UNIVERSITY OF SOUTHAMPTON

FACULTY OF Medicine

Cancer Sciences

Investigation into the heterogeneity of tissue resident T cell populations in human solid tumours.

by

James Clarke

Thesis for the degree of Doctor of Philosophy

31st September 2018

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

The role of the immune system in the control of malignancies is now well established. The presence of cytotoxic T cells (CTLs) has been shown to be prognostic in a wide range of malignancies. The extent of the heterogeneity of the anti-tumour between discrete locations in a single patient's tumour and across multiple patient's is currently unknown. Therefore, this thesis aimed to investigate the biological variability in immune signatures, both within and between patients.

In the first section of this thesis I investigated the intra-patient variability in the anti-tumour immune response, over time and at different locations within the tumour, in Head and Neck Squamous Cell Carcinoma (HNSCC) (Chapter 3). This data suggested that at a global level the immunological signature of each patient was similar, although from preliminary single cell analysis there appeared to be significantly heterogeneity masked in the bulk signature. Concurrently, our group had generated a transcriptome profiling of $CD8^+$ T cells isolated from human Lung Tumours, which suggested a significant proportion were tissue resident ($CD103^+$) $CD8^+$ cells. Therefore, the next stage was to complete flow cytometry investigations into the protein expression of these cells (Chapter 4). This analysis suggested that $CD103^+$ $CD8^+$ cells represented a population of tissue resident T cells in human solid tumours, RNA-seq of purified populations (Chapter 5) and single cell RNA-seq (Chapter 6) was completed to understand this population in more detail. This further transcriptomic profiling highlighted a unique population of tumour specific T_{RM} that appeared to be highly proliferative and expressed superior functional properties (cytokines, chemokines and cytotoxic molecules).

In conclusion, we have begun to clarify the signature of these prognostic tissue resident cells in human anti-tumoural immunity at the bulk and single cell level, highlighting a unique subpopulation. Future work will aim to complete functional investigates and additional genome-wide methodologies to study these population in-vivo.

FACULTY OF MEDICINE

Cancer Immunology

Thesis for the degree of Doctor of Philosophy

Investigation into the heterogeneity of Tissue Resident T cell populations in human solid tumours.

James Clarke

Table of Contents

List of Tables

List of Figures

Declaration of Authorship

I, James Clarke

declare that this thesis and the work presented in it are my own and has been generated by me as the result of my own original research.

Investigation into the heterogeneity of tissue resident T cell populations in human solid tumours.

I confirm that:

1. This work was done wholly or mainly while in candidature for a research degree at this University;
2. Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
3. Where I have consulted the published work of others, this is always clearly attributed;
4. Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
5. I have acknowledged all main sources of help;
6. Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
7. Parts of this work, or methodologies, used have been published:
8. Mellone, M., Hanley, C.J., Thirdborough, S., Mellows, T., Garcia, E., Woo, J., Tod, J., Frampton, S., Jenei, V., Moutasim, K.A.,… **Clarke, J.,**… et al., (2016). Induction of fibroblast senescence generates a non-fibrogenic myofibroblast phenotype that differentially impacts on cancer prognosis. Aging. *9*, 114–132.

   Layfield, D., **Clarke, J.,** Savelyeva, N., Cutress, R., and Ottensmeier, C. (2016). Response to vaccination via tumour draining lymph nodes does not demonstrate a tumour suppressive effect in patients with breast cancer: A peri-surgical window study. Eur. J. Surg. Oncol. *42*, S43.

   Wood, O*., **Clarke, J*.,** Woo, J., Mirza, A.H., Woelk, C.H., Thomas, G.J., Vijayanand, P., King, E., and Ottensmeier, C.H. (2017). Head and neck squamous cell carcinomas are characterized by a stable immune signature within the primary tumor over time and space. Clin. Cancer Res. *23*, 7641–7649. * Designates joint equal contribution

   Ganesan, A.-P., **Clarke, J.,** Wood, O., Garrido-Martin, E.M., Chee, S.J., Mellows, T., Samaniego-Castruita, D., Singh, D., Seumois, G., Alzetani, A., et al., (2017). Tissue-resident memory features are linked to the magnitude of cytotoxic T cell responses in human lung cancer. Nat. Immunol. *18*, 940–950.

Patil, V.S., Madrigal, A., Schmiedel, B.J., **Clarke, J.,** O'Rourke, P., de Silva, A.D., Harris, E., Peters, B., Seumois, G., Weiskopf, D., et al. (2018). Precursors of human CD4 + cytotoxic T lymphocytes identified by single-cell transcriptome analysis. Sci. Immunol. *3*, 8664.

Johnston, H.E., Carter, M.J., Larrayoz, M., **Clarke, J.,** Garbis, S.D., Oscier, D., Strefford, J.C., Steele, A.J., Walewska, R., and Cragg, M.S. (2018). Proteomics Profiling of CLL Versus Healthy B-cells Identifies Putative Therapeutic Targets and a Subtype-independent Signature of Spliceosome Dysregulation. Mol. Cell. Proteomics *17*, 776–791.

Currently under review.

Clarke, J., Panwar, B., Madrigal, A., Singh, D., Wood, O., Chee, SJTW., Awad, AS., King, E., Alzetani, A., Seumois, G., Manuscript under Review.

Garrido-Martin, E.M., Mellows, T.W.P., **Clarke, J.,** Ganesan, A.-P., Wood, O., Cazaly, A., Seumois, G., Chee, S.J., Alzetani, A., King, E., et al., Manuscript under Review.

Singh, D., Ganesan, A.-P., Panwar, B., Hanley, C.J., Madrigal, A., Suastegui, C.R., **Clarke, J.,** Wood, O., Garrido-Martin, E.M., Chee, S.J., Seumois, G., et al., Manuscript under Review.

Signed:

Date: 31st September 2018

Acknowledgements and dedications

The completion of this thesis work would not have been possible without a number of people who have contributed to this work.

Primarily this includes Professor Christian Ottensmeier and Professor Pandurangan Vijayanand who have provided support, leadership and development that was vital for this project.

In Southampton University and La Jolla institute, the training and support has primarily been given by Mr Wood, Professor Friedman and Ms Patil; who I would like to express my sincere gratitude to for their training, emotional support and scientific input.

This work would be impossible without a large and exceptionally dedicated team which has been interfaced by Dr Chee, Mr Johnson and others, with the help large number of surgeons and pathologists who have been vital. Importantly none of this work could be completed without the willingness of the patients to provide material. Although Professor Thomas, Prof Ay, and Ms King are not named supervisors of this project their input has been vital throughout.

This work is dedicated to several members of the Clarke family who have suffered from the diseases discussed here during this project.

Chapter 1: Literature review & research Focus.

1.1 Generation of a naïve T cell.

1.1.1 Introduction to Immunology and the adaptive immune response.

The presence of an immune system is fundamental for survival of every living multi-cellular organism. Both plants and animals have co-evolved with continuous challenge from pathogens such as viruses, bacteria, fungi and protozoa. This selection pressure has culminated in the development of a highly heterogeneous and remarkable 'defence system' in all animals; with the more complex organisms having more complex defence systems (Davis, 2012). The 'defence function' of the immune system has many highly specialised groups of cells each contributing to the overall process in unique ways. While all cell types have an innate capacity to defend themselves against microbial attack, the initial cellular response is not specific to a particular microbe but is activated by a variety of molecular patterns that the cell senses are "dangerous" (discussed later). This so-called innate immune response generates rapid protective mediators. In addition to the universal self-protective responses, other more specialised cells and proteins such as granulocytes, NK cells, macrophages and soluble defence proteins augment and enhance the effectiveness of the innate immune response (Holgate, 2012). The innate immune responses form a first line of defence responsible for containing or eliminating microbial invaders in the early days of an infection (Gajewski et al., 2013).

However, the overall immune defences are supplemented by the slower but more specific mechanisms of the adaptive immune response. The adaptive immune responses are effected by various types of T lymphocyte and also various classes of antibodies produced by B lymphocytes. The cellular components of the adaptive immune system reside in lymphoid organs, but once activated they enter the circulation and traffic through tissues and organs in order to deliver their effector mechanisms at the localizations at which they are required (Förster et al., 2008).

Classically, T cells are the initial responders and controllers of the adaptive immune response. However, before they are activated by contact with their cognate antigen(s) these cells are somewhat catatonic, termed naïve T cells. Only specialized so-called "professional" antigen-presenting cells (APCs) possess the machinery capable of giving the series of activation signals required to activate the naïve T cells. Antigen processing is where a particular macromolecular antigen is taken through the antigen-processing pathways, resulting in antigen-derived peptide(s) being presented in physical association with Major Histocompatibility Complex (MHC) molecules –

in this thesis this complex of peptide and MHC will be referred to as pMHC. This complex is the fundamental initial requirement for activation of a naïve T cell (Davis et al., 2011). Structurally this macromolecular grouping consists of a peptide cradled in the groove of the MHC; the pMHC complex is then exposed on the surface of the antigen-presenting cell (APC). The antigen-processing pathway has distinct differences depending upon the initial origin of the peptide, which is discussed later. When APCs detect foreign material, and detect that it is potentially of dangerous origin (via specialised danger signalling receptors), they take up the material and carry it to the regional lymphoid organs to present the antigen to the dormant naïve T cell populations to see if any T cells recognise it (Neefjes et al., 2011). During their migration to the lymphoid organs, the APC undergo maturation, increasing the expression of the various "tools" they will use both to present antigens and to activate any naïve T cell able to recognize the antigen/MHC complex. (Attaf et al., 2015a, 2015b; Sewell, 2012)

If a T cell possesses the specific T cell receptor (TCR) able to recognize and bind the relevant pMHC, the TCR gives an activation signal, the first of 3 powerful signals that initiate the activation and subsequent proliferation and differentiation of the T cell (Vyas et al., 2008). The specific peptide component of the whole antigenic protein is known as the epitope. The second critical signal is given via a group of so-called "co-stimulatory" molecules on the surface of the APC, interacting with dedicated counter-receptors on the T cell (Chen and Flies, 2013; McKinney et al., 2015). Signals through these co-stimulatory receptors play an important role in determining how the T cell differentiates and what the final phenotype may be in response to the local environment. The third signal is delivered via soluble mediators called cytokines produced by the APC and, depending on exactly which mediators are produced, contributes to the final pathway of differentiation of the activated T cell.

In response to this set of activation signals, the T cells proliferate, increasing rapidly in number to generate a clone of identical T cells all expressing the same TCR and hence, able to recognize the same pMHC complex. During this clonal expansion, the cells differentiate into different types of T cells (discussed in more detail in 1.1.3). Some cells are terminally differentiated and therefore undergo controlled cell death as the response wanes, in a phase termed the contractile phase (Farber et al., 2014). A sub-population of the cells become long-lasting memory cells that migrate to particular areas in the body. Upon challenge by re-exposure to their cognate antigen, these long-lasting cells can proliferate and expand rapidly and may acquire the ability to function as effector cells. This second presentation of cognate antigen can be performed by a wider range of cells including resident tissue cells capable of expressing antigen-presenting MHC molecules but not needing to give signals 2 and 3. This secondary response of memory cells to their specific antigen allows for a more rapid response able to deliver effective defences in a much more

rapid timescale (Vyas et al., 2008).

In order to discuss this in more detail, the literature review has been broken into sections. The first section focuses upon the origin of naïve T cells and the development of the specialised T cell receptor. The next section discusses the cellular and molecular biology that underpins T cell activation and differentiation from these naïve states, into activated $CD8^+$ and $CD4^+$ cells. Following sections will discuss some of the more recent advances that shed new light into these areas and how this has been applied to cancer immune-oncology. The final section introduces recent advances in genomic and bioinformatics approaches that underpin this thesis.

1.1.2 Selection of non-auto reactive T cells.

It is clearly critically important that the immune system should not be able to react against the body's normal protein components – it must be "tolerant of self". Hence, potentially self-reactive T cells must be eliminated. Precursor T cells migrate through the thymus, where they undergo various forms of selection (Klein et al., 2014). This selection process is achieved by signals given on the basis of whether the T cells express specific a particular T cell Receptor (TCR) that recognize self-antigens or not. To achieve this, the immature T cells have to generate their TCRs by splicing together segments derived from a number of genes. The combinations of different segments from many alleles is able to generate a huge diversity of TCR structures with the purpose of recognising the potentially vast number of foreign antigens with which the host may come into contact (discussed in depth later). The T cell selection process occurs in discrete thymic microenvironments – positive selection conserving pMHC responsive cells occurring in the medulla, and negative selection removing overly auto-reactive cells, occurring in the cortex (Klein et al., 2014). This means that each compartment contains unique cell types that interact with classical T cells mediating this selection process.

The aim of positive selection is to ensure the initial survival of T cells that are able to bind to MHC molecules, and which therefore, are potentially useful for the host. Initially, these cells are negative for CD4 and CD8 – hence called "double negative" cells in this setting. T cell precursors lacking CD4 and CD8 enter the thymus from the peripheral blood stream (Figure 1) and as they migrate through the cortex from the outer capsule, they develop into double positive (i.e. expressing both CD4 and CD8) cells (Fu et al., 2013; Klein et al., 2014). Also, during this stage, these double positive cells generate unique TCRs (discussed below). Double positive cells interact first via their TCRs with pMHC complexes on cortical thymic epithelial cells (Capone et al., 2001; Laufer et

al., 1996) which express both MHC-Class-1 and Class-2. The cells are selected to only express one of the markers as a result of the subsequent interaction between CD8 binding to MHC Class-1 or CD4 binding to MHC-Class 2. Consequently, this process means that only single positive cells that express the relevant CD8 or CD4 molecule are able to facilitate binding of the TCR to a particular pMHC, and therefore will be able to pass onto the next stage of selection. While the extracellular domains of CD4/8 bind to MHC, the internal cytosolic domain associates with LCK (Lymphocyte-specific Protein Tyrosine Kinase) (Van Laethem et al., 2013). Consequently, these co-receptors block non-MHC-restricted TCRs during selection, by sequestering LCK away from the complex unless in the presence of a relevant MHC molecule (Kappes, 2007; Stepanek et al., 2014; Takada et al., 2015)

If a cell fails to engage its TCR with any pMHC, it will undergo apoptosis, termed "death by neglect". This process has been demonstrated to be mechanistically reliant upon TCR affinity (Klein et al., 2014). The presentation of antigens is primarily mediated through cortical thymic epithelial cells, which are specialised APC in that they express *Psmb11*, a unique proteasome subunit. The presentation of antigens by these epithelial cells is supported by a number of more conventional dendritic cells, creating an environment that maximises potential for TCR-based stimulation. The majority of T cells that undergo cell death in the thymus can be attributed to failure to survive positive selection. Upon receiving the TCR-mediated signal following engagement with pMHC, the now single positive cells migrate into the thymic medulla (via CCR7 gradient to CCL21) to undergo elimination of auto-reactive cells by negative selection (Fu et al., 2013; Klein et al., 2014).

Negative selection is fundamental in preventing survival of auto-reactive T cells that target tissue restricted self-antigens. To facilitate this deletion of the potentially auto-reactive TCR clones, self-antigens are potentially 'seen' by single positive T cells during a 4-5-day residence in the medulla. If a cell receives repeated high affinity TCR stimulation, the cell undergoes apoptosis. In the medulla, T cells can receive pMHC stimulation from both resident dendritic cells and/or thymic epithelial cells. This elimination of T cells is highly conserved across species, with the mouse thymus auditing ~50 million double positive thymocytes daily (Klein et al., 2014). These naïve cells then migrate out of the thymus into peripheral lymphoid organs to undergo the next stage in the 'life cycle' of a T cell.

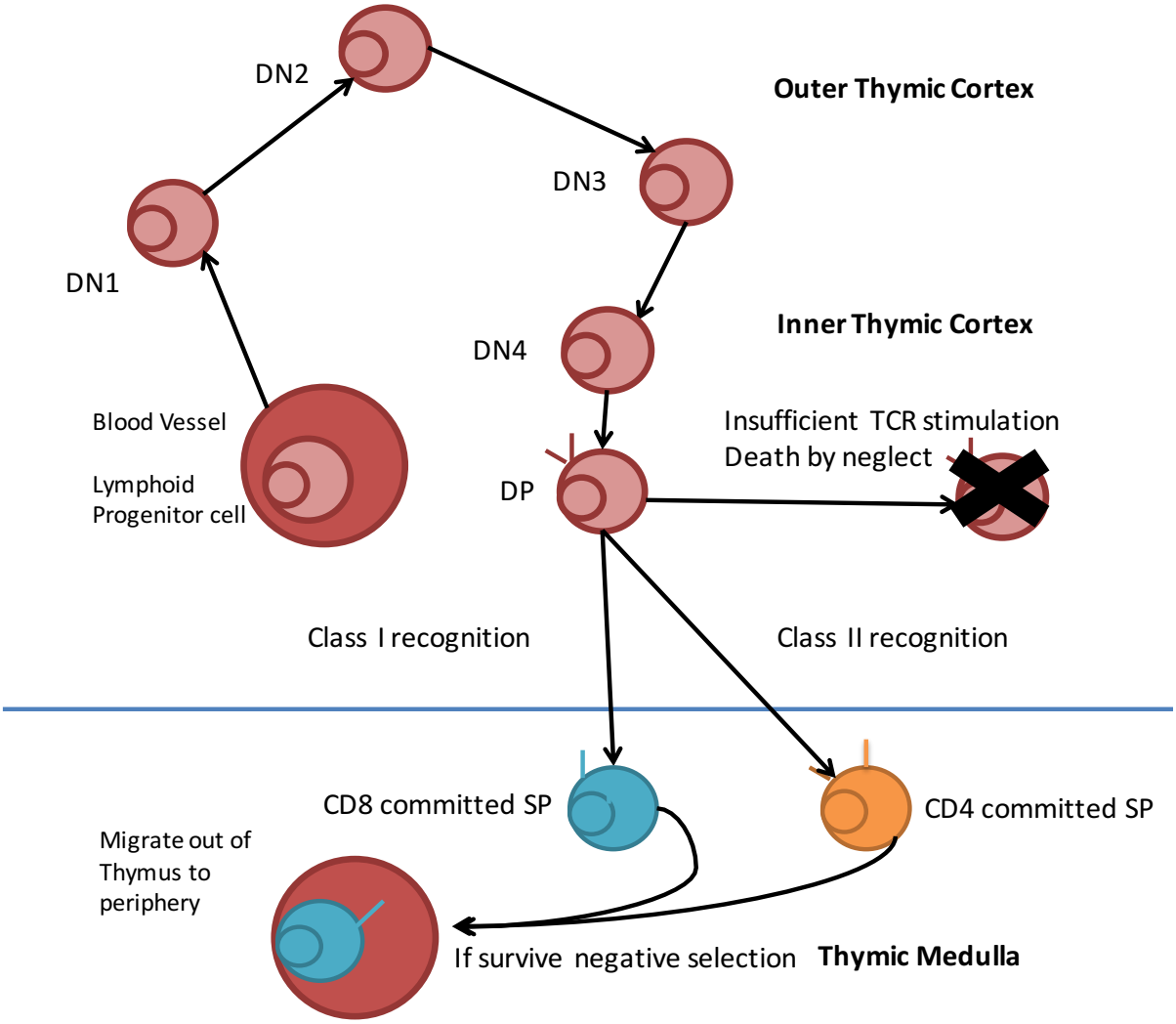


Figure 1. Early 'life cycle' of a naïve T cell.

