen, E., Fields, P. E., Flavell, R. A. & Sanjabi, S. Lack of Sprouty 1 and 2 enhances survival of effector CD8+ T cells and yields more protective memory cells. *Proceedings of the National Academy of Sciences* **115,** E8939–E8947 (2018).

199. Ohteki, T., Yoshida, H., Matsuyama, T., Duncan, G. S., Mak, T. W. & Ohashi, P. S. The Transcription Factor Interferon Regulatory Factor 1 (IRF-1) Is Important during the Maturation of Natural Killer 1.1+ T Cell Receptor–$\alpha/\beta$+(NK1+ T) Cells, Natural Killer Cells, and Intestinal Intraepithelial T Cells. *The Journal of experimental medicine* **187,** 967–972 (1998).

200. Kumar, B. V., Ma, W., Miron, M., Granot, T., Guyer, R. S., Carpenter, D. J., Senda, T., Sun, X., Ho, S.-H., Lerner, H., *et al.* Human tissue-resident memory T cells are defined by core transcriptional and functional signatures in lymphoid and mucosal sites. *Cell reports* **20,** 2921–2934 (2017).

201. Lähnemann, D., Köster, J., Szczurek, E., McCarthy, D. J., Hicks, S. C., Robinson, M. D., Vallejos, C. A., Campbell, K. R., Beerenwinkel, N., Mahfouz, A., *et al.* Eleven grand challenges in single-cell data science. *Genome biology* **21,** 1–35 (2020).

202. Yuan, G.-C., Cai, L., Elowitz, M., Enver, T., Fan, G., Guo, G., Irizarry, R., Kharchenko, P., Kim, J., Orkin, S., *et al.* Challenges and emerging directions in single-cell analysis. *Genome biology* **18,** 1–8 (2017).

203. Wang, X., He, Y., Zhang, Q., Ren, X. & Zhang, Z. Direct comparative analyses of 10X genomics chromium and smart-seq2. *Genomics, Proteomics and Bioinformatics* **19,** 253–266 (2021).

204. Hausmann, F., Ergen, C., Khatri, R., Marouf, M., Hänzelmann, S., Gagliani, N., Huber, S., Machart, P. & Bonn, S. DISCERN: deep single-cell expression reconstruction for improved cell clustering and cell subtype and state detection. *Genome Biology* **24,** 212 (2023).

205. Cheng, Y., Ma, X., Yuan, L., Sun, Z. & Wang, P. Evaluating imputation methods for single-cell RNA-seq data. *BMC bioinformatics* **24,** 302 (2023).

206. Lopez, R., Regier, J., Cole, M. B., Jordan, M. I. & Yosef, N. Deep generative modeling for single-cell transcriptomics. *Nature methods* **15,** 1053–1058 (2018).

207. Liu, S., Thennavan, A., Garay, J. P., Marron, J. & Perou, C. M. MultiK: an automated tool to determine optimal cluster numbers in single-cell RNA sequencing data. *Genome biology* **22,** 1–21 (2021).

208. Bergen, V., Soldatov, R. A., Kharchenko, P. V. & Theis, F. J. RNA velocity—current challenges and future perspectives. *Molecular systems biology* **17,** e10282 (2021).

209. Zhu, C., Preissl, S. & Ren, B. Single-cell multimodal omics: the power of many. *Nature methods* **17,** 11–14 (2020).

210. Zhang, D., Deng, Y., Kukanja, P., Agirre, E., Bartosovic, M., Dong, M., Ma, C., Ma, S., Su, G., Bao, S., *et al.* Spatial epigenome–transcriptome co-profiling of mammalian tissues. *Nature* **616,** 113–122 (2023).

211. Vandereyken, K., Sifrim, A., Thienpont, B. & Voet, T. Methods and applications for single-cell and spatial multi-omics. *Nature Reviews Genetics* **24,** 494–515 (2023).

212. Athaya, T., Ripan, R. C., Li, X. & Hu, H. Multimodal deep learning approaches for single-cell multi-omics data integration. *Briefings in Bioinformatics* **24,** bbad313 (2023).