T cells originate from hematopoietic stem cells in the bone marrow before migrating into the thymus. This pro-T cell potentially differentiates into γδ T cells or more commonly the αβ T cell. From here the αβ T cells undergo VDJ recombination to generate a particular TCR. From this point, the cell migrates into the thymic cortex, transforms from a double negative to a double positive cell expressing both CD4 and CD8, that undergoes positive selection. At this point the cells are selected towards a single positive (either $CD4^+$ $CD8^-$ or vice versa) forming the differentiation into the $CD4^+$ or $CD8^+$ lineages with distinct downstream properties discussed later. At this point the cell can undergo negative selection which can tolerise (Yu et al., 2015) or delete auto-reactive cells. DP (Double Positive), SP (Single Positive) Adapted from (Nemazee, 2006).

1.1.3 Types of T cells.

Classical T cells arise from a common precursor and share a common differentiation pathway (Figure 1) (Attaf et al., 2015b). Following selection for being non-auto reactive, T cells migrate out of the thymus and migrate to the lymphoid organs. In the lymphoid organs, these cells can undergo further differentiation in response to cytokine stimulation and transcription factor levels, leading to generation of a number of subtypes of $CD4^+$ and $CD8^+$ cells (Holgate, 2012; Iadecola and Anrather, 2011). Thus, $CD4^+$ cells can differentiate into several types of "helper" cells called Th1, Th2, Th9, Th17 and Th22 according to the cytokines they produce (Crotty, 2014; Mucida et al., 2013; Zhou et al., 2009). They can also give rise to "Regulatory T cells" (Treg). $CD8^+$ cells differentiate into cytotoxic T cells (CTLS), which produce effector molecules enabling them to kill target cells. The $CD4^+$ helper cells control the other cells, helping the cytotoxic $CD8^+$ cells to destroy any identified target cells, while $CD4^+$ Treg cells mediate turning down, or blocking excessive or unwanted activity of effector cells. The adaptive immune response peaks a few days following initial exposure to antigen and ebbs overtime, leaving a few long-lived residual cells (Seder et al., 2008). These memory cells are vital as they form an enduring reservoir from which, upon re-exposure to their cognate antigen, the body is able to respond rapidly. There are also 'non-conventional' non-$\alpha\beta$ T cells (Table 1), which are discussed later, which are more akin to innate immune cells (Attaf et al., 2015b).

Table 1. T cell subtypes.

T cell Subtype	***Target***
αβ T cell $CD8^+$	pMHC I
αβ T cell $CD4^+$	pMHC II
γδ T cells	CD1(?); cellular stress?
MAIT T cells	MR1
NKT	CD1d
Each subtype expresses a different TCR with a unique target protein. Peptide major histocompatibility complex (pMHC) remains the classical T cell system. CD1 is a family of glycoproteins structurally similar to MHC Class-1, on the surface of antigen presenting cells. MR1 (Major Histocompatibility Complex, Class I-Related) is another similar molecule (Attaf et al., 2015b).	

1.1.4 The T cell receptor and co-receptors.

The TCR is a heterodimer of two chains combined; in the majority (~95%) of T cells this is a combination of α and β, but there are γ/ δ chains on ~5% of the peripheral T cells. Each of these TCR chains comprises multiple domains. The constant region is the most proximal part of the polypeptide chain to the T cell, inserted into the T cell's outer membrane; it is joined through the J region to the variable region (Attaf et al., 2015b). In β and δ chains, there is an additional D (diversity) region between the J and V regions (Figure 2). This diversity region is the site of much genetic mutation and/or recombination to produce different antigen-recognition structures. The chains are folded and intertwined so that, at the more distal parts of the structure are 3 areas each involved in physical contact and recognition of the antigenic peptide (epitope) as presented in the groove of the associated MHC molecular complex. These areas are called complementarity determining regions and are named CDR1, CDR2 and CDR3 (Bhati et al., 2014; Hawse et al., 2014; Smith et al., 2014). CDR3 is the most closely involved in the highly specific recognition of the pMHC complex. It is therefore unsurprising that it is this area of the TCR that has the greatest potential diversity. The mechanism that mediates this diversity is discussed below.

The antigen-recognition part of the TCR – the heterodimer described above, associates with a cluster of transmembrane proteins, which together form the CD3 complex. CD3 has multiple sub chains called ε, γ, δ and ζ which combine together. These chains interact with the TCR on the T cell membrane and, after the antigen-recognition part of the structure engages with its cognate target, the CD3 complex allows further proteins to engage, which are required for signal transduction. In

order for the TCR to complete the engagement with the pMHC structure, an additional co-receptor molecule is involved in the complex: on $CD4^+$ cells it is the CD4 molecule and on $CD8^+$ cells it is the CD8 molecule. These co-receptors, CD4 and CD8, restrict the type of MHC complex with which the antigen-recognition component of the TCR can engage. Thus $CD4^+$ T cells can only interact with and recognize antigenic peptides when they are presented by MHC II structures; $CD8^+$ T cells can only interact with antigenic peptides presented by MHC-I structures (see 1.2.1) (Wang et al., 2009).

1.1.5 Generation of diverse CDR1, CDR2 and CDR3 motifs through V(D)J recombination.

The polypeptide chains that make up the TCR comprise 4 regions: Constant (C), Joining (J), Diversity (D) (on β and δ chains) and variable regions (V). Each of these regions is encoded by separate genes located on different chromosomes. There are 46 functional V genes, and 50 J genes, on the alpha T cell receptor chain, while on the beta chain, there are 48 functional variable regions, two D segments and ~13 functional J segments. The remarkable specificity of the human adaptive immune response results primarily from the TCR diversity generated through V(D)J recombination in naïve T cells (Shetty and Schatz, 2015). Only one of each chain are spliced together to form the genomic background to the final polypeptide chain structure used by an individual T cell.

Each individual lymphocyte uses a different combination of these gene segments to encode its antigen-recognition receptor. Furthermore, any one of these TCR pairs may recombine with any of the others, potentially producing a huge number of possible TCRs. During the recombination process variable (V) and junctional (J) genes from the TCR-α locus (*tra; 14q11)* are recombined with a constant segment (D). This also happens for the V and J regions of the β chain (*trb; 7q35),* but in addition, a diversity segment D and two further constant regions are used during recombination (Attaf et al., 2015b). The combination with the additional D segment allows for a higher number of potential sequences for the entire β chain. This process is initiated through the binding of RAG1 & RAG2 (recombination-activating genes) enzymes to recombination signal sequences. Hairpin loops are cleaved by endonucleases, and random nucleotides added essentially stochastically to the genomic region. These modifications of the genetic sequence greatly increase the variability in the final nucleotide sequence and hence the final structures of the TCR. Upon successful generation of a β chain, a T cell will begin recombination of the α chain (Attaf et al., 2015b). If an error occurs during this process, such as the processing ending up out of frame, rearrangement will begin on the alternative chromosome. Through the mechanism of allelic exclusion, it is possible to down-regulate the 'failed' chromosome (Abbey and O'Neill, 2008). Upon a second β chain failure, the T

cell will be potentially unable to survive selection while undergoing auditing in the thymus (Figure 1).

These newly generated TCRs are now able to bind their designated pMHC complexes through the previously mentioned contact regions, CDR1, 2 and 3. The CDR1 and 2 loops are germ-line encoded using the original genomic information, whilst the hyper variable CDR3 is generated quasi-randomly by further somatic mutation. Crystallography-based structural studies have demonstrated the CDR1 and CDR2 loops create the main contacts with the MHC helices, whilst the CDR3 loops have greater flexibility, allowing engagement of the particular peptide in the MHC groove (Figure 2) (Birnbaum et al., 2014; Rudolph et al., 2006). The importance of the CDR1 and CDR2 loops in the binding has been demonstrated with particular polymorphisms in the T cell *receptor* loci significantly altering the response against presented epitopes (Gras et al., 2010). A level of cross-reactivity is required in order to combat the wide variety of potential epitopes on pathogens, and the significance of this variety for human health is discussed later However, this cross-reactivity has a potential negative implication, being that a particular TCR may be auto-reactive and therefore could be at risk of reacting with many pMHC complexes and activating responses against 'normal' cells.

Following this overview of the biology that generates naïve $CD8^+$ or $CD4^+$ cells that can potentially engage an MHC molecule in the periphery, the next section will focus upon the mechanisms that induce a T cell response. This begins with the biology of the pMHC, which is the initiation point from which a catatonic naïve cell converts into an activated immune effector.

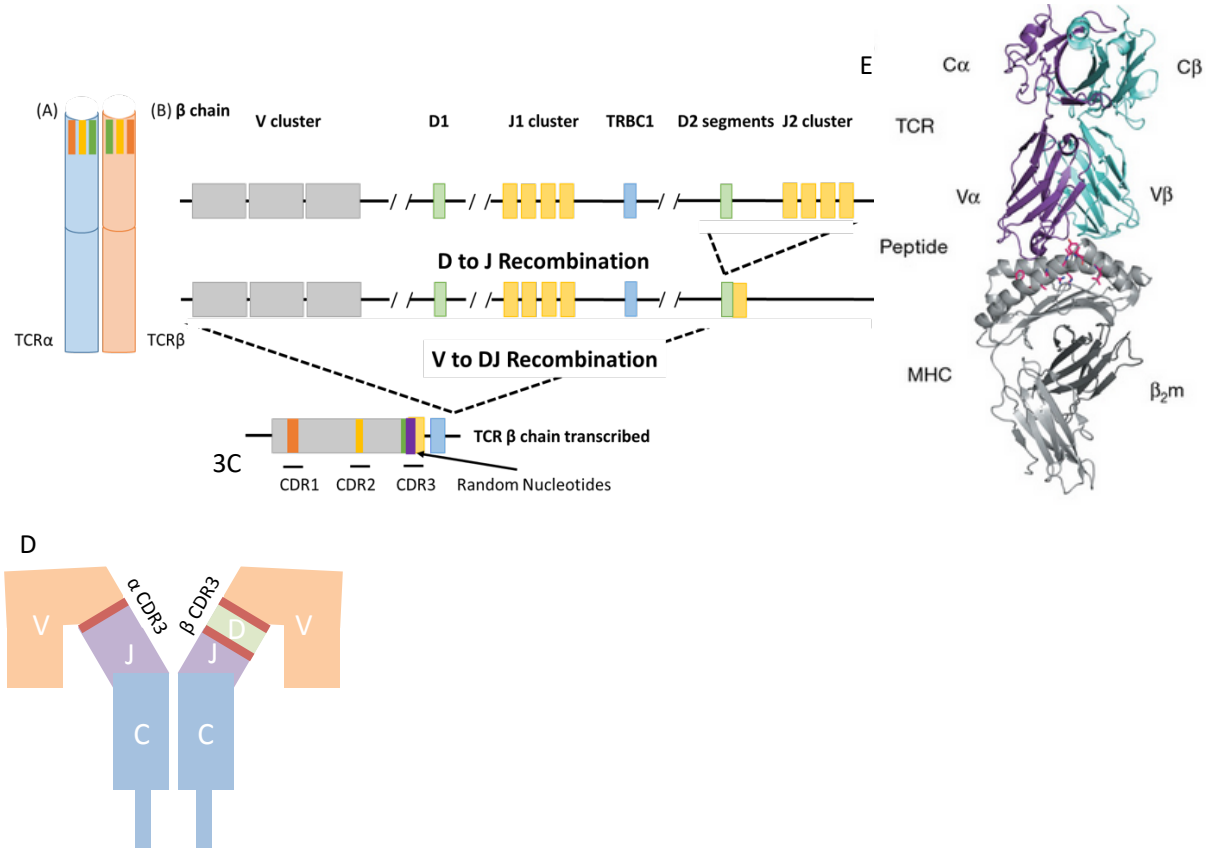


Figure 2 Graphical representation of V(D)J recombination.

The human TCR orchestrates immunity by responding to potentially billions of different ligands. The most studied isoform is the αβ TCR (A) which is pseudo-randomly generated in precursor T cells for auditing in the thymus. On the 'fingertips' of the TCR there are 6 CDR (Complementarity determining regions; 3 per chain). CDR 1 and 2 are predefined from the V gene selected during recombination, whereas CDR3 is generated by the combination of multiple segments of the genomic DNA combined with random nucleotides. The β chain initially undergoes D to J recombination combining the later section of the genomic DNA (C), and then undergoes V to DJ recombination. The α chain does not include a D segment, meaning the V and J clusters are combined in a singular step (Figure adapted from (Abbey and O'Neill, 2008; Attaf et al., 2015b)). This chain is represented as a graphic (D), showing the CDR3 on the membrane distal end of the TCR. A generated TCR is then able to bind to pMHC, an example structure from Borg, et al 2007, shows a TCR-pMHC1 complex with the MHC in grey, the peptide in magenta and the two TCR chains in purple and blue respectively (E) (Borg et al., 2007).

1.2 Mechanisms underpinning activation of a naïve T cell.

1.2.1 Loading of peptides onto MHC class I for presentation to $CD8^+$ T cells.

The MHC is a set of heterodimers with 2 heavy chains, an α and β, which are expressed on the cell surface. MHC class I is supported by a protein called β2 microgobulin to form the final protein structure on the surface of all nucleated cells. These molecules are bound together in such a way that the distal end contains a groove where a peptide can be cradled (Roche and Furuta, 2015). These molecules also have a constant region where CD4 and CD8 expressed on the T cell can bind to the MHC. The TCR on $CD8^+$ T cells engages with antigenic peptides with a two-point system: first the TCR engages with the complex of the cognate peptide lying in the groove of an MHC class I molecule, and second, the CD8 molecule itself also binds to the external structure of the MHC I. Therefore, in order to be presented to $CD8^+$ T cells, the APC must process the antigen so that the antigenic peptides are associated with MHC I molecules. Intracellular peptides (derived from tumour or viral antigens as well as the cells own structural components) are loaded onto MHC class I in all nucleated cells (Neefjes et al., 2011), resulting in an extracellular snapshot of the intracellular peptidome that a single cell contains at the time point at which a T cell interacts with the individual cell. In contrast, extracellular peptides are presented on MHC class II molecules (Kobayashi and van den Elsen, 2012).

Intracellular proteins are initially processed by multiple subunits of the immunoproteasome, which poly-ubiquitinates proteins, facilitating their proteolytic cleavage and transport of the resulting peptides into the endoplasmic reticulum (via TAP) (Elliott and Neefjes, 2006). These peptides will be loaded into the grooves of nascent MHC class I molecules, but they usually are required to be trimmed to the perfect length to fit in the MHC. To facilitate generating the correct length, peptides may be further trimmed by the enzyme ER aminopeptidase associated with antigen processing (ERAP) and transported back into the cytosol by the Endoplasmic-Reticulum-Associated Protein Degradation (ERAD) system (Nagarajan and Shastri, 2013). The peptides are recycled back into the ER until they are the 'correct' size (8 or 9 amino acids for MHC class 1). Additionally, each peptide must be able to be anchored stably in the groove with anchor residues that help maintain the peptides binding to MHC. At this point in the processing pipeline the peptides are associated with the peptide loading complex (including Tapasin, MHC class I heterodimer, Calreticulin) (Elliott and Neefjes, 2006), and can now traffic to the cell surface for recognition by the appropriate TCR. In humans, the MHC class I heavy chains are encoded by three heavily polymorphic genes allowing for an extremely high degree of variability in how a singular peptide is presented between different individuals, emphasising the specificity of singular TCR clones.

1.2.2 Loading of peptides onto MHC class II for presentation to $CD4^+$ T cells.

MHC class II complexes are primarily expressed by professional APCs such as dendritic cells. There are further levels of complexity in this pathway as proteins destined to become associated with MHC class II are sampled from the extra-cellular milieu including from surrounding cells. Consequently, the exogenous antigen must be taken in through the external membrane in order to be cleaved into the constitutive peptides before being combined with the class II MHC. This entry route involves the protein antigen being taken up into an endocytic vesicle and broken apart into constitutive smaller parts in a phagolysosome, leading to a peptide ready for combination with MHC. Although this was classically considered to be the role of dendritic cells there is growing evidence that many cells including epithelial cells, fibroblasts and other stromal cells appear to be capable of antigen presentation, under particular stressful conditions such as inflammatory conditions (Neefjes et al., 2011). The distinction between intracellular and extracellular processing is slightly blurred by the ability of specific dendritic cells to 'cross-present' antigens (Joffre et al., 2012). Cross presentation allows cells to uptake extracellular antigens and express them as intracellular ones, which can become associated with MHC Class-I, allowing the induction of $CD8^+$ T cell responses.