213. Lee, M. Y., Kaestner, K. H. & Li, M. Benchmarking algorithms for joint integration of unpaired and paired single-cell RNA-seq and ATAC-seq data. *Genome Biology* **24,** 244 (2023).

214. Yang, K. D., Belyaeva, A., Venkatachalapathy, S., Damodaran, K., Katcoff, A., Radhakrishnan, A., Shivashankar, G. & Uhler, C. Multi-domain translation between single-cell imaging and sequencing data using autoencoders. *Nature communications* **12,** 31 (2021).

215. Cao, Z.-J. & Gao, G. Multi-omics single-cell data integration and regulatory inference with graph-linked embedding. *Nature Biotechnology* **40,** 1458–1466 (2022).

216. Zhu, B., Wang, Y., Ku, L.-T., van Dijk, D., Zhang, L., Hafler, D. A. & Zhao, H. scNAT: a deep learning method for integrating paired single-cell RNA and T cell receptor sequencing profiles. *Genome Biology* **24,** 292 (2023).

217. Comiter, C., Vaishnav, E. D., Ciapmricotti, M., Li, B., Yang, Y., Rodig, S. J., Turner, M., Pfaff, K. L., Jané-Valbuena, J., Slyper, M., *et al.* Inference of single cell profiles from histology stains with the Single-Cell omics from Histology Analysis Framework (SCHAF). *BioRxiv,* 2023–03 (2023).

# Acronyms

**AE** autoencoder. 113

**APCs** antigen-presenting cells. 5

**AT** adipose tissue. 14, 15, 17, 18

**CC** cell-cell communication. 36

**CPM** counts per million. 27

**CTLs** cytolytic T cells. 7, 10

**DCs** dendritic cells. 5, 6, 17, 23

**DEG** differentially expressed genes. 34

**DGE** differential gene expression. 31, 34, 68, 69, 86, 146, 147

**gd T** gamma delta T. 57

**GO** Gene Ontology. 35

**GRN** gene regulatory network. 37

**GSEA** gene set enrichment analysis. 35

**HC** hepatocytes. 19

**HVG** highly variable genes. 29, 46, 47

**ILCs** type 2 innate lymphoid cells. 18

**KCs** Kupffer cells. 19, 21

**KEGG** Kyoto Encyclopedia of Genes and Genomes. 35

**KNN** K-Nearest Neighbour. 30, 31

**MAIT** mucosal-associated invariant T. 7, 12, 23

**MHC** major histocompatibility complex. 7, 9, 10, 21, 37

**NK cells** natural killer cells. 5, 18, 21, 55

**ORA** Over representation analysis. 35

**PAGA** Partition-based graph abstraction. 32, 33, 52

**PCA** Principle Component Analysis. 29, 32, 47

**QC** quality control. 26, 27, 29, 46

**RNA-seq** RNA sequencing. 23

**scRNA-seq** single-cell RNA sequencing. 24, 27–30, 34, 36, 37, 39, 40, 44, 55, 107

**scTCR-seq** single-cell TCR sequencing. 37, 38, 40, 55

**SI** small intestine. 22, 55, 91

**snRNA-seq** single-nucleus RNA sequencing. 24

**t-SNE** t-distributed stochastic neighbor embedding. 29

**Tcm** central memory T. 12

**TCR** T cell receptors. 7, 8, 12, 13, 17, 37, 44, 53, 59

**TCR-seq** T cell receptor sequencing. 39

**Teff** effector T. 10

**Tem** effector memory T. 12

**Tfh** follicular helper T. 11

**TI** trajectory inference. 32

**Tregs** regulatory T cells. 11, 16

**Trm** resident memory T. 12, 13

**UMAP** uniform manifold approximation and projection. 30, 31, 39, 47, 48, 52, 74

**VAT** visceral adipose tissue. 15–18, 55, 63, 106

**WNN** weighted-nearest neighbor. 112

# List of Figures


1 **Overview of scRNA-seq analysis steps.** Image modified from [113]. . 25

2 **Overview of single-cell experimental design** . . . . . 55

3 **Overview of global cell lineages.** A: UMAP of initial clustering revealing three global cell lineages: T cells, Myeloid cells and NK/Innate cells and the Heatmap of the mean expression of cell type markers. B: Cell numbers of each global cell lineage in different samples . . . . . 56