1.2.3 Experimental technique: tetramers and flow cytometry.

The major approach to investigate lymphocyte populations uses the principle of determining the presence of molecules of interest (such as CD8) on the surface of the cell under study, by labelling the target molecule with a fluorescently tagged antibody. The presence of the fluorescent antibody is then detected by flow cytometry. This technology passes cells in a fine stream through the beams of lasers, each of which emit light of a different but specific wavelength. Each laser causes its' specific light absorbing fluorochrome to fluoresce. The intensity of the fluorescence is quantified by detectors; based on the levels of each florescent label, the co-existence of multiple labelled molecules can be determined and the instrument (flow cytometer) allows separation of each cell population based on the intensity of a particular colour. In order to study TCR: pMHC interactions, the common tools employed are tetramers or less frequently, pentamers. These are groups of 4 or 5 covalently linked pMHC molecules bearing the relevant peptide epitope in their grooves. The tetramers of pMHC bind to the cognate TCRs on the surface of the T cells expressing them. The tetramers include a fluorescent molecule that therefore allows cells carrying the fluorescent tetramers to be identified and separated based on the levels of fluorescence intensity (Lee et al., 2008). Through multiple combinations, it is possible to investigate co-expression of

different fluorescent markers on each cell individually. Additionally, this tetramer technology has been used as a surrogate marker for measuring binding affinity using “tetramer off” rates, although these sometimes fail to reflect the true kinetics due to the multitude of factors that influences TCR:pMHC binding (Dolton et al., 2015; Stanford et al., 2012). A recent technology called CyTOF, has taken to using heavy metal conjugated antibodies, allowing for analysis of ~30 markers on the same cell without the limitation of spectral overlap but with the issue of heavily metal impurities and poorer resolution and throughput.

1.2.4 T cell responses to signal 1, 2 and 3.

1.2.4.1 Signal one.

The first signal a naïve T cell receives to initiate its awakening and differentiation/maturation is delivered via the TCR when it binds to its target pMHC complex. Unlike the membrane-bound B cell receptor (Rickert, 2013), the TCR has no intrinsic enzymatic functions such as protein kinase activity, which would enable it to give intracellular signals (Dolton et al., 2015; Morris and Allen, 2012). Instead, it depends on the recruitment of multiple molecules for signal transduction. This activation is largely mediated through the previously discussed interaction between the TCR and the pMHC, combined with CD3 chains, which are phosphorylated on tyrosine (Immunoreceptor Tyrosine-based Activation Motif, ITAMs) by intracellular LCK (lymphocyte-specific protein tyrosine kinase); in combination with the relevant extracellular CD4/CD8 molecule (Brownlie and Zamoyska, 2013; Sewell, 2012; Stanford et al., 2012). High resolution imaging has identified clustering of TCR-CD3 complexes which aggregate to form micro clusters of TCR-CD3. This immune synapse increases the likelihood of pMHC:TCR binding (van der Merwe and Dushek, 2011), where ‘synapse’ means the specialised structure formed between two plasma membranes in close contact. Initially it was defined as central supramolecular activation cluster (cSMAC) and peripheral supramolecular activation cluster (pSMAC). The spreading of the T cell in response to binding with pMHC can be seen within two minutes and the development of the different sections of the mature immune synapse are well defined within 10 minutes (Alarcón et al., 2011). The SRC family kinase members LCK and FYN (FYN Proto-Oncogene, Src Family Tyrosine Kinase) are the primary downstream molecules that mediate this signal transduction (Palacios and Weiss, 2004). To initiate intracellular signalling cascades kinase activity of SRC family kinases, particularly LCK, interact with ZAP70 (Zeta-chain-associated protein kinase 70) (Brownlie and Zamoyska, 2013; Zhang and Bevan, 2011). This signalling induces cytoskeletal rearrangements and induces the specialised effector functions of an individual T cell to occur. Owing to the low binding affinity of the TCR to pMHC complexes, it is perhaps unsurprising that additional co-regulatory molecules (Signal 2) are required for optimal

signal induction, this is also important as it allows context-specific control of T cell function (Figure 3).

1.2.4.2 Signal two.

Signal 2 is also provided by professional APCs such as DCs. This is vital as non-professional APCs cannot provide this co-stimulation and therefore cannot induce full activation, instead leading to T cell anergy. Signal 2 comprises two main parts the first of which is CD28 binding to CD80/CD86 on the DC, this induces a number of changes in the T cells. CD28 signalling leads to the activation through a number of mechanisms, such as through P13K and subsequently mTOR and NF-κB signalling. These leads to expression of IL2 (a key cytokine for T cell survival) as well as CD25, which is the high affinity part of the receptor for IL2. There are several key transcription factors in this process, including NFAT (nuclear factor of activated T cells), the nuclear localised NF-κB protein and AP-1 which co-ordinate to initiate the gene expression program associated with activated T cells. IL2 also induces T cell proliferation through cell-cycle progression, up-regulation of co-stimulatory receptors and co-inhibitory (which are discussed in their own section later) and up-regulating anti-apoptotic signals (Liao et al., 2013; Mayya and Dustin, 2016). This provides a positive feedback look further benefiting the activated T cell(s), leading to initiation of a rapidly expanding, immune response (Mayya and Dustin, 2016). Signal two also includes CD40L, binding to CD40 on the dendritic cells, which provides signalling through multiple members of the TNF Receptor Associated Factor (TRAF) family, which 'licences' DCs so that they can become more immunostimulatory (Deluca and Gommerman, 2012). This induces expression of IL12, the absence of which leads to defects in $CD8^+$ memory cells (Deluca and Gommerman, 2012). In the absence of these co-stimulatory molecules the T cells primed in this defective state produce limited IL2 and have limited proliferative capacity.

1.2.4.3 Signal three.

The third signal remains fundamental for T cell activation. It is reliant upon the local inflammatory context, which is therefore dependent upon the innate immune response mentioned earlier. Medzhitov et al identified the first human homologue of the Drosophila receptor that detects pathogens; hence termed the pattern recognition receptors (PRRs) (Medzhitov et al., 1997). These PRRs are loosely subdivided into microorganism-associated molecular patterns and damage-associated molecular patterns. The triggering of these signalling pathways induces T cell activation indirectly through induction of pro-inflammatory cytokines from the APC (Mills, 2011). For example, mature dendritic cells can detect LPS through Toll-like receptor 4 (TLR4) and induce pro-inflammatory cytokines (IL23, IL12, TNFα, etc.) that induce differentiation of naïve $CD4^+$ cells into Th17 or Th1 classes of $CD4^+$ cells. This is an example of how cell differentiation (discussed in detail

later) is skewed by supplementary parts of the immune response, to favour a particular outcome. This allows variation induced by pathogenic challenge to induce a highly regulated cascade for optimal responses against a particular pathogen type.

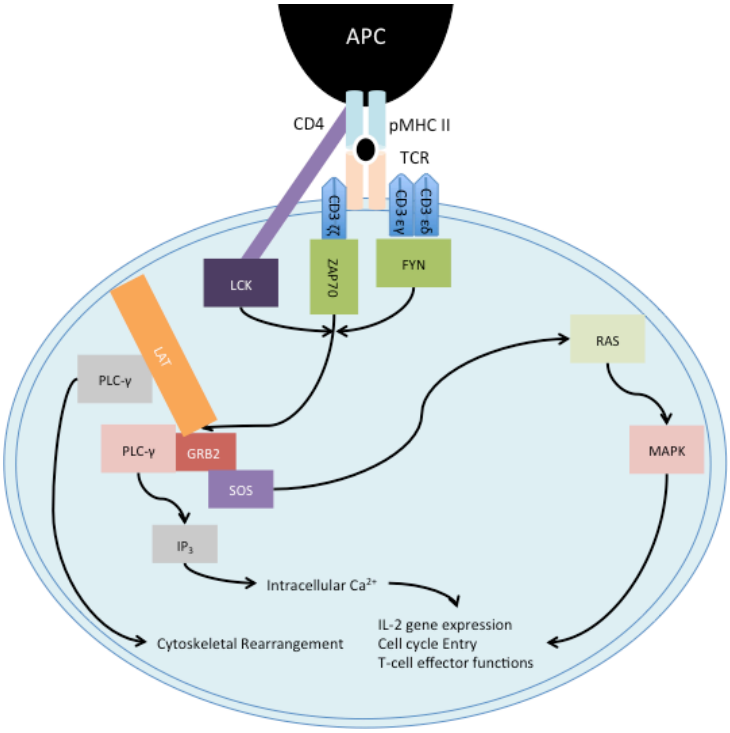


Figure 3. TCR signalling pathway.

The TCR binds to pMHC class I or II with the addition of a co-receptor (CD8 and CD4, CD4 shown here). This binding forms a complex of multiple sets of CD3 dimers (εγ, εδ, ζζ) that, combined with LCK, phosphorylates ZAP70 leading to downstream signalling cascades. LCK engagement is mediated through the co-stimulatory molecule CD4 (or CD8) on the cell surface. This ZAP70 driven cascade induces the survival, proliferation and cytoskeletal rearrangement for T cells. This cascade is mediated through the activation of the LAT-signalosome (Linker of Activated T cells) which when in combination with Phospholipase C (PLC) isoforms, adds phosphate groups on PIP_2 (not shown) forming Inositol trisphosphate (IP_3) which releases calcium from intracellular stores inducing cellular signalling. Additionally, signalling originating from SOS and GRB2 section of the LAT complex (Son of Sevenless/Growth factor receptor-bound protein 2) induces a phosphorylation cascade through the RAS driven pathways. Adapted from (Brownlie and Zamoyska, 2013).

1.2.5 T cell subtypes – CD8$^+$ cytotoxic cells and generation of CD8$^+$ effector, central and resident memory cells.

Cytotoxic T cells, one of the main subtypes of T cells, are the primary effectors in the destruction of virally-infected or malignant cells. These cells are the primary effector cells of the adaptive immune response alongside antibody-producing B-cells. CTL were originally shown in the 1970's to have the ability to lyse cells with a foreign MHC phenotype. This was quickly expanded to the recognition that CTL could also lyse virally-infected cells, introducing the concept of detection of 'altered self'. These T cells have an astonishing ability to expand (500,000 fold) in response to antigen-specific stimulus (Arsenio et al., 2014; Gerlach et al., 2010). These populations of antigen-specific CD8$^+$ effector cells are mostly a mass of short-lived effector cells, which expands and then ebbs, leaving a minimal number of memory precursor effector cells, primarily controlled through the transcription factor Tbet. Tbet has been shown to control differentiation of many cells of the immune system, both innate and adaptive. Through following a single cell which has been genetically engineered to express a marker which will be unique to its progeny, it has been demonstrated that a single CD8$^+$ cell can diversify into effector and memory cells (Arsenio et al., 2014). Other work has suggested that asymmetric division of the Tbet transcription factor between the daughter cells affected the differentiation of the progeny (Chang et al., 2007). The development of a strong CTL immune response is also dependent on multiple cytokines including IL2, IL12, IL21, IL27 for inducing the expression of the transcriptome repressor BLIMP-1, among others, which programs a cell towards CTL differentiation (Cui et al., 2011; Xiao et al., 2009).

CD8$^+$ CTLs induce cell lysis through the formation of an "immunological synapse" through which they interact with target cells. The secretory granules fuse with the presynaptic membrane and release perforin (*PRF1*) and granzymes (*GZMA*/*GZMB*/*etc.*) into the synaptic cleft, perforating the cell membrane (perforin) and inducing apoptosis (granzymes) in the target cell (Voskoboinik et al., 2015). These CTLs have been imaged grouping up and co-operating to destroy virus-infected cells. An alternative pathway is through FASL on an effector cell inducing apoptosis through FAS on a target cell. Surprisingly detailed mechanistic insight into these divergent functions are relatively poorly understood, additionally, there are a number of differences between the human and the murine system. For instance, humans have 5 granzymes, while mice have 10, the two most studied being granzyme A and B. Additionally, the most studied two molecules granzyme A and granzyme B work through different mechanisms (Kaiserman, et al. 2006), with granzyme A being caspase independent, compared to granzyme B which has several mechanisms of cell-killing including caspases. Granzyme B initiates apoptosis through cleavage of BID, caspases (3 & 10) and direct induction of Caspase-activated DNase, which initiate apoptosis inducing pathways.

Granzyme A works through alternative mechanism, inducing disruption of the electron transport complex 1 leading to DNA damage (Lieberman, 2010) and finally apoptosis.

A large number of the CTLs generated in an immune response are terminally differentiated and therefore die following antigen clearance, although a residual number of these cells survive. This memory compartment of $CD8^+$ CTL includes multiple CTL subtypes which were originally identified in the blood based on the expression of CD62L and CCR7. Based on the levels of CD62L and CCR7, a particular T cell can follow a gradient towards particular locations (Förster et al., 2008). For instance, for a T cell to migrate to the lymph nodes and blood, the cell must be high in CCR7 and CD62L expression (Förster et al., 2008). CD62L binds to adhesion molecules such as Glycam-1, allowing transient T cell adhesion. This is followed up through CCR7 binding to it ligands CCL21 and/or CCL19; this allows T cells to 'stick' and then migrate. The cells that 'patrol' the blood and lymph nodes, are called central memory cells. It has been demonstrated that a single $CD62L^{hi}$ cell transferred to an appropriate recipient animal, can provide protection against bacterial challenge, demonstrating the ability of a single cell to proliferate in the context of an immune memory response (Gattinoni et al., 2011; Stemberger et al., 2007, 2014). In contrast, T effector memory cells have a less proliferative capacity (and are CCR7/$CD62L^-$), which lies somewhere between the proliferative capacity of central memory on the one hand and terminally differentiated T cells on the other. In addition to the CD62L and CCR7 markers, the other key markers of naiveté or differentiation are isoforms of CD45, CD45RA and CD45RO. Naïve cells express CD45RA, while memory cells express CD45RO, which are different isoforms of the same gene (*PTPRC*) (Patil et al., 2018; Tian et al.). Another key molecule is KLRG1 which is used as a surrogate molecule of differentiation status as cells with a more differentiated and less 'memory' phenotype express more of this molecule (Herndler-Brandstetter et al., 2018).

More recent research has identified a third subset called 'tissue resident' memory cells (T_{RM}) which were originally missed in the research conducted on the blood. In order to investigate these cells, the original method was to serially transplant the tissue between multiple animals. This was neatly demonstrated through use of fluorescently marked immune cells that were retained in tissue when transplanted between mice. In order to investigate the levels of trafficking, the next approach used parabiosis experiments, in which two mice were attached and their circulations connected. The limited recirculation of these cells (Mackay et al., 2013, 2016; Nizard et al., 2017) was reflected in the lack after 4 weeks, of immune cells from the parabiotic partner mouse in the tissue of the indicator mouse. These tissue-resident cells are a subset that reside in the epithelial barrier tissues, positioned to rapidly respond to pathogen challenge (a hallmark of memory responses). Serial transplantation of tissue cannot be performed in a human setting. Tissue-resident cells became

apparent in humans with the analysis of the immune infiltrate of organs during transplantation (characterised primarily through higher expression of CD103 and CD69). There appears to be a further sub-population of effector memory (TEMRA) cells, which are $CD45RA^+$, $CCR7^-$ cells that have low proliferative potential, but greater cytokine effector functions and which were thought to be terminally differentiated (Patil et al., 2018).

1.2.6 Signalling controlling T cell retention in peripheral tissues.

It has been suggested that $CD8^+$ resident memory precursor cells are $CD103^+$ $KLRG1^-$ at the point that they enter the tissue following a chemokine gradient. These precursor cells develop in response to micro-environmental cues such as IL15 and TGFβ (Förster et al., 2008) from resident epithelial or resident hematopoietic cells, allowing long-term retention and survival of T_{RM} cells (Mackay et al., 2013). This residency was shown in mice to be mediated and controlled through the transcription factors *HOBIT* and *BLIMP1* (Mackay et al., 2016). CD69 is associated with peripheral retention of T_{RMs} which inversely correlates to S1PR1 and CCR7 expression, through Kruppel-like factor 2 (KLF2) signalling (Skon et al., 2013). CD69 inhibits KLF2/KLF3 (transcription factors), leading to the downregulation of S1PR1 and the unique homing profile of a resident cell. The ligand, S1P, is expressed at low levels in peripheral tissues but at high levels in the blood; this allows selective migration of T_{RM} and effector cells along chemotactic gradients. It was suggested recently, that in a small cohort of patient samples, the human lung resident cells were induced by *RUNX3* and *NOTCH* signalling and were not *BLIMP1* driven as in the mouse system (Hombrink et al., 2016). These resident T cells represented ~35% of the T cells in the Lung, with ~2% $CD103^+$ in the blood, suggesting that a small number of these cells spill into the blood (Hombrink et al., 2016). Interestingly, that paper suggested that the *BLIMP1* and *HOBIT* transcription factors, thought to drive tissue residency in mice, are not replicated in humans. Hombrink et al. compared the peripheral blood effector memory cells to the tissue $CD103^+$ cells. CD103 expression in comparison to CD69, assessed by high resolution CyTOF single cell analysis, appears to be a marker of particular tissue locations, for instance in $CD8^+$ cells isolated from Lung tissue, but not liver (Wong et al., 2016).

1.2.7 The generation of organ-specific $CD8^+$ T resident memory responses.

It is important to differentiate memory subsets which are present in the peripheral tissue at the time of sampling from T_{RM} cells, which do not recirculate, with unique TCRs and residency properties for a particular organ. CD103 has been used as a surrogate surface marker (Pan et al., 2017; Schenkel and Masopust, 2014). CD103+ cells are enriched for a TCR repertoire partially reflective of the individual microbial challenge of that particular organ. Consequently, the TCR repertoire of skin T_{RM} remains fundamentally different from those of lung or gastrointestinal

populations (Park and Kupper, 2015). The anatomical location of T cell priming is reflected in the expression of homing molecules that direct the expanded population to the particular location of priming. For example, naïve T cells activated in skin draining lymph nodes, express a set of chemokine receptors such as CCR4, CCR8 and CCR10 and CLA (cutaneous homing antigen) which binds E-selectin in the skin; T cells activated in the gut express $\alpha_4\beta_7$ which binds to MAdCAM1 (mucosal vascular address in cell adhesion molecule 1 (Park and Kupper, 2015).

1.2.8 The origin of the response - memory stem cells?

It has also been suggested that there is an additional memory state beyond that of central, effector and resident cells, termed memory stem cells (Gattinoni et al., 2011; Su et al., 2013). These are thought to form the original precursor cells during an immune response, although the current understanding of the signatures and migratory patterns for these cell stages remains limited (Chang et al., 2014). It remains unknown if memory stem cells represent distinct cell states or a pre-defined 'hard-wiring' of the cell that slowly differentiates between different stages. Furthermore, the control and differentiation of these cell stages are defined following multiple environmental cues (Figure 4) (Chang et al., 2014). TCR activation strength remains fundamental for any T cell with higher TCR affinity, skewing a T cell towards a more differentiated state. However, there are multiple other factors that have been demonstrated in the literature, for instance, a fundamental mechanism that allows variable subtypes of cells is the asymmetric distribution of the transcription factors between daughter cells during cellular division, leading to functionally non-identical daughter cells (Arsenio et al., 2015). Further environmental signals include the metabolic and inflammatory context at the point of cell division has been shown to determine cell division (Farber et al., 2014).

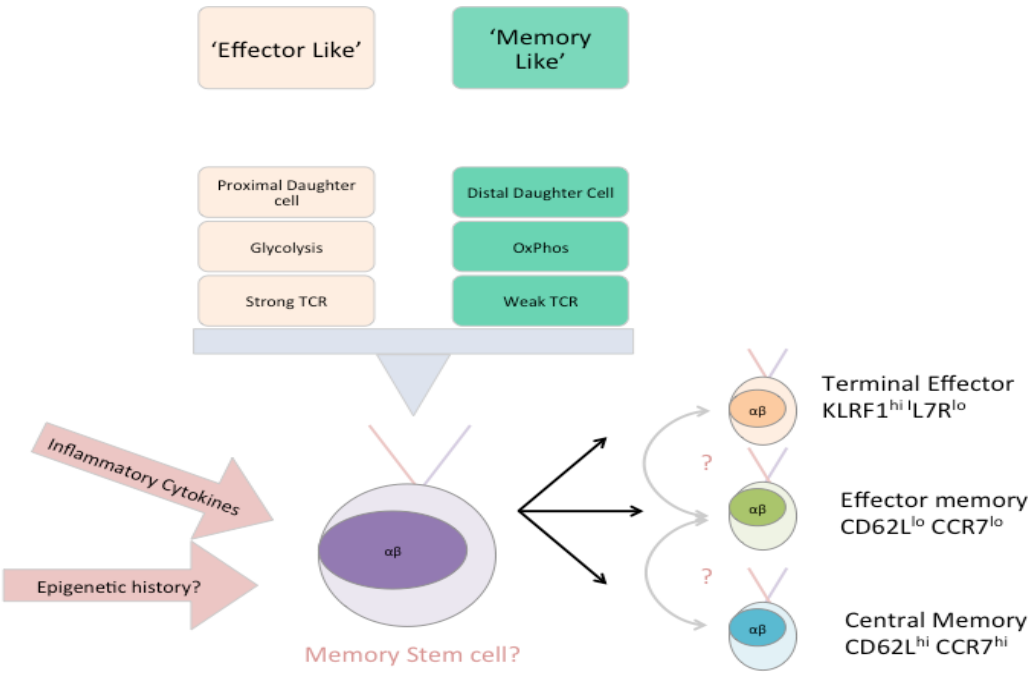


Figure 4. Development of T cell memory into the classical T_{CM} and T_{EM} subtypes.

There are multiple contexts that alter the balance, which determines if an individual cell develops into an effector-like or a memory-like cell. The metabolic state of an individual cell and its epigenetic background allow context specific outcomes. It has become clear that, for example tryptophan metabolism and glycolysis are important regulator of T cell function (Man and Kallies, 2015). Furthermore, the binding affinity of a TCR affects the outcome T-cell activation: higher affinity TCR's are skewed towards a stronger effector response. It has been demonstrated that uneven distribution of transcription factors during cellular division can allow for significant differences in cellular expression profiles (Chang et al., 2014). Effector and central memory cells are generally separated by flow cytometry on CD62L and CCR7 expression, with T cells high in these markers being found in the lymphoid tissue. These central memory cells represent a more senescent, long-lived memory reservoir (Gattinoni et al., 2011). The original 'stem cell' from which these cells originate and the mechanisms that control their development, are currently incompletely characterised. Quite where T_{RM} and exhausted T cells (discussed later) fit into this is still subject to debate if these represent distinct lineages or more of an adapted phenotype from one of these branches.

1.2.9 $CD4^+$ T cell subtypes – "Helper" $CD4^+$ cells.

In the immunological response, the CD4 helper cell represents the 'conductor' of the immune orchestra, guiding the other components of the immune response. A naïve $CD4^+$ cell expands into multiple lineages in a transcription factor-dependent process (Figure 5; based upon the immunological microenvironment (Stubbington et al., 2016). The original classification was based on two T helper subtypes (Th1/2), with each lineage responsible for the adaptive response targeted at both intracellular (typically viral) and extracellular (typically parasitic) pathogens respectively (Lloyd and Hessel, 2010). The original classification system was then supplemented by the recognition of an additional subtype, Th17 cell. It should also be noted that $CD4^+$ cells can develop into terminally differentiated cells that undergo apoptosis during the contraction stage of the response, while residual memory cells are maintained, akin to behaviour of CD8 T-cell subsets (Figure 4).

The Th1 subset is fundamental in driving responses against intracellular pathogens such as viruses. In order for a naïve $CD4^+$ T cell to differentiate into a Th1 subtype it must receive cytokine stimulation from a dendritic cell, through IL12 and IL27. These lineage-polarising cytokines induce an intracellular signalling cascade through STAT4 and STAT1 respectively, leading to an up regulation of the key transcription factor Tbet (Bluestone et al., 2009). The large amounts of IFNγ produced induce effects in the surrounding immune cells such as anti-retroviral responses or further immune cell recruitment (Bluestone et al., 2009). The Th2 subset is driven by IL4, rather than the IL12. IL4 induces expression of the transcription factor GATA3 through STAT6-mediated signalling. Th17 cells were demonstrated to follow a unique development process, distinct from that of the Th1 and Th2 subsets. Th17 cells are dependent on the transcription factor RORγT. The expression of this transcription factor is induced by IL6 and TGFβ through STAT3 (Wynn, 2005; Zou and Restifo, 2010). Beyond Th1, 2, 17 CD4 T-cells, further subtypes are being defined (1.2.11, Figure 5).

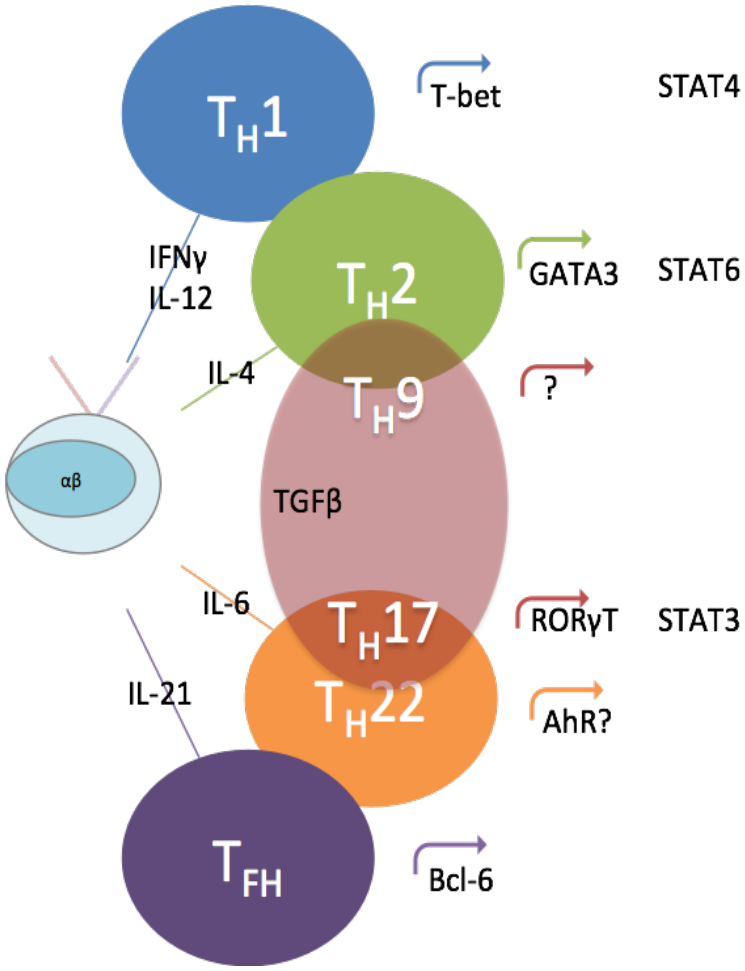


Figure 5. CD4⁺ 'helper' T cell lineages allow the optimisation of an immune response through skewing the surrounding cell down a particular lineage.

CD4⁺ T cells are divided into multiple subtypes, based on the expression of particularly transcription factors that drive a cell down a particular lineage, under the control of particular STAT (*Signal transducer and activator of transcription)* molecules. These cells then express a unique cytokine signature that induces further cells to develop particular signatures; the idea being these feedback loops helps modify an overall system towards a particular response. Adapted from (Coghill et al., 2011; DuPage and Bluestone, 2016).

1.2.10 CD4⁺ T cell subtypes – T follicular helper cells.

T follicular helper cells (T_{FH}) are a particular CD4⁺ subset involved in mediating B-cell maturation and antibody class switching. These cells migrate to the lymph node in a *BCL-6* (IL6 driven) dependent manner and express CXCR5 and CXCL13, critical for their migration into the lymphoid follicle. Upon migration into the lymph node these T cells down-regulate multiple surface proteins such as CCR7 and PSGL1, allowing chemotactic migration into the germinal centre (Crotty,

2014; Tangye et al., 2013). In the germinal centre of the lymph node these T_{FH} cells interact with dendritic and B cells, which present antigens to them. This ongoing TCR/pMHC stimulation is necessary as T_{FH} cells require constant antigenic stimulation for cell division. Additionally, these cells require a source of ICOSL, which necessitates B cell help to provide the required stimulation (Crotty, 2014). Finally, as the T_{FH} cells migrate into the germinal centre they appear to become a highly heterogeneous T_{FH} population that generates memory phenotypes, and also includes a regulatory subtype (termed T_{FR}) through a process only partially characterised, although the cells high in a marker called PD1 (discussed later) (Crotty, 2014; DuPage and Bluestone, 2016).

1.2.11 Recently identified CD4 T cell subtypes – Th9, Th22 and cytotoxic $CD4^+$ cells.

The recently discovered Th9 and Th22 memory/effector subtypes remain less characterised. It is hoped that more sensitive analytics will clarify the effector mechanisms of this cell state[47]. Th9 cells are currently defined on the basis of their expression of IL9. This IL4-STAT6 and IL2-STAT5 dependent phenotype appears to be generated through exposure to TGFβ cytokines during development (Kaplan et al., 2015). STAT6 supresses TGFβ-induced FOXP3 and Tbet, pushing the developing cell down a unique lineage pathway, which expresses a pleiotropic cytokine profile that can enhance/augment immune cell migration (DuPage and Bluestone, 2016). Th22 are a specific subset found in the skin and mucosae that express IL22, but not IFNγ or IL17, differentiating them from Th1 and Th17 cells respectively (Kim et al., 2012). These populations represent the helper side of the $CD4^+$ lineage, however $CD4^+$ cells also can differentiate into the regulatory subtype, with intensive immune suppressive capabilities. Finally, papers have also proposed the presence of $CD4^+$ CTLs which perhaps suggests a level of pluripotency from an individual cell under specific environmental circumstances although where these cells fit into the lineage profile is currently unknown (Patil et al., 2018).

1.2.12 $CD4^+$ T cell subtypes – "Regulatory" $CD4^+$ cells.

Regulatory T cells (Treg) are vital for systemic immune homeostasis; through a diverse set of functions these cells modulate the immune response. Tregs are IL2 dependent similar to other T cell lineages, but have limited expression of effector cytokines and a unique metabolite profile (DuPage and Bluestone, 2016; Li and Turka, 2010; Li and Zheng, 2015; Walker and Sansom, 2011). Expression of the transcription factor FOXP3 is the most specific feature of Treg lineages. Germ line deletion of FOXP3 leads to Treg deficiency and fatal autoimmunity. These regulatory cells have been classified into two subtypes: tTreg cells and iTreg cells. The tTreg population appears to be thymus-derived, to suppress T cells which have a relatively higher binding affinity and hence, to reduce

chances of autoimmunity. The iTreg subtype are induced regulatory cells and are derived from effector cells in the presence of TGFβ; the exact subtypes and the level of heterogeneity in these populations is currently unproven. An important protein in the Treg phenotype is CTLA4 which is a target gene of the FOXP3 transcription factor; additionally, CTLA4 engagement on Tregs enhances their immunosuppressive potential. CTLA4 represents a vital effector function of these regulatory cells. There are multiple potential mechanisms for CTLA4's role in regulating immune responses. Primarily, it competes for CD80/CD86 with higher affinity than CD28, hence, repressing CD28 induced signalling (Pardoll, 2012). Additionally, CTLA4 appears to sequester CD80/86 from the surface of the APC (Qureshi et al., 2011); though these mechanisms are not necessarily mutually exclusive. The importance of targeting CTLA4 has become significant with the advances in immunotherapy (1.4.2.1).

Having discussed the induction of a long-term memory response, the next section of the introduction will discuss some of the more recent advances in the field that shed new light on these concepts.

1.3 The evolving field of T cell biology.

1.3.1 Non-conventional T cells.

Since the discovery of the T cell receptor in seminal work by Mark Davis, the main focus has been on the α/β-expressing cells. However, there are a number of TCR subtypes that have specialised roles. It is therefore possible to use the TCR as a molecular indicator of the rarer non-αβ subtypes (Han et al., 2014). The γ/δ T cells account for 1-5% of the peripheral circulating T cell population, but in the human mucosa they represent a highly a specialised and abundant population. This concentration in mucosae may be a reflection of the receptors' specialised role in engaging CD1 restricted lipids (Attaf et al., 2015b). Studies have proposed a role of these cells in tumour surveillance, although exact mechanisms are currently not understood (Carding and Egan, 2002; Oberg et al., 2014; Wu et al., 2014). Rarer T lymphocyte types also include 'unconventional' mucosal-associated invariant T cells (MAIT) (Held et al., 2015) and invariant natural killer T cells (iNKT) (Table 1) (Attaf et al., 2015b). Additional complexity is caused by the observation of additional receptor hybrids which appear to be MHC-restricted but which recognise currently unknown antigens. This suggests the T cell compartment is far more fluid than the original peptide MHC-restricted model had predicted, and that actually, there is a spectrum of innate and adaptive lymphocytes (Fan and Rudensky, 2016). Although this thesis focuses on the α/β variety it must be emphasised that these do not represent the entire T cell pool.

Some T cells express a δ/α combination as δ and α are on the same locus (similar location in the genome). Interestingly the TCR-δ chain with multiple D segments has the largest diversity, with the ability to be translated in multiple reading frames (Attaf et al., 2015b). Despite this large diversity, limited information is available regarding this receptor subtype. Furthermore, there is limited current understanding of the selection of invariant T cells; for instance, γ/δ selection is in a TCR-dependent manner, although how the postulated antibacterial and stress-induced phosphorylation ligands for these non-classical, non-α/β (i.e./δ) are expressed in the thymus is still under investigation (Bowen et al., 2014; Godfrey et al., 2015; Holland et al., 2012). The fact that these cells are identifiable suggests that alternative methods must be employed during selection to control for the auto-reactive immune cells.

1.3.2 Recent insights: mathematical modelling of TCR diversity.

As V(D)J recombination is able to induce such a large number of variable T cells, profiling of the T cell receptor repertoires could provide insight into disease pathogenesis. Experimental evidence has demonstrated the diversity of the TCR is vital in the optimal response to immune challenge. Therefore, understanding the baseline level of TCR diversity in humans is required to understand how this varies in disease states and in the healthy state (Cha et al., 2014; Föhse et al., 2011). In order to understand TCR diversity, there are various factors that must be considered. T cells can potentially be expressing a second α/β or γ/δ chain if a cell underwent two sets of V(D)J recombination (Stubbington et al., 2016), although only one V(D)J recombination is expected to be functional. Sequencing of bulk T cell RNA (even if enriched for TCR genes) unfortunately fails to accurately combine the TCR sequences as they come from unique transcripts, so the alpha chains cannot be 'matched' to its β chain. Consequently, to understand both chains in the same cell the use of single cell methodologies remains the optimal method, although the cost and throughput are currently limited. The power of single cell sequencing now allows investigation of the variation in the paired clonotypes (Skowera et al., 2015) at higher resolution than a more traditional PCR based methods, using singular α or β chain bulk population diversity.

An alternative approach is to use statistical modelling, such as done by the Chain group (Best et al., 2015). This has identified the technical variability of bulk sequences induced through PCR, which could be overcome through single-molecule barcoding (UMI – unique molecular identifiers) and subsequent bioinformatic normalisation (Best et al., 2015). As the field expands, the potential for clonotype sequencing has begun to realise its potential in a diagnostic and predictive setting (Kirsch et al., 2015; Weng et al., 2013), although cost remains an issue. In order to understand the dynamics of TCR variability in patient samples, it is also vital to develop rapid and accurate methods to interpret the large volume of data generated.

Recently, several interesting papers have implemented tools to quantify TCR diversity (Ramesh et al., 2015; Stubbington et al., 2016; Thomas et al., 2013). Statistical methodologies previously used in ecology and microbiology have been applied to the analysis of TCR diversity; examples are the Shannon entropy, species richness and Simpson index (Pallmann et al., 2012). Each of the statistical tests outlined above model the data using different assumptions in particular on population distribution. The species richness calculation emphasises the number of rare species, while the Shannon entropy weights species according to proportional abundances and the Simpson concentration index focuses on the dominance of particular populations.

1.3.3 Aberrant human TCR diversity correlates to immunological pathology.

Li et al., 2016 quantified the TCR repertoire landscape of 9142 patient samples of tumour infiltrating lymphocytes (TILS) and found a correlation between T-cell clonotype diversity and mutational load in the tumour. The authors demonstrated a similar CDR3 length, distribution and amino acid contribution in the TILS investigated when compared to frequencies in PBMC's from healthy individuals from previous publications. 4.8% of the TILS were γδ, although there was large variation between tumour types (Li et al., 2016).