4 **Cluster annotation of T cell subsets.** A: UMAP showing 16 T cell sub-types including 6 CD4+ T cells, 6 CD8+ T cells, three mixed T cell groups, and gd T cells. B: Heatmap of the core markers for T cell sub-type annotation. . . . . 58

5 **Distribution of T cell sub-types in blood, VAT, liver, and SI.** A: UMAPs showing T cell cluster annotations in each tissue: blood, VAT, liver, and SI. B: Heatmap of cell type proportions in different samples. . . . . 59

6 **The presence of alpha-beta TCR in each T cell group** . . . . . 60

7 **TCR clonal expansion.** The clonal expansion of T cell receptors was classified into different level of groups: Hyperexpanded (100 < X), Large (20 < X <= 100), Medium (5 < X <= 20), Small (1 < X <= 5), and Single (0 < X <= 1). UMAP with the clonal expansion groups was shown (left) and the stacked barplot (right) presented the number of cells for each tissue in every T cell sub-type. . . . . 61

8 **TCR overlap between tissues.** The clonotypes from clonally expanded CD4+ T cells (left) and CD8+ T cells (right) were compared for overlapping between tissues. The connected path within the circle plot represents shared clonotypes and the width of the connection indicates the number of shared clones. . . . . 62

9 **Adaptation of CD4+ T cells by shared clonotypes.** All the comparisons here were based on the cells with shared clonotypes. A: Volcano plots of the DGE results for VAT vs blood (left), liver vs blood (middle), and liver vs VAT (right). B: Upset plot showing the intersection of different DEG sets: "Liver_high" represents the upregulated genes comparing liver to VAT, "Liver_up" represents the upregulated genes comparing liver to blood, and similar for "VAT_high" and "VAT_up". C: Top 10 enriched pathways of the tissue-common adaptation genes using the Reactome database. . . . . 64
10 **Tissue adaptation signatures in all CD4+ T cells.** . . . . . 67
11 **Tissue signatures of CD4 Trm cells.** A-B: Volcano plots showing the DGE results of CD4+ Trm cells in VAT (A) and liver (B) compared with all other CD4+ T cells in the data. C: Venn diagram of the VAT resident genes and liver resident genes. D: Dotplot of key markers of tissue-resident signatures . . . . . 69
12 **Sub-clustering of CD4+ T cells.** A: UMAP plot of CD4+ T cells sub-clustering annotation. B: Heamap showing the expression of core naive and resident markers of CD4+ T cells. . . . . 72
13 **CD95 and CD62L gating of naive CD4+ T cells.** . . . . . 73
14 **Trajectory inference of CD4+ T cells.** A: UMAP plot of velocity projection. B: PAGA graph connecting different T cell groups. The weights of the edges represented the confidence of the connection between the two groups. . . . . 74
15 **Trajectory inference of tissue CD4+ T cell clusters.** A: UMAP of latent time in tissue-enriched clusters (Naive (3), Intermediate, Trm (1), Trm(2), and Th17 cells). B: Directed PAGA graph for the tissue trajectory. C: Heatmap showing top-ranked driving genes in every tissue-enriched cluster. . . . . 76
16 **Scatter plots of selected DE genes along latent time.**: The first two rows represented genes that expressed more in VAT and the others expressed more in liver. . . . . 77
17 **Tissue adaptation model.** . . . . . 78
18 **Volcano plots of DGE results in naive CD4+ T cells.** A-B: Volcano plots showing the DGE results of non-expanded CD4+ naive T cells in VAT (A) and liver (B) compared with cells from blood. . . . . 79