The most work on TCRs diversity has been undertaken in monozygous twins with concordant (both) or discordant (only one) disease states, as a method to investigate the role of TCR diversity, in genetically identical people. Through bulk TCR sequencing in monozygous twins and before undergoing selection, T cells have similar V gene repertoires (Zvyagin et al., 2014). However, after selection, the J usage is skewed and discordant between individuals. It has been suggested that this may be due to the variability in the processes that select and recombine the gene segements in TCR rearrangement. Interestingly the number of shared CDR3 motifs/clonotypes are similar between monozygous twins and unrelated individuals (Zvyagin et al., 2014). Yu, et al, demonstrated that even in monozygous twins individuals have private CDR3 against shared epitopes (Yu et al., 2007), suggestive of the stochastic environmental nature of this process as opposed to a hardwired genetic outcome. The repertoire is not fixed, but is plastic and malleable, shaped during the exposures of an individual during their life. This is visible in HLA-identical siblings in whome the CD8 Vβ dissimilarity increases over time (Attaf et al., 2015a). Although there is currently limited investigation of the TCR diversity in cancer, the profiling of TCR expansion would potentially indicate clonally activated (and hence enrich for tumour antigen specific) TCRs.

1.3.4 Co-stimulatory & Co-inhibitory receptors help tune a T cell to local context.

The original paradigm of a model of TCR/pMHC (signal 1) and B7-CD28/CD40-CD40L (signal 2) has become increasingly complex with the modern advances in molecular and structural biology (Chen and Flies, 2013; McKinney et al., 2015). There also appear to be additional molecules that are vital for optimising T cell function in response to environmental cues. These molecules broadly fall into two categories: the immunoglobulin superfamily and Tumour Necrosis Factor Receptor Superfamily (TNFRSF; Figure 7). Upon naïve $CD8^+$ T cells encountering antigens, the effector arms rapidly expand, as previously discussed (1.2.5). After this peak of an immune response, the surviving cells form a reservoir of memory cells. However, upon chronic stimulation, particularly with chronic

viral or cancer antigens, T cells undergo a change in their phenotype to a more "exhausted" state as they fail to clear the relevant antigen. Markers of this "exhausted" state were initially identified in the response to a virus, LCMV clone 13, which induced a longer infection than the wild type, leading to a T cell phenotype that was termed "exhausted" (McKinney et al., 2015; Wherry, 2011). An exhausted T cell is characterised by a progressive loss of effector functions, proliferative capacity and production of effector molecules IL2, TNFα and IFNγ (Wherry, 2011). This loss of effector function coincides with expression of markers including PD1, LAG3, TIM3, BTLA. PD1 and other inhibitory molecules have become highly significant in the context of cancer immunotherapy, as discussed later (Couzin-Frankel, 2013; Pardoll, 2012; Wherry, 2011). The progressive loss of effector function, altered metabolism and a unique transcriptome, develops gradually with different functions diminishing at different rates. IL2, cytotoxicity and proliferation diminish rapidly, whilst IFNγ is maintained for longer (Wherry, 2011).

"Exhausted" T cells overexpress multiple cell-surface inhibitory receptors such as PD1, targeting of which has recently emerged as a dominant treatment option in cancer (discussed later) (Wherry, 2011). The downstream TCR signalling molecules (LCK and NFATc) were also significantly-down regulated in the exhausted cells compared with the normal cells. Additionally, the genes for chemotaxis, migration, metabolism and adhesion are altered in exhausted cells. The physiological role of this exhaustion phenotype is presumed to be important for protection against autoimmunity-induced tissue destruction in the event of long-term antigenic stimulus.

PD1 acts primarily through an intracellular ITIM motif, which inhibits RAS and PI3K signalling, which reduces T cell proliferation and survival respectively. The promoter of the PD1 gene (*PDCD1*) is in an epigenetically repressed status in chronically antigen-exposed T cells, which does not always reset even when the antigen stimulation is removed (Philip et al., 2017). This was shown by epigenetic analysis using the assay for transposase-accessible chromatin using sequencing (ATAC-seq) (Corces et al., 2017; Satpathy et al., 2018; Scott-Browne et al., 2016). This suggests that in some as yet ill-defined contexts, the T cell is irretrievably reduced in its functional potential. John Wherry's group demonstrated a distinct gene profile of exhaustion in a model of lymphocyte anergy (Wherry, 2011). An important feature of these exhausted effector cells is their reduced survival and functionality, it is thought that during the initial response to LCMV-13 these cells are partially exhausted and maintain a pool of stem-cell like cells (discussed earlier) that have high expression of TCF1, enabling long term survival. Over time these cells begin to irretrievably differentiate into terminally differentiated effectors cells with poor survival and effector functions (exhausted cells; (Philip et al., 2017). Although PD1 has been best described, there are several additional markers that can be used as surrogates of this exhausted phenotype.

LAG3 is an additional marker of exhaustion. This molecule has a 20% similarity with the human CD4 molecule, and binds to MHC class II. It is not expressed by resting T cells but begins to be up-regulated several days after T cell activation. PD1 and LAG3 are also expressed on Treg cells, which opens the possibility/likelihood that targeting these proteins clinically, might affect multiple T cell populations (Nguyen and Ohashi, 2014). Targeting TIM3 has a similar impact on *in-vivo* models; when TIM3 is knocked out in murine models, autoimmunity develops (Anderson et al., 2016; Fourcade et al., 2010; Huang et al., 2015). In contrast to this inhibitory function, there is some evidence that, depending on the context, TIM3 can promote immune responses (Anderson et al., 2007). As with LAG3, TIM3 expression has been noted on both $FOXP3^{+}$ Treg populations and exhausted $CD8^{+}$ cells (in response to chronic LCMV infection). These findings suggest that targeting TIM3 or LAG3, although attractive in cancer immunotherapy, may affect both the exhausted T cells and the regulatory compartments (June et al., 2017).

The co-activators of T cells include the TNFRSF members such as 4-1BB (CD137), OX40, CD27, DR3, GITR and HVEM. These receptors synergise with TCR-CD3 signalling to promote cell cycle progression, cytokine expression and T cell survival in a variety of immunological contexts (Buchan et al., 2018; Croft et al., 2009; Kim et al., 2013a, 2015). 4-1BB remains an interesting marker due to being upregulated following TCR recognition of antigen, making it an attractive marker during flow cytometry analyses (Kniemeyer et al., 2016). TNFRSF4 (OX40) is transiently expressed following TCR ligation and, like 4-1BB, is being investigated in both cancer and auto-immunity (Croft, 2009). GITR is highly expressed on Treg and T effector cells but appears to have distinct effects upon activation, depending on the context of the cell stimulated (van Olffen et al., 2009). This demonstrates the complexity of the possibilities with multiple surface co-inhibitory and co-stimulatory receptors mediating cellular dynamics.

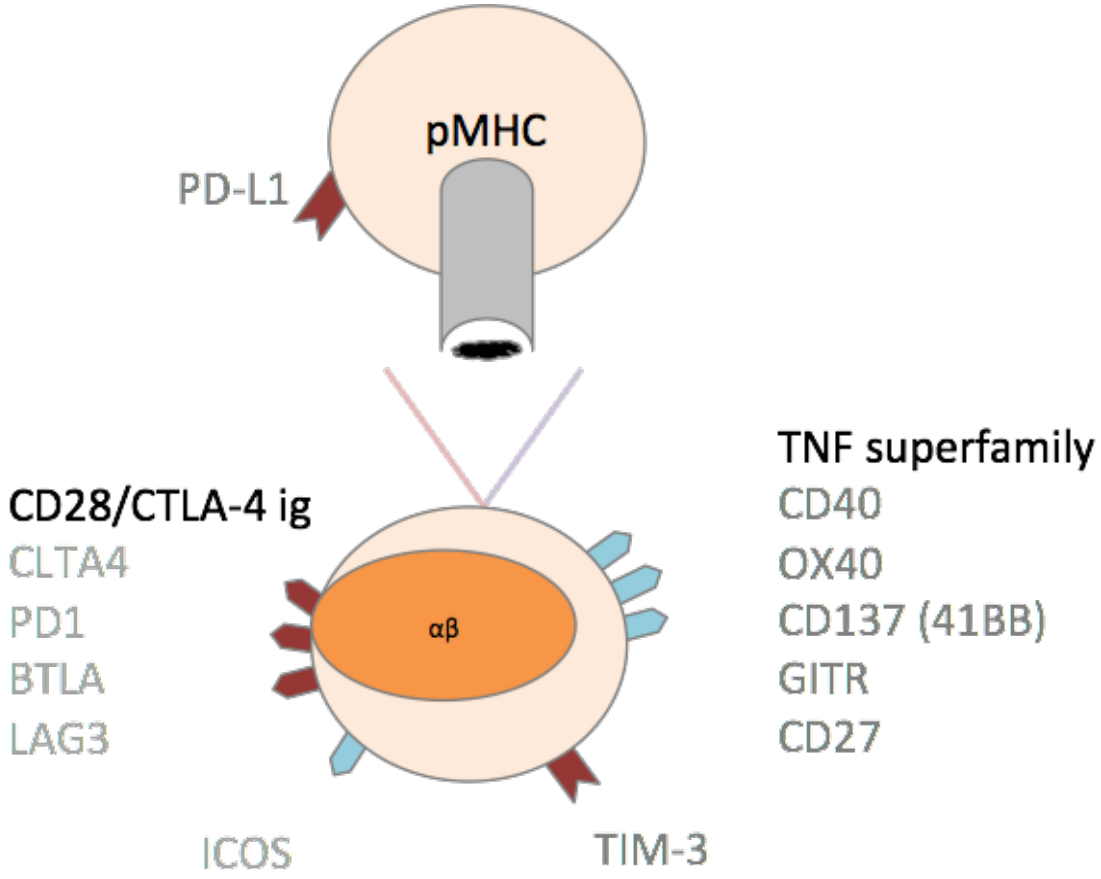


Figure 6. T cell co-stimulatory and co-inhibitory receptors.

T cells are controlled through multiple co-stimulatory and co-inhibitory receptors that help 'fine tune' the T cell to a micro environmental context. There are two main subfamilies the tumour necrosis factor receptor superfamily (TNFRSF) and the CD28/CTLA4-like families. The TNFRSF family generally includes molecules that, upon binding, induce a greater immune response. It should be noted that several of these molecules can have both stimulatory and/or inhibitory functions dependent upon the context of the cell activated. The interest in these markers has recently increased owing to the findings that the inhibition mediated by these receptors can be blocked to unleash the T cell response to mediate tumour destruction in patients for outstanding clinical benefits. Although PD-L1 is shown here on the APC, it can also be expressed on tumour cells, adapted from (Pfirschke et al., 2016; Sharma and Allison, 2015). CD28 superfamily contains CD28 and CTLA4 which compete for CD80/86 binding and several other surface proteins that are up regulated upon chronic TCR stimulation.

1.4 Exploitation of T cell biology in Oncology.

1.4.1 The critical importance of the immune response against cancer.

Cancer develops from dysplastic cells that slowly acquire mutations and develop resistance to programmed cell death. Traditionally, cancer was considered a disease of the (epi)genome, with the original 6 'hallmarks', presented by Hanaham and Weinberg (2000), focusing upon: evading apoptosis, self-sufficient growth signals, sustained angiogenesis, tissue invasion, limitless replicative potential and insensitivity to anti-growth signals (Hanahan and Coussens, 2012; Hanahan and Weinberg, 2000, 2011). Recently extra hallmarks were added: the avoidance of immune destruction and deregulation of cellular energetics (Hanahan and Weinberg, 2011). These properties are facilitated by tumour-promoting inflammation and genome instability, which drives multiple sub-clones, that diverge following Darwinian evolutionary principles from selection pressure (Alizadeh et al., 2015; de Bruin et al., 2014). That immune surveillance results in cells being constantly monitored (Galon et al., 2006; Matsushita et al., 2012; Sharma and Allison, 2015) resulted from the observation of a 3-fold increase in the rate of malignancies in immune compromised individuals. This suggests that, in all probability, many potentially malignant cells are detected and eliminated with no overt trace of cancer. Consequently, for a tumour to be clinically noticeable it must represent a cell clone that has survived immune editing and eventually escaped immune attack.

Seminal research by Robert Schreiber's group formally demonstrated this concept of immunoediting by showing that tumours induced in immunocompromised models induced a far greater immune response when the tumours were transferred to new recipients than those generated in immune-competent mice (Dunn et al., 2002; Matsushita et al., 2012). This suggested that this immune response could be quantified by the number of immune cells in the tumour. Further seminal work by Galon (Galon et al., 2006) first demonstrated that patient survival could be stratified on the basis of immune cell infiltrate levels. In keeping with this, previous research from our laboratory and others, has identified the importance of the levels of these tumours infiltrating lymphocytes (TIL's) in predicting patient outcome for a number of malignancies (Ward et al., 2014; Wood et al., 2016). There are also other cell populations infiltrating into these tumours such as innate immune cells and fibroblasts (Gajewski et al., 2013).

1.4.1.1 Stromal components inhibiting T cell responses.

Cancer associated fibroblasts (CAF's) and myeloid derived suppressor cells (MDSC) have been demonstrated to participate in inhibiting the optimal immune response (Draghiciu et al., 2015; Hanley et al., 2015). The altered collagen matrix produced by the conscripted stromal cells has been demonstrated to be prognostic through inhibiting migration of immune cells (Hanley et al., 2015; Mellone et al., 2016). It has become apparent that the immune cells in the centre of the stroma are different to the areas not within stroma (Bindea et al., 2013; Restifo, 2013). The Galon group has demonstrated in multiple high impact papers that the stromal immune response and the internal area of the tumour are distinct features, with individual impacts of patient prognosis (Bindea et al., 2013; Restifo, 2013).

1.4.1.2 Cancer cell mutational profiles predicting TIL infiltrate levels – does a lack of antigenicity prevent strong anti-tumoural response?

Generation of neo-antigens in cancers provides targets against which immune attack may be generated. Previous work had identified the importance of neo-antigens combined with TIL infiltrate levels in predicting survival (McGranahan et al., 2016; Rooney et al., 2015; Tran et al., 2015). To interrogate this further, Rooney et al., used bioinformatic interrogation to investigate mechanisms that define a tumour's visibility. They analysed the high-resolution ('multi-omic') data from the Cancer Genome Atlas (RNA-seq/proteome/genomic data from thousands of tumours). Through the generation of a gene expression 'score' (Granzyme A and Perforin) that quantified the level of cytotoxic activity in a tumour, the authors demonstrated a high degree of variability. Cervical and clear cell kidney cancers had the highest 'score' and prostate cancer the lowest. Interestingly, melanoma had an average score, however with a high degree of variability, perhaps making the heterogeneity of outcome following immunotherapy understandable. Interestingly, the highest score was associated with either a viral infection or latent retroviral expression or a high number of mutations where neo-antigens generate immunogenic epitopes. This suggests that the neo-antigen landscape links the levels of immune infiltrates, although this does not definitely imply causation. Although the number of Class I predicted epitopes was considered vital, a recent paper from Sahin's lab suggests that the number of Class II epitopes contribute to $CD4^+$ cell attack (Kreiter et al., 2015).

1.4.2 The history of cancer immunotherapy in liquid and solid tumours.

The potential of treating malignant cells through immune attack was originally identified in 1893 when William Coley used live bacteria to treat cancer. It took until 1991 for the first identification of a CTL targeting a neo-epitope to be characterised. In the next year, IL2 was approved as an anti-cancer therapy as the field began to emerge (Couzin-Frankel, 2013; Drake, 2010). Exploiting the genetic instability of tumour cells and the neo-antigens they express, is more clinically appealing and potentially more selective that through the use chemotherapy or radiation. Active immunotherapy fall into a number of categories (Table 2) and current approaches are starting to combine different therapies in the clinic. Cytokine therapy (such as IL2) appears to boost expansion of particular populations of immune cells without selecting for specificity (Liao et al., 2013; Mayya and Dustin, 2016; Stevenson et al., 2004). Antibodies can be targeted to remove malignant cells (i.e. CD20-specific mAb's rituximab (Couzin-Frankel, 2013; Zhou et al., 2008)) or to boost immune responses by blocking T cell 'checkpoint' (i.e. CTLA4 (Couzin-Frankel, 2013; Tumeh et al., 2014; Walker and Sansom, 2011; Wolchok et al., 2013)). TIL expanded in vitro can be given as an adoptive therapy (Kalos and June, 2013; Restifo et al., 2012) and vaccines aim to expand tumour antigen specific T-cells in the patient (Stevenson et al., 2004). These therapies represent a significant breakthrough that has already had clinical benefits in a range of malignancies for which treatment options were limited and clinical outlook had been poor (Couzin-Frankel, 2013; Robert et al., 2015a; Wolchok et al., 2013).

Table 2. Types of immunotherapy.

Method	Examples	Purpose
Cytokine Therapy	IL2	Non-specific boost to immune system
Targeted Therapy - targeted antibodies	Rituximab (anti-CD20)	Destruction of cells expressing particular markers
Immune activating antibodies	Ipilimumab (anti-CTLA4)	Increasing immune responses through inhibiting T cell 'checkpoints'
Vaccines	DNA vaccines, RNA vaccines	Immune responses against particular immune epitopes
Adoptive T cell transfer	TIL expansion and return material to the patient	Expansion of potentially anti-tumour immune cells and return to host

Solid tumours derived from epithelial tissue are much less distinct cellular populations, making targeting significantly more difficult without simultaneously targeting normal tissue. An alternative to monoclonal antibody-based treatments is the use of chimeric antigen receptors (CAR). These molecular constructs are a high affinity antibody domain attached to the internal signalling molecules of a TCR. Allowing for a strong T cell response to be induced against a particular antigen of interest. In theory, any cell surface molecule can be targeted, however outside of B cell malignancies these treatments have seen less investigation owing to the potential to destroy large numbers of bystander cells (Srivastava and Riddell, 2015).

An alternative approach has been pioneered by Steven Rosenberg and others, using expanded TIL's and engineering higher affinity T cells targeting somatic mutations (Nguyen et al., 2010; Tran et al., 2015; Wölfl and Greenberg, 2014). These cell therapy approaches are also demonstrating impressive clinical benefit although at risk of fatal immunological cross reactions (June et al., 2017). Rosenberg's group have published several high impact papers elegantly demonstrating the mechanism through which these expanded adaptive immune cells targets cancer antigens (Gros et al., 2016; Stevanović et al., 2017; Tran et al., 2015).

1.4.2.1 Breakthrough – 'checkpoint blockade'.

The main breakthrough of cancer immunotherapy in solid tumours, was the realization that instead of targeting the cancer cells via tumour associated antigens, greater success could be obtained by focusing upon markers of T cell regulation (initially targeting CTLA4; discussed 1.3.4) to

boost a patient's immune response. Potentially, the advantage of targeting the immune system, rather than the myriad of genetic alterations, prevents treatment selection of resistant cancer cells which has been a significant issue in targeted therapies (Vanneman and Dranoff, 2012). Through inhibiting the pathways preventing activation of dysfunctional T cells, it was possible to demonstrate improved survival benefits (Reck et al., 2016; Robert et al., 2015a). This was considered by the journal Science as the breakthrough of the year in 2013 (Couzin-Frankel, 2013).

This was the first time that a survival benefit was shown in a randomised clinical trial including patients with metastatic melanoma, with nearly a quarter surviving at least 2 years (Hodi et al., 2010). However, it immediately became apparent that responses were highly heterogeneous; interestingly, patients who survived for two years rarely relapsed, suggestive of a potential 'cure' for cancer (Rojas et al., 2015; Tumeh et al., 2014). Review of the initial cases identified that poorly immunogenic tumours did not respond to treatment, suggesting that the immune system had failed to "identify" these tumours (Stevanović et al., 2017; Tran et al., 2015). Following CTLA4 therapy, anti-PD1 therapies have revolutionized cancer treatment by inducing a higher frequency of responses with less toxicity than anti-CTLA4 antibodies. Given the association of PD1 with exhaustion and the description of CTLs expressing PD1 in human cancers, exhausted CTLs are generally assumed to be the cells reactivated by anti-PD1 therapy. However, definitive evidence for this is actually lacking in humans and PD1 is not always a marker of exhaustion but instead functions as a mechanism for a T cell to survive constant antigen stimulation (such as T_{FH} cells) (Crotty, 2014).

1.4.3 Potential reasons for failure of sustained anti-tumour immune responses under 'checkpoint blockade'.

A major problem that surrounds the use of checkpoint inhibitors, which inhibit the inhibitory molecules on immune cells, such as anti-PD1, is the heterogeneity of responses seen. It is crucially important that the reasons for this heterogeneity are established. Evidence is emerging suggesting that the heterogeneity of response is a reflection of the variable expression of the checkpoint molecules by the tumour-infiltrating immune cells. Thus, in a murine model of lung adenocarcinoma, Koyama et al, identified that progressing tumours that failed to respond to anti-PD1 therapy had a greater number of T cells (both $CD4^+$ and $CD8^+$) expressing alternative immune checkpoint markers (Koyama et al., 2016). Therefore, in theory, analysis of the expression of these other markers in patient samples could allow a personalised application of the relevant immunotherapy agent for a particular patient. Consequently, more effective selection of appropriate anti-checkpoint therapies requires an understanding of the phenotype of anti-tumoural immune cells from human malignancies both at original diagnosis and during checkpoint

blockade therapy. However, the $CD8^+$ cells represent one part of a complete ecosystem which must be considered in the analysis of human anti-tumoural responses.

1.5 Development of 'next generation sequencing' for transcriptome profiling.

1.5.1 Microarray analysis.

Up to the mid 2000's, the analysis of gene expression by assessments of mRNA levels was mainly performed through the use of microarray-based techniques. This involves the use of microarray "chips" on which thousands of DNA sequences, representing all the genes of interest (up to 55,000 genes), are applied as spots on the chip. The mRNA from an experimental cells is isolated and reverse-transcribed to cDNA. The cDNA is labelled with fluorescent tags and is applied to the microarray chip where individual sequences hybridize with the complementary gene sequence, thus labelling the spot. The intensity of fluorochrome at each spot thus reflects the quantity of mRNA derived from that gene. The different expression levels of mRNAs (in the form of the labelled cDNAs) derived from different experimental conditions could thus be interpreted through the differences in fluorescent intensity. This method dominated the field until around the mid 2000's when the first 'next generation sequencer' arrived. This technology allowed the simultaneous sequencing of millions of individual DNA fragments. By attaching an individual molecular identifier to each DNA fragment in a well, it is possible 'barcode' each sample and to subsequently use computational techniques to work out from which sample each DNA fragment came. This meant it was possible to sequence multiple samples at the same time, increasing throughput dramatically. Furthermore, this came with a 50,000-fold drop in the price of sequencing, with a rapidly increasing capacity and reduced cost. The cost of a sequencing an individual genome has dropped below $1000 at high speed and throughput (Goodwin et al., 2016). This technological advancement has heralded the ability to sequence an entire transcriptome (i.e. all the mRNA expressed in a particular sample), relatively rapidly and cheaply.

1.5.2 Experimental technique: developments with RNA-seq analysis.

The sequencing of mRNA (RNA-seq) has rapidly bcecome the method of choice for the unbiased analysis of gene expression through the development of techniques allowing the sequencing of the entire transcriptome (Braggio et al., 2013; Langfelder and Horvath, 2008; Zhao et al., 2014). In RNA-seq or microarray the analysis may focus on an individual gene and then quantifying the difference in average expression between binary conditions (Anders and Huber, 2010). Using a pre-defined data distribution (negative binominal (Anders and Huber, 2010), differences between two categorical groups are read out. As the technology has improved, more complex methods for microarray analysis (i.e. network analysis (Langfelder and Horvath, 2008; Luo et al., 2015)) have been developed to allow more stringent interpretations (Luo et al., 2015; Mitchell et al., 2015). Network approaches correlate every variable (gene) with all the others to form a correlation matrix, looking for genes correlating together, termed a 'module'. Then, by correlating any experimental variations to these modules, it is possible to understand subtle shifts that would be lost when simply comparing condition A, to condition B. As these network methods (Anders and Huber, 2010) provide a focus upon the differential correlation, this identifies significant differences that are not apparent with the differential expression based techniques (Oldham et al., 2006). Additionally, through analysing modules of co-regulated genes, it is possible to identify cell process-specific clusters of genes that interlink, co-express and potentially co-regulate each other (Ottensmeier et al., 2016). This analysis is called network analysis which is discussed later. However, despite this rigorous network analysis, often signals from individual cells or cell types are lost in bulk RNA-seq (Gaudillière et al., 2014; Newell and Davis, 2014), necessitating the development of single cell RNA-seq or 'microscaled RNA-seq' on an isolated population (Shalek et al., 2014; Treutlein et al., 2014; Trombetta et al., 2014; Tsang et al., 2015) (Figure 7 & Figure 8).

1.5.3 Examples of the Network Approaches.

A "real life" example of a network in public use would be the London underground map, which represents nodes (stations) and edges (connections). A hub gene is one with a larger amount of connectivity: in the case of the London tube analogy; King's cross or Earl's court. If this analogy is extended, it is perhaps unsurprising that these hubs reflect the status of the surrounding ones. For instance, if Embankment is busy at a particular time it is likely that the surrounding stations follow this pattern. Understanding biological data using methodologies such as this allows an investigation into the overall data while allowing for investigations into individual genes in the network.

The primary advantage of looking at a module of tightly co-expressed genes is that it improves the statistical power and reproducibility of the findings as it is not focused on individual variables. Secondly, differential gene expression fails to interrogate if there are differentially correlated genes between subtypes (rather than pure differential expression). Hovarth's (Oldham et al., 2006) seminal paper demonstrated the significance of this methodology when comparing between human and chimpanzee neurological systems; using traditional individual gene expression, they failed to find a high degree of difference that could be translated into functional differences between the species. However, when using differential connectivity, the differences became far more apparent, demonstrating the power of network theory when applied originally to microarray datasets, and now being applied to the emerging field of RNA-seq driven transcriptomics.

There have been several powerful and convincing papers using weighted gene co-expression network analysis (WGCNA) to interrogate complex datasets in the context of immunology (SLE). In a particularly strong recent example (Banchereau et al., 2016) the authors identified sub-modules that correlate to clinical parameters, the identification of these genes would not have been possible only with differential expression. Upon generation of the network this then allows the detection of hub genes that have high 'intra-module connectivity'. Modules are a more effective (and robust) way of understanding biology, as the analysis groups together genes that link to an individual pathway. The approach moves away from traditional differential gene expression of individual genes; instead, using a systems level view of the transcriptome isolated for a particular experiment. In order to generate a representative of a particular module, the concept of an "eigengene" is used. This represents a mathematical construct describing the behaviour of a particular group of genes which can then be quantified in the light of a particular biological phenotype (Langfelder and Horvath, 2007). Furthermore, by limiting the number of comparisons, the chance of a false positive (often partially corrected for in differential gene expression) is reduced (Anders and Huber, 2010).

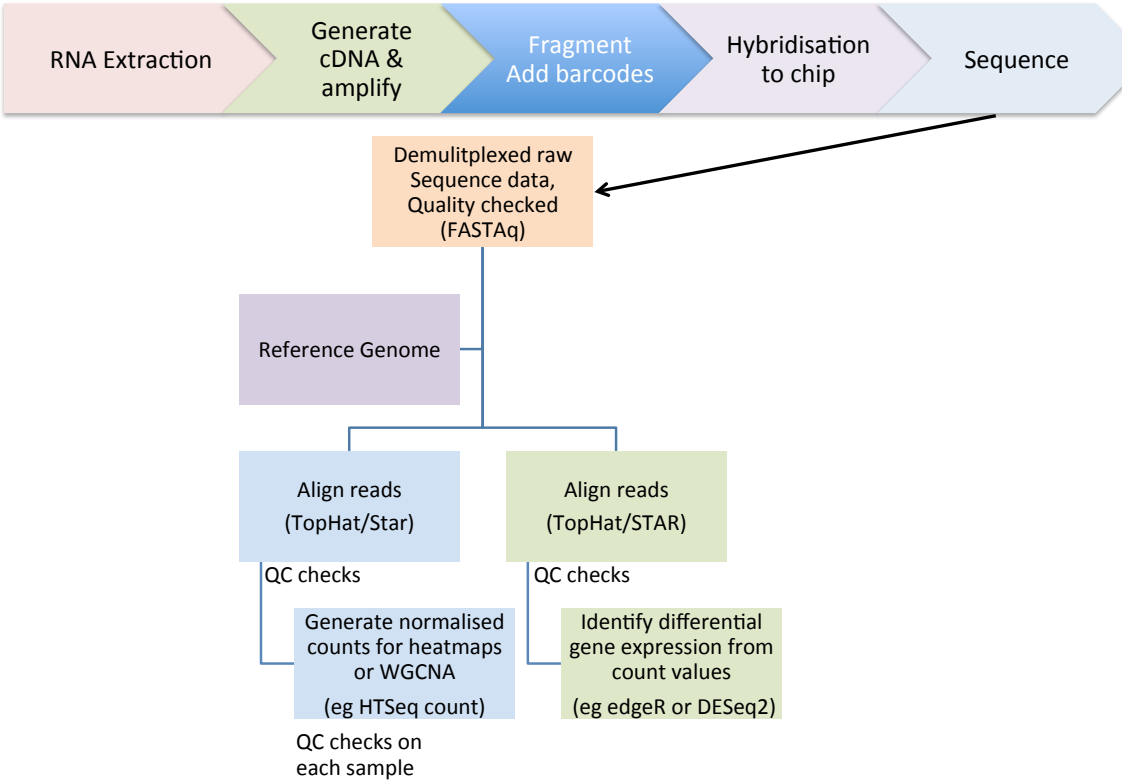


Figure 7. 'Bulk' RNA-seq pipeline.

RNA methodologies require multiple steps using RNA extraction methods to isolate the RNA from a tissue of interest, generating a library and then sequencing that on a number of available platforms. The data that comes off these platforms are de-multiplexed, aligned to the human reference genome, checked for alignment and then analysed downstream with systems modelling methods (blue box) (Van Verk et al., 2013). Through principle component and clustering methodologies, outliers can be identified and removed. To identify differentially expressed genes raw counts are normalised for size and differential gene expression is analysed following a negative binominal distribution (green box) (Gierliński et al., 2015; Robinson et al., 2010).

1.5.4 Advantages of single cell RNA-seq (scRNA-seq) compared to bulk expression profiling.

scRNA-seq has recently emerged as the highest resolution methodology for understanding complex heterogeneous tissue. This ever-improving technology makes use of advances in single cell barcoding and library preparation through microfluidic technologies (Klein et al., 2015; Macosko et al., 2015). For instance, an impressive study by Macosko (Macosko et al., 2015) barcoded, sequenced and undertook bioinformatic analysis of 44,808 retinal cells and identified 39 transcriptionally distinct subsets which included several novel types. This methodology was also applied to embryonic cells and demonstrates the ever-improving maturity of this technology (Kolodziejczyk et al., 2015; Sato et al., 2009; Yu and Lin, 2016). As 'next generation' sequencing becomes increasingly suitable and cost efficient, the biggest issue will be developing effective stringent statistical analysis methods (Brennecke et al., 2013). The next stage is the incorporation of network theory such as WGCNA for the transition from the developments in bulk RNA-seq into scRNA-seq (Luo et al., 2015). This has begun to occur in non-immunological contexts such as neurosciences (Tasic et al., 2016). This revolution has led to an increasing number of statistical techniques to model complex single cell datasets (Trapnell et al., 2014). For instance, modelling the 3D spatial location of a single cell in a developing animal (Satija et al., 2015) or the temporal space in a developing organ (Amir et al., 2013; Bendall et al., 2011; Trapnell et al., 2014).

1.5.5 Application of single cell gene expression technology in immunology.

scRNA-seq has begun to emerge for interrogation of immunologically important areas including analysis of the heterogeneity of Th17 populations. It is anticipated that given time, scRNA-seq will emerge as the highest available resolution for any immunological questions (Arsenio et al., 2015; Gaublomme et al., 2015), primarily due to the small, highly heterogeneous subpopulations controlling any immune response (Shalek et al., 2014). When analysing individual T cells with a view to defining the sequence and structure of the TCR, one initial issue was the inability to align the sequenced reads to a suitable reference transcriptome to accurately identify the TCR sequence from single cell RNA-seq data (Stubbington et al., 2016). However, recent bioinformatics techniques have now solved this issue. This means that rather than the previous single cell PCR-based methods which were limited to lower efficiency and at best, combined with semi-quantitative gene expression (Han et al., 2014), it is now possible to complete scRNA-seq and obtain the clonotype coupled to the >24,000 dimensions of potential gene expression. These methods can then be analysed through multiple pipelines that aim to reduce the dimensions to visualise the individual cells in a 2 dimensional "gene space" (Figure 8) (Amir et al., 2013; Newell and Davis, 2014).
Unsurprisingly, given the power of the technology, these approaches have been rapidly adapted by immunology researchers. Owing to the increasing power of these techniques there is growing

evidence beginning to question the classical distinctions of immune cells into broad subtypes (such as T cells being part of the adaptive response) by the finding of memory NK, NK-like B cells, iNKT cells and innate T cells; however, the implication and function of these rare cell subtypes is relatively unknown (Reeves et al., 2015; Wang et al., 2016). The finding of these cells does reflect the increasing power of more modern techniques for the investigation of these heterogeneous cells (Amir et al., 2013; Gaublomme et al., 2015). For instance, a recent single-cell transcriptomics paper has further identified a level of heterogeneity in the T_H17 lineage with variation in both pro-inflammatory and regulatory molecules, suggestive of a spectrum of biology underpinning this auto-immunity linked subtype rather than a specific binary fate decision (Gaublomme et al., 2015). A further example is a paper from the Teichman lab, describing single cell ex-vivo transcriptomics, in which they identified that naïve $CD4^+$ T cells have up regulation of *IFNGR2, STAT5* and *IL21R* compared to other ex-vivo generated $CD4^+$ subtypes; the finding is possibly suggestive of a greater potential to respond to local cytokine stimulation (Stubbington et al., 2015). This demonstrates how the individual naïve T cells are primed to be ready for a particular response, a finding which would be lost when analysing the average of the entire population.

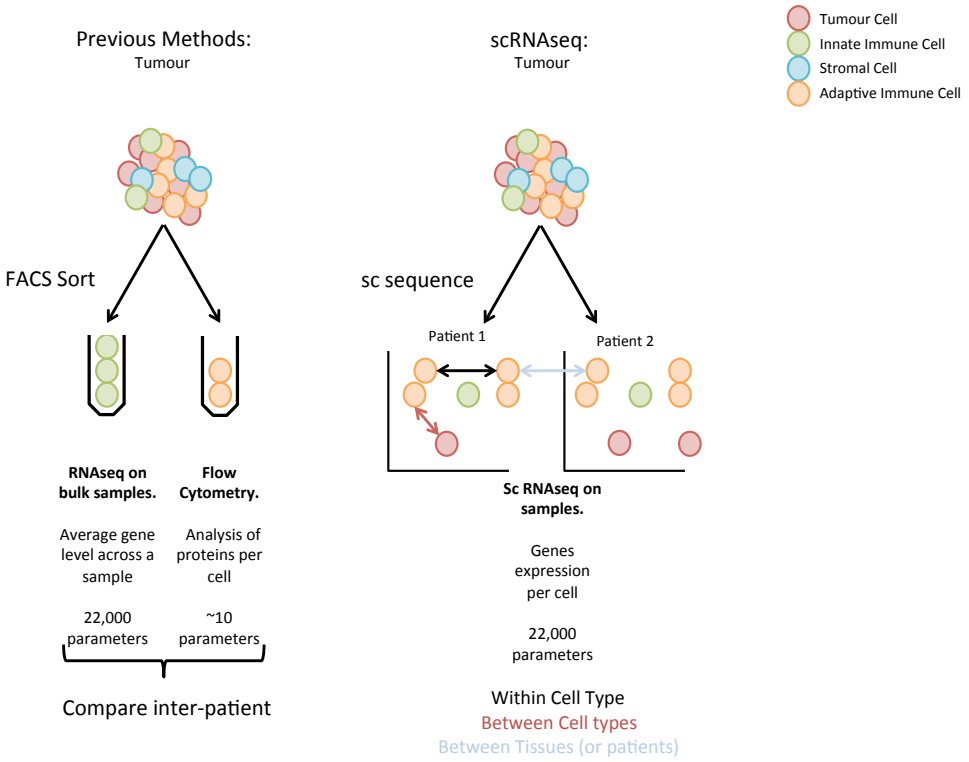


Figure 8. Comparison of traditional RNA-seq or Flow Cytometry (FACS) with single cell RNA-seq.