19 **Overview of cell-cell communication between CD4+ T cells and other non-T cells** A: number of interactions between CD4+ T cell, NK cells and Myeloid cells in VAT (left) and liver (right) respectively. B: The information flow was defined by the sum of communication probability among all pairs of cell groups in the inferred network. The relative information flow in liver and VAT was shown in the stacked barplot. . 83
20 **IL6 signaling network** . . . . . 84
21 **IL6 signaling experimental validation.** A: qPCR of the receptor and ligands of IL6 signaling, including IL6, IL6R, and IL6ST. B: Flow cytometry analysis of BATF expression in naive CD4 T cells along time with or without IL6 stimulation. The fold change is measured compared to time zero. . . . . 84
22 **Adaptation of CD8+ T cells with shared clonotypes.** All the comparisons here were based on the cells with shared clonotypes. A: Volcano plots of the DGE results for VAT vs blood (left), liver vs blood (middle), and liver vs VAT (right). B: Upset plot showing the intersection of different DEG sets: "Liver_high" represents the upregulated genes comparing liver to VAT, "Liver_up" represents the upregulated genes comparing liver to blood, and similar for "VAT_high" and "VAT_up". C: Top 10 enriched pathways of the tissue-common adaptation genes using the Reactome database. . . . . 86
23 **Sub-clustering of CD8+ T cells.** A-B: UMAP of all CD8+ T cells colored by tissue (A) and cluster annotations (B). C: Heatmap showing the top 5 markers of each sub-type of CD8+ T cells. . . . . 90
24 **DEGs of SI CD8 Trm cells**. . . . . 92
25 **Tissue adaptation of NK cells.** A: Heatmap showing DEGs comparing each tissue to others. B: Venn diagram of the intersection of tissue DEG lists. . . . . 96
26 **Tissue adaptation of Myeloid cells.** A: Heatmap showing DEG comparing each tissue to others. B: Venn diagram of the intersection of tissue DEG lists. . . . . 99

# List of Tables


1 Donor information . . . . . 43
2 Core marker genes of T cell types . . . . . 49

3 Tissue common adaptation signatures of CD4+ T cells . . . . . 65
4 Tissue-specific adaptation signatures of CD4+ T cells . . . . . 66
5 Common tissue resident signatures of CD4+ Trm cells . . . . . 70
6 Tissue-specific resident signatures of CD4+ Trm cells . . . . . 71
7 Upregulated DEGs in VAT naive CD4 T cells . . . . . 80
8 Upregulated DEGs in Liver naive CD4 T cells . . . . . 81
9 Tissue common adaptation signatures of CD8+ T cells . . . . . 87
10 Tissue-specific adaptation signatures of CD8+ T cells . . . . . 88
11 DEGs in small intestine CD8+ Trm cells . . . . . 93
12 Tissue adaptation signatures of NK cells . . . . . 97
13 Tissue adaptation signatures of myeloid cells . . . . . 100

# Acknowledgments

Firstly, I want to give my deepest appreciation to my doctoral supervisor, Prof. Dr. Stefan Bonn, who welcomed me to join his excellent research team and provided me with the opportunity to deepen my expertise in bioinformatics. His guidance and feedback during my doctoral research were crucial to the development of my academic skills and the completion of this dissertation.

Furthermore, I am very grateful to my collaborators Prof. Dr. Nicola Gagliani and his team members Jonas Wagner and Guido Rattay, for working together on this interesting research project. They not only conducted web lab experiments and provided the source of datasets for this project, but also enriched my understanding of the biological implications of our findings. Without their data and biological guidance, I wouldn't have had the chance to complete my dissertation with this novel project.

Besides, I want to thank my colleagues at the IMSB, especially Dr. Yu Zhao, Dr. Sonja Hänzelman, and Dr. Zeba Sultana for their valuable suggestions, empathy, and encouragement during my entire doctoral study. I am deeply appreciative of their efforts in creating such a productive and supportive work environment that made my time not only educational but also enjoyable.

Lastly, I would like to give my heartful thanks to my dearest friends Alexander Heuer and Te Liu, for their unwavering emotional encouragement whenever I was depressed or anxious during the completion of my thesis. My great appreciation also goes to my parents for their care and selfless support throughout my academic pursuits in Germany.

# Eidesstattliche Erklärung

Hiermit erkläre ich an Eides statt, dass ich die vorliegende Dissertationsschrift selbst verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe.

Ort, Datum Hamburg, 2024 Signature 