Previous methodologies such as microarrays or more recently RNA-seq, allowed the quantification of the entire transcriptome from bulk tissues. Our group have recently applied this technique to the isolated immune cell compartments of Lung cancer which forms the background to Chapter 4. The conventional methodology of flow cytometry using fluorescently tagged antibodies, which offers single-cell resolution but is limited by the number of colours that can be distinguished from each other in the instrument. RNA-seq allows the quantification of a many parameters in each cell population. Single cell (Sc) RNA-seq combines the best of both bulk RNA-seq (many variables) and FACS (single cell resolution). When analysed, it is possible to visualise each cell compared to all other cells in the experiment in 2D gene space.

These single cell methods allow for multiple sets of analysis such as within cell types (black arrow), which investigates the variability of the entire transcriptome. Furthermore, scRNA-seq allows comparison of individual cell types (red arrow on the principle component plot), particularly with the investigation of novel cell subtypes lost in the bulk average of tissues. Finally, by comparing between tissues and/or patients (light blue arrow on the principle components), it is possible to investigate differences in cell-type composition between matched cell type between patients, although this is currently prohibitively expensive (Sandberg, 2013).

1.6 Main Cancer types investigated in this thesis.

1.6.1 Lung cancer.

Lung cancer is divided into multiple subtypes based upon histology; there are both Small Cell (~15%) and Non-Small Cell lung cancer (NSCLC; ~85%). The ~85% NSCLC subtypes can be divided further into large cell (~15%), adenocarcinoma (~40%) and squamous cell carcinomas (~30%). More than 80% of lung cancer are caused by smoking; this can be primary or secondary exposure. Some of the other causes are Radon gas which is a naturally occurring radioactive gas present at low level in all rocks and soils, which causes a further ~3% of the lung cancers in the UK (Gridelli et al., 2015). Further causes can be air pollution and chemicals such as silica. A predisposing factor is having a weakened immune system from previous radiotherapy, HIV or immunosuppressive treatment regimens following organ transplants. Lung cancer is further grouped into stages with similar clinical outcome based on the size, nodal and metastatic invasion. Lung cancer identified early, such as at stage 1A, has a 58-73% 5-year survival; unfortunately, the later stage disease (Stage 3-4) has between a 2-10% 5-year survival (Gridelli et al., 2015). This is the second worst major tumour type behind pancreatic cancer, and combined with the high incidence rate represents a substantial disease and socioeconomic burden.

If a particular patient's cancer is smaller sized (generally lower stage), the main clinical treatment approach is surgery with curative intent. However, for larger cancers which have begun to invade surrounding tissue, or for unfit patients, such surgery is unsuitable. Frequently, as these cancers are identified in patients with a history of heavy smoking and are at a later stage, this limits the ability to survive the removal of a large proportion of the lung tissue. Patients can receive palliative chemotherapy or radiotherapy to help reduce symptoms. Of the patients who receive surgery with curative intent, a large proportion (>50%) relapse. Set against this terrible survival rate after standard therapy, immunotherapy that induces sustained responses even in the late stage disease is a major breakthrough (Couzin-Frankel, 2013; Gridelli et al., 2015; Remark et al., 2015). An important aspect for research into lung cancer is the fact that during the tumour resection, patients frequently have adjacent uninvolved "background" lung removed too.

1.6.2 Head and neck squamous cell carcinoma (HNSCC).

HNSCC comprises two major types: ~25% are driven by oncogenic human papilloma viruses (HPV) while ~75% are non-virally driven. Non-virally driven HNSCC have a similar disease aetiology to lung cancer (~85%) (Lawrence et al., 2015; Wood et al., 2016). HPV-driven HNSCC has a better survival that the non-virally driven tumours. Our team have shown that this is conferred by the greater immune response levels (quantified by number of tumour infiltrating lymphocytes) (Lawrence et al., 2015; Ward et al., 2014; Wood et al., 2016). Similar to Lung cancer, the non-HPV driven tumours are dominated by smoking induced mutations. Consequently, p53 mutations are frequently found in combination with other mutations that allow a particular cell to avoid cell cycle restriction (such as *CDKN2A*). This makes HNSCC a potentially informative cancer to investigate owing to its strong similarity to the squamous disease of the lung; furthermore, the treatment strategies are similar to lung cancer with aggressive surgery with curative intent being the main option.

1.7 Area of study.

Due to incomplete current understanding and limitations of preceding methods, there remains a pressing need to develop molecular and computational methods to analyse changes in the tumour and the T cell population at high sensitivity. Application of the new methods to investigation of the anti-tumour immune responses following immunotherapy is likely to elucidate the mechanisms of action and to improve our ability to select the most appropriate type of immunotherapy. Previous work with whole tumour RNA-seq has demonstrated systems investigation of the entire tumour microenvironment (Ottensmeier et al., 2016) which has been informative especially when combined with large amounts of clinical information. This thesis was initially focused on quantifying the level of variation, between and within patients, at the whole tumour transcriptome level. However, bulk RNA-seq is still the average of every population analysed. Consequently, this thesis then began to focus on looking at the purified population level to increase the resolution within a particular cell population. A preliminary cohort of scRNA-seq was created to investigate the level of variability at the single cell level of gene expression and TCR. Concurrently, our group had begun increasingly to focus on Tissue Resident T cells owing to findings of another ongoing project. Consequently, the next stage was to investigate these subtypes of $CD8^+$ cells at the protein level through flow cytometry (Chapter 4) and in-silico, using microscaled RNA-seq on purified $CD8^+$ subpopulations (Chapter 5) and then at the single cell level (Chapter 6). Future work will focus upon functional investigation into the particular novel targets identified from these genome-wide assays.

Chapter 2: Material and methods.

2.1 Cell biology.

2.1.1 Snap frozen tissue.

Samples were manually during surgery for the identified representative areas. Small ($0.2mm^3$) pieces were gently submerged sample first into liquid nitrogen and held for 10 seconds then dropped fully into the nitrogen. Cryosections (10µm) were cut and used for RNA isolation with the RNeasy Mini Kit (Qiagen Ltd., Manchester, UK) for whole tumour RNA-seq

2.1.2 Immunohistochemistry.

This technique allows imaging of particular proteins of interest through antibodies targeted against that protein. Frozen tumour sections taken immediately adjacent to the tissue analysed by RNA-seq were stained with haematoxylin and eosin (H&E); tumours were assessed as high (TIL^{high}), moderate (TIL^{mod}) and low (TIL^{low}) by an accredited pathologist; Professor Thomas and/or Dr Easton, as per (King et al., 2014; Ward et al., 2014; Wood et al., 2016). Staining of CD8, CD3, CD103 and CD20 were all completed using a "TRS high buffer" (Dako) in collaboration with a GCP certified clinical pathology department, as per our previous publications(Wood et al., 2016). A TRS high buffer is a 2mM EDTA pH8 buffer with the pH established with NaOH.

Cell counts were counted manually from images taken with an Olympus Ckx41 with an average of 5 high-power fields to account for any tumour heterogeneity. This method was identical to that in Wood et al, 2016 (for Chapter 3). Due to technological developments, this project moved to an automated cell counting program for the later work (Chapter 5) as discussed later. The 10x cohort (Chapter 6) was defined as per our recent paper (Ganesan et al., 2017).

2.1.3 PBMC isolation.

Blood samples were taken before anaesthesia and PBMC's isolated by layering 10ml of blood on top of 10ml of lymphoprep (Stemcell) solution; the samples underwent centrifugation at 500g for 23 minutes using the slowest acceleration and deceleration available; the mononuclear cells at the interface were harvested with a Pasteur pipette and washed in RPMI (Thermo) with 450g, 5-minute centrifugation. The cells were then re-suspended in freezing medium into cryovials at a concentration of ~8×10^6 per vial (McCann et al., 2016). cRPMI was then mixed with 40% inactivated AB Human Serum (Sigma) and 10% DMSO (Sigma) final percentage.

For making the cPRMI the 500mL final volume RPMI contained sodium pyruvate (1mM final) (Sigma), Pen-Strep (100units/ml) (Sigma), L-glutamine 2mM (Sigma) together with 50ml Fetal Bovine Serum (Sigma).

2.1.4 Tumour dispersion.

Tumour-infiltrating T cells and background non-involved Lung T cells were isolated using a combination of mechanical and enzymatic dissociation. The tumour tissue was cut into small fragments using a scalpel. Tumour fragments were then incubated at 37^{o}C for 15 minutes in an orbital shaker with 1mL RPMI 1640 medium (Gibco, Fisher Scientific UK Ltd., Loughborough, UK) containing 20 units/mL Liberase DL (Roche Diagnostics Ltd., Burgess Hill, UK) and 800 units/mL DNase I (Sigma-Aldrich Co. Ltd., Gillingham, UK). The tumour cell lysate was then passed through a 70µm filter with ice-cold RPMI 1640 medium and centrifuged at 1500rpm for 5 minutes. Cells were re-suspended in Red Blood Cell Lysis (Qiagen) for 10 minutes on ice, as per manufacturer's instructions. Cells were re-suspended in PBS (1xPBS containing 2mM EDTA (pH 8.0)) and the volume adjusted so that in combination with antibodies and FC-block the final volume would be exactly 200µL.

2.1.5 Flow cytometry: phenotypic analysis.

Using a flow cytometer, it is possible to analyse the expression of multiple proteins on each cell. Compensation was completed using "ultra comp" compensation beads (ebioscience) with 1µL of antibody. For antibody staining, cells were incubated with 10µL FcR block (Miltenyi Biotec Ltd., Bisley, UK;) (10 minutes @4 °C - dark)) per 100µL of cell suspension. Then, the relevant antibody cocktail (30 minutes @4 °C - dark)) was added to a final volume of 200 µL with relevant FMO (fluorescent minus one) controls. The tube was filled up with MACS buffer and centrifuged for 10 minutes @450g. The sample was then re-suspended in 300 µL MACS buffer (1xPBS containing 2mM EDTA (pH 8.0) and 0.5% BSA) and maintained @4 °C (dark) for immediate acquisition on a FACS Canto II (Becton Dickinson) system. To account for the level of cell death in the tumours, the live/dead stain was added at 2x the recommended amount before Fc blocking for Aqua based live/dead discrimination. For surface only analysis not using an Aqua based live/dead discrimination, a 10,000-event acquisition (as a negative control) was completed. Following this, the DAPI (2µl of a 200µg/mL stock solution, per million cells) or PI (at 4 times manufacturers instructions) was added and the sample analysed on the flow cytometry instrument

For intracellular staining, the dispersed surface-stained cell suspension was spun for 5 minutes (450g at 4 °C), the pellet was gently re-suspended and then add 0.5ml Fixation Buffer was added drop wise on a vortex mixer (BioLegend) following an Incubation in the dark for 20 minutes at room temperature. The cells were then spun down at 450 g for 10 minutes at RT. *(450g and 10 minutes was important for these low cell number samples).* For permeabilisation, the 10X Intracellular Staining Perm Wash Buffer (Biolegend) was diluted to 1X in sterile distilled water. The fixed cells were re-suspended in 4 ml of Intracellular Staining Perm Wash Buffer and centrifuged at 450 *g* for 10 minutes.

For intracellular staining, fixed/permeabilised cells were suspended in 130 µL Intracellular Staining Perm Wash Buffer with a small aliquot taken for a fluorescence minus one (FMO) control. 20 µL of Fc block was added Intracellular antibodies were added for 30 minutes in the dark for a final volume of 200 µL. The cells were then washed with 4 ml of Intracellular Staining Perm Wash Buffer with the centrifuge set at 450 *g* for 10 minutes. Cells were re-suspended in 0.2 ml MACS Buffer in FACS tubes and analysed with appropriate controls.

For stimulation, the sample was divided into two falcon tubes, (± stimulation) and re-suspended at $4.0x10^6$ cells/ml cRPMI in a 15 ml Falcon pre-warmed to 37°C with or without PMA (Sigma) and Ionomycin (Sigma) at the appropriate concentration (Table 4). The tubes were then Incubated at 37°C (with CO_2) at a 45-degree slant and with the lid slightly loosened for two hours, the Brefeldin A (5µg/µL) was then added for a further two hours. Importantly, every 30 minutes, the tubes were vortexed gently to limit cells from clumping. Data was exported from FACS Diva and analysed in FlowJo (10.2).

2.1.6 Antibodies.

Samples were stained with the following antibodies at the volume (Table 3) indicated for a final volume of 200 µL. Example plots are included in each figure. Also note CD107A staining must occur during stimulation. These antibodies were added in excess owing to the large variability in cell numbers and cell viability per sample.

Table 3. Antibody list.

Epitope	Fluorochrome	Company
GRANZYME A	Alexa Fluor647	BIOLEGEND
CD197 (CCR7)	Alexa Fluor700	BIOLEGEND
CD4	Alexa Fluor700	BIOLEGEND
CD45	Alexa Fluor700	BIOLEGEND
CD103	APC	BIOLEGEND
KLRG1	APC	BIOLEGEND
CD3	APC	BIOLEGEND
CD3	APC-Cy7	BIOLEGEND
IFNG	BV421	BIOLEGEND
PERFORIN	BV421	BIOLEGEND
PD1	BV421	BD
CD4	BV510	BIOLEGEND
CD62L	BV510	BIOLEGEND
CD4	BV510	BIOLEGEND
TIM3	BV605	BIOLEGEND
CD45	FITC	BIOLEGEND
CD39	BB515	BD
CD4	PE	BIOLGEND
GRANZYME B	PE	MILTENY BIOTECH
CD107a (LAMP-1)	PE	BIOLEGEND
CD49A	PE	BIOLEGEND
CD49A	PE	BD
CD39	PE	BIOLEGEND
GRANZYME B	PE	eBiosciences
4-1BB	PE	BIOLEGEND
CD103	PE-Cy7	BIOLEGEND
CD103	PE-Cy7	BIOLEGEND
CD3	PE-Cy7	BIOLEGEND
CD3	PE-cy7	BIOLEGEND
CD103	PE-cy7	BIOLEGEND
CD38	PE-cy7	BIOLEGEND
CD8A	PerCP-cy5.5	BIOLEGEND

Table 4. Stimulation reagents.

Stimulation Reagent	Final Concentration
PMA	20µM
Ionomycin	1µM
Brefeldin A	5µg/µL

2.1.7 Flow cytometry sorting.

Using the principles of flow cytometry, it is possible to separate out cellular populations into individual tubes based on expression of particular markers. Compensation and sample processing was completed as per phenotypic analysis above. The sample was then re-suspended in 400 µl and maintained @4 °C (dark) for immediate acquisition on a FACS Aria. The Aria sorter was optimised with Accudrop beads (BD) and the samples sorted into 4 °C 750 µL trizol tubes or 4 µL lysis buffer for single cells or for ultra-low cell numbers (cells following simulation) (Engel et al., 2016; Wood et al., 2016).

2.1.8 Flow cytometry analysis.

Analysis was completed in Flowjo, data was exported and analysed in Graphpad Prism 6 or 7.

2.1.10 Computational image analysis of tissue microarrays.

This technique allows a computer to count the number of positive cells for a particular protein marker. Tissue Arrays were scanned in using a Zeiss AxioScan Z1 (As per manufacturers' instructions) and each core was imaged. Data is then transferred to a high-performance computer where the following script was written and run using the software Qupath 0.1.2 downloaded from https://github.com/qupath/qupath and analysed with a custom script. This generated a .txt file that has the name of the image alongside the number of positive cells, number of total cells and the total area. As each sample was an average of 5 cores, the 5 were averaged (mean) which represents the final IHC value for a particular patient.

2.2 Molecular biology.

2.2.1 RNA extraction.

RNA was extracted from Trizol using miRNAeasy (Qiagen) as per manufacturer's instructions with an elution volume of 16μL Tris-EDTA (10 mM Tris-HCl, 1 mM EDTA, pH 8.0 – referred to as TE). This was to allow for 'bulk RNA-seq' as used for the $CD103^+$ or $CD103^-$ $CD8^+$ subpopulations in Chapter 5. For the single cells and lysis buffer sorts, samples underwent reverse transcription and were quantified after amplification to check the quality of the cDNA generated with greater security.

2.2.2 qPCR to quantify mRNA amounts from extracted samples for bulk RNA-seq.

This quantifies the amount of RNA in a particular sample used in Chapter 5. RNA was extracted and 1.5μL of the total RNA volume extracted was added to 2.8μL molecular grade water containing 2.5μM OligoDT, 5 ng/μL Random Hexamers (all Invitrogen) and 1.2 Units/μL RNAse out (Clontech). The plate was then compared to a standard curve of known amounts of human T cell RNA. The plate was incubated at 65°C for 5 minutes and then placed on ice for one minute. 5.25μL of cDNA synthesis 10x first strand buffer was added to each sample containing 5mM $MgCL_2$, 10μM DTT, 0.5mM dNTPs, 1.2 units/μL RNAseOUt and 10u/μL Superscript III reverse transcriptase (Invitrogen). The plate then underwent qPCR PCR (primers below) using 25°C for 10 minutes, 50°C for 50 minutes and then 85°C for 5 minutes. A 5μL SYBR Green Master Mix containing final 0.3 μM forward and reverse primers was added alongside 16.3μL water, per sample. The level of RNA was quantified versus a standard curve using a 95°C (10 minutes), 40 cycles of 95°C, (15 seconds) 65°C for 1 minute set of cycles using a QuantStudio6 (Thermo).

B2M FWD: 5’- CTGCCGTGTGAACCATGTGACTTT-3’;

B2M REV: 5’-TGCGGCATCTTCAAACCTCCATGA-3’

2.2.3 RNA-seq – Smartseq2 for micro-scaled assays (1-200 cells).

2.2.3.1 Preparation of the single cell plate for reverse transcription.

This enriches for the polyadenylated RNA leaving an overhang for use later in the protocol. Importantly this entire procedure until post amplification was completed continuously. Upon use of a plate for scRNA-seq the work bench, pipettes and any additional kit was cleaned with RNAzap and 70% Ethanol. Furthermore, wearing a mask during this procedure is required and all tips were molecular grade. The plate was transferred out of the freezer, the ice wiped away, the plate was then transferred onto ice and then placed directly into a 4°C centrifuge. After centrifugation at 1000g rpm for 5 minutes, the plate was placed onto a cleaned metallic rack. 1µl of Oligo30dt primers (10µM) (Picelli et al., 2014) was added directly to the liquid, the plate was spun briefly in a centrifuge, vortexed and then spun again. It was imperative that at all stages the plastic lid was resealed tightly and the vortexing was sufficient to mix the pellet but not to allow any liquid to spill onto the lid. The plate was then incubated at 72°C for 3 minutes and then placed on ice for 3 minutes. The sample was then pulsed briefly in a centrifuge and then placed briefly on ice (5 minutes). At this point the oligo-dt primer was hybridised to the poly(A) tail of all the mRNA molecules.

2.2.3.2 Reverse transcription for Single cell RNA-seq.

The RNA is converted to cDNA using a template switching Oligonucleotides which generate full length coverage cDNA. At the 5’ end of the RNA template the enzyme leaves a few additional nucleotides which exploited as the template switching site. Upon base pairing between the Template switching oligo and the appended stretch, the reverse transcriptase switches from the cellular RNA to the alternative strand. With the overhang generated in the previous poly-adenylation step this allows the addition of a primer sequence of choice for use in the pre-amplification step (Picelli et al., 2014). 5.7µl of the RT mix (Table 5) was added, to the RNA. The samples were vortexed, spun briefly and then incubated in a thermal cycler with a heated lid (105°C) for 42°C (90 minutes); 10 cycles of 50°C (2 minutes), 42°C (2 minutes); then 70°C for 15 minutes with a 4°C hold.

Table 5. Reverse transcription reaction for single cell RNA-seq.

RT Mix	Final Concentration
Superscript II RT	100U
RNAse inhibitor	10U
Superscript II first strand buffer	1x
DTT	5mM
Betaine	1M
$MgCl_2$	0.6mM
TSO	1µM
PCR quality water	To volume
RT mix was prepared in this order; RNAse inhibitor, buffer, DTT, Betaine, Magnesium, then the TSO (Template-Switching Oligonucleotide) and nuclease free water, spun, then the superscript II enzyme added.	

2.2.3.3 PCR pre-amplification and purification.

The low amounts of cDNA are amplified further using primers specific for the Oligo(dT) primers added during the polyadenylated RNA selection. This cDNA mixture contains large amounts of primer dimers which are removed through clean-up. The PCR mix (KAPA hiFI Hotstart Ready-mix; KAPA biosystems) was generated and 15µL of the mix (Table 6)was added to each well, the plate was spun briefly, vortexed and pulsed again before a PCR reaction with a 98°C for 4 minutes then 18-23 cycles of 98°C for 20 seconds, 67°C for 15 seconds and 72°C for 6 minutes. The reaction was then held at 72°C for 15 minutes before being placed at 4°C. It must be noted that as the Ready-mix buffer has a higher salt concentration, this affected the DNA melting and consequently the denaturation was completed at 98°C. For bulk RNA-seq 18-19 cycles was used depending on starting amount of RNA as calculated from the B2M PCR, discussed previously, for single cells 23 cycles were previously established as most suitable for T cells (Patil et al., 2018).

Before the purification steps, the Ampure XP beads (Beckman Coulter) were incubated for 15 minutes at room temperature and vortexed before use. 80% (vol/vol) Ethanol solution (made with RNAse-DNase free water) was briefly vortexed before use. 80% of starting volume of the beads was added to each sample, a new plate lid was attached, the plate spun, vortexed, spun again followed by an incubation at room temperature for 4 minutes. The plate was then attached onto the magnetic stand and left for 5 minutes for the beads to equilibrate. Following this, all the liquid removed for all the wells. 190µL of 80% ethanol solution was added, incubated for 30 seconds and then removed; again, with a second pipette to confirm the removal. This wash was repeated three times with an additional QC step using 20µL pipette tips. The plate was then left for 5 minutes, 25µL of pre-warmed (60°C) TE buffer was added and the plate spun, vortexed and spun again making sure the colour was homogeneous with no liquid on the lid. After a 2-minute incubation off the magnet, this entire purification was then repeated eluting in 20µL of warm TE. The 20µL was then transferred into a low retention plate for storage at -20°C until the cDNA quality controls steps.

Table 6. Pre-amplification mix.

Pre-amp mix	**Final concentration**
Kapa HiFI Hostart Ready Mix	1X
PCR Primers(Engel et al., 2016) (IDT technology)	0.1 µM
Molecular Grade Water	To volume
Total	15µL

2.2.3.4 Quality control (pico green and bioanalyser).

Both of these methods are completed after the pre-amplification and library preparation steps with different objectives. The picogreen assay is a quantitative assay using a dye for double stranded DNA to accurately quantify the amount of DNA in a well. The bioanalyser quantifies the quality of the cDNA generated and length of cDNA fragments.

For the picogreen assay, a TE solution was created with 50ml water, 500µL Tris pH 8 and 10µL EDTA and 18µL pipetted into each well of a plate. For each run a new solution of N x 200µL of TE, with 2µl of picogreen dye per ml was generated (as per manufacture instructions; Invitrogen) with 190µL of this pico-TE solution added to a black, flat bottom, 96 well plates Using a sonicated standard lambda DNA (Sigma) dilution at 40ng/µL (Thermo) with serial 1:4 dilutions' standard curve using a blank negative control well. 1/10 diluted (TE) of samples were added and quantified (as per manufacturer's direction)

The bioanalyser cDNA quality analysis was completed as per manufacturer's instructions (Agilent), using the quantity of cDNA and the average size of the fragments (and quality) from the Bioanalyser data it is then possible to accurately dilute the sample to the required amount for library preparation.

2.2.3.5 NexteraXT- library preparation (single and bulk RNA-seq).

Library Preparation adds a unique molecular barcode to an individual set of cDNA which then means it can be pooled with many other samples and sequenced at the same time with a unique marker for each well. In a fresh, RNA/DNA-free plate, 1.5ng of template pre-amplified cDNA was diluted in 5µL high quality TE (i.e. freshly made and warmed to 60°C). The 'Tagmentation' library preparation reaction was made up for 100 single cell wells (Table 7) or the required number of bulk samples (Table 8) on ice. If a single cell failed to reach the given amount of cDNA then the maximum amount available was taken, for bulk RNA-seq this sample would not be continued.

Table 7. Tagementation reaction for single cell RNA-seq

	Volume µL	
Tagmentation buffer 2x	1.6	Make together for 100 wells and pipette into plate on ice
Amplicon tagmentation mix (Illumina)	0.8	
Total	2.4	
cDNA template	0.8	

Table 8. Tagmentation reaction for bulk samples.

	Volume µL	
Tagmentation buffer 2x	10	Make together for X wells and pipette into plate on ice
Amplicon tagmentation mix	5	
Total	15	
cDNA template	5	

The plate was spun, vortexed and spun again before the tagmentation reaction was completed in a thermal cycler at 55°C for 5 minutes before cooling on ice. 0.8 µL of NT (Neutralize Tagment Buffer; Illumina) buffer was added and the plate was vortexed and pulsed again before a 5-minute room temperature incubation. 2.4 µL of the Nextera PCR Master Mix (Illumina) and 0.8µL of each of the two barcode indices (1.6µL total) was added to each well and the plate was pulsed, vortexed pulsed before 9 -12 cycles of a PCR amplification reaction. For bulk samples, 5 µL NT buffer was used, and 25 µL of PCR mix.

The PCR reaction was 72°C for 3 minutes; 95°C for 30 seconds; then 9 or 12 cycles (bulk or single cell) of 95°C (10 seconds), 55°C (30 seconds), 72 °C (30 seconds); 72°C for 5 minutes and a 4°C hold. The plate was then diluted with molecular grade water to a final volume of 40 µL (32 µL added); 32 µL of the purification beads were added and the plate purified once as previously outlined, before the library prepared cDNA was suspended in 24 µL of suspension buffer. Each well underwent picogreeen and each plate 11 well bioanalyser QC steps, as previously described before pooling.

2.2.3.6 Pooling, final QC and sequencing on Illumina Hiseq2500.

This combines multiple samples together at equal amounts for sequencing. Samples were pooled as per our manuscripts from our group (Engel et al., 2016); all four NexteraXT libraries were combined aiming to have a final concentration of 2-10molar in molecular grade water which was then hybridised to the sequencing chip. This was to allow accurate pooling for the cBot on a Quanti Studio Flex 6 (Applied Biosystems; as per manufacturer instructions). Samples were sequenced in-house on a HISeq 2500 (Illumina) using standard 50bp read length, single end sequencing.

2.2.4 RNA-seq – TruSeq bulk tumour samples (Bulk Chapter 3) & TCR-seq.

This is an alternative methodology suitable only for larger amounts of RNA such as whole tumour RNA. RNA quality was assessed using the Agilent 2100 Bioanalyser (Agilent Technologies UK Ltd., Stockport, UK); Total RNA was converted into a library for sequencing on the HiSeq 2500 (Illumina Inc., San Diego, USA) using the TruSeq stranded mRNA Sample Preparation Kit (Illumina Inc.). This followed a similar procedure to the Smart-seq procedure upon amplified cDNA creation, briefly, poly-A mRNA was purified from total tumour RNA (100ng) using the Poly(A) Purist Mag Kit (Life Technologies Ltd., Paisley, UK), reverse transcribed, amplified and sequenced according to the manufacturer's instructions.

Please note the bulk HNSCC TruSeq work was not completed by James Clarke

TCR-seq was completed as described using TRU-seq single index barcodes as described in the methods paper (Shugay et al., 2014).

2.2.5 10x Genomics.

10x Genomics was completed as per manufacturers recommendations and was sequenced on an HiSeq 2500 or 4000 (Illumina).

2.3 In-silico bioinformatics.

2.3.1 Alignment, count estimation and differential expression for chapter 3 and 5.

This calculates a relative expression of each gene and then compares it between particular sets of samples. RNA-seq data was mapped against the hg19 reference using TopHat (Trapnell et al., 2009) by the La Jolla Bioinformatics Core or Dr Panwar, SE reads were then mapped to the human genome (hg19). for visualisation in Chapter 3 RPKM was used which was the standard methodology at the time for gene length normalised expressed. In the later chapters, TPM was used rather than RPKM, all visualisations in chapters 5 and 6 were generated with custom scripts in R unless otherwise stated.

2.3.2 Correlation analysis and heat maps of whole tumour RNA-seq.

This calculates the level of similarity between each sample or gene using correlation analysis (either Spearman or Pearson as discussed per figure). This also includes the visualisation of the heat map.

Raw counts from RNA-seq were processed in Bioconductor package DESeq2, variance was estimated and size factor normalised. Correlations and statistics were performed in Corrplot 0.73, in an R statistical environment (3.1). Principle component analysis and Euclidean distance calculations were performed on data using the setting: mean = 0, variance = 1 normalization in the Qlucore Omics software (3.1.19). In order to identify the 4000 most variable genes, briefly genes with a standard deviation (SD) less than a specified variance cut off from the maximal SD were removed using Qlucore Omics Explorer (3.1.19) (as per (Engel et al., 2016; Seumois et al., 2016)). This removes genes that have a standard deviation less than a particular cut off. Data were visualized in heat maps where each row represents normalised gene expression values for a given gene; each column represents the gene expression for a given tumour: red shading denotes greater gene expression; blue shading denotes lower gene expression. Heatmaps were originally constructed in Qlucore in chapter 3, but in the following chapters all figures were generated in R using ggplot2, heatmap2 and other custom scripts.

Hierarchical clustering (cluster method=average linkage; distance measure=Pearson correlation) of genes and tumours based on their expression profile is reflected in the dendrograms to the left and the top of the heat map when used, as notated in the figure legend.

As a measure of similarity, the analysis was completed through quantified Euclidean distances (heat map distance measure = Pearson correlation) between all replicates from the same patient to all other replicates. This measures distances between samples in the hierarchical tree with smaller distances between replicates indicating a closer relationship. The median distances were calculated for replicates from all other cases (interpatient) and compared to the median distances from intrapatient replicates. T cell gene lists were downloaded from AmiGO (www.amigo.geneontology.org/) and key cancer immunotherapy markers were added (Vanneman and Dranoff, 2012).

To identify genes expressed differentially by various cell types, we performed negative binomial tests for paired comparisons by employing the Bioconductor package DESeq2 (Ganesan et al., 2017; Love et al., 2014) (1.14.1) disabling the default options for independent filtering and Cooks cut-off (Ganesan et al., 2017). We considered genes to be expressed differentially by any comparison when the DESeq2 analysis resulted in a value Benjamini-Hochberg–adjusted *P* value of

< 0.05 and a Fold Change of at least 2. Union gene signatures were calculated using the online tool jVenn.(Bardou et al., 2014) GSEA was completed as previously described in Ganesan et al., 2017 using SNR setting. Visualisations throughout were generated in ggplot2 and heatmap2 using custom scripts, expression values were visualised in Graphpad Prism7. tSNE visualisation were calculated using the first 10 principle components using most variable genes allowing for unbiased visualisation.

For microarray analysis, datasets were downloaded from GEO and data was analysed using Limma (Ritchie et al., 2015) using thresholds of adjusted *P* value of <0.05 and a Fold Change of 2 as completed in the RNA-seq.

Please note Limma Microarray analysis was completed by Dr Panwar.

2.3.3 Additional single cell quality control steps, followed by visualisation of single tSNE plots and single cell clustering.

This reviews the single cell sequencing quality scores and then from the cells that pass the analysis, the analysis generates a 2D representation of the whole single cell (or bulk) dataset for easier interpretations. Each was run through an internal quality control analysis pipeline (Engel et al., 2016; Seumois et al., 2016) and the normalised counts exported. In chapter 3 single cell data DESeq2 normalised counts were used, while chapter 6 used transcripts per million (TPM). To expand on the automated pipeline, the automated parameters were adopted from Gaublomme et al., 2015 (Gaublomme et al., 2015; Yosef et al., 2013). The following parameters from the mapping statistics were used for excluding the cells from further analysis. Analysis sequentially excluded the cells with total raw counts < 250,000 reads, Total Uniquely Mapping Reads < 100,000, sequencing coverage bias and percentage mapping from Picard tools. Based on these parameters cells were classified as low, medium and high quality and eliminated all the low-quality cells from further downstream analysis.

The latest version of Seurat was used (2.1) throughout, downloaded directly from https://github.com/satijalab/seurat (Macosko et al., 2015). Unbiased clustering of single cells was performed using Seurat allowing for clustering and visualisation of data in 2D gene space (Patil, et al 2018).

Monocle was completed with version 2.6.1 and Sincell 1.8.0, using default settings, using the most variable genes identified in Seurat. Average expression was calculated in Seurat, using the *AverageExpression* function, heatmaps were clustered using average linkage.

2.3.4 Single cell differential expression (SCDE and MAST).

This analysis allows comparison of two sets of cells based on Bayesian modelling theory. Analysis was completed as per directed in the manuscript for Smartseq2 data (Fan et al., 2016; Kharchenko et al., 2014). For 10x data MAST was used. To account for different number of cells the SCDE parameter *min.nonfailed* was set to 20% of the total cell number and other parameters were used at default settings. Two-sided P-values were calculated from the Benjamini-Hochberg multiple testing corrected Z-score (cZ) using the normal distribution as null hypothesis. Results are adjusted for multiple comparisons using Benjamini-Hochberg procedure. For the 10x data MAST was used with default settings through converting the 10x data into CPM (counts per million), which was suggested by the developers.

2.3.5 TraCer TCR recovery from single cell RNA-seq.

This calculates the TCR from each singular cell. Tracer was downloaded as per latest version from https://github.com/teichlab/tracer, installed and run using this script as outlined by Stubbington et al,. 2016(Stubbington et al., 2016, 2016). In brief, reads originating from TCR mRNA were extracted from the sequence data for each cell and assembled into contigs (overlapping reads) representing the full-length TCR sequences expressed in each cell. Clonally related cells were identified by finding instances where more than one cell had identical TCR sequences. To be stringent we selected only cells that shared both alpha and beta chains as related clones. T cells were mapped onto single cell transcriptome in R statistical environment.

2.3.6 Gene set enrichment analysis.

T cell gene lists were downloaded from AmiGO (www.amigo.geneontology.org/); key cancer immunotherapy markers were added. Pathway images were generated using Ingenuity Pathway Analysis (IPA). Gene set enrichment analysis was completed using the free tool *Toppfun (*https://toppgene.cchmc.org/enrichment.jsp) or Cluego (using default average settings)(Bindea et al., 2009).

The Qlucore Omics Explorer software package was used for GSEA analysis. GSEA was used to further assess whether specific biological pathways or signatures were significantly enriched between two groups. GSEA determines whether an *a priori* defined 'set' of genes (such as a signature) show statistically significant cumulative changes in gene expression between phenotypic subgroups. In brief, all genes are ranked based on their differential expression between two groups. Next, a

running enrichment score (RES) is calculated for a given gene set based on how often its members appear at the top or bottom of the ranked differential list. 1000 random permutations of the phenotypic subgroups are used to establish a null distribution of RES against which a normalised running enrichment score (NES) and FDR-corrected q values are calculated using Kolmogorov-Smirnov statistic, which quantifies differences in data distributions.

2.3.7 WGCNA analysis.

We employed WGCNA version 1.51 in R 3.4.0 throughout. We created a reference weighted signed network from the T-CD103$^+$ dataset or NL-CD103$^+$ dataset (Langfelder and Horvath, 2008; Oldham et al., 2006; Ottensmeier et al., 2016) using genes that passed a quality metric (gsg) in WGCNA in every dataset and were expressed greater than 2 TPM in at least 25% of the total samples. We calculated the weighted correlation matrix by raising the correlation to the power 8. We transformed this into a matrix of connection strength termed an "adjacency matrix", to calculate the topological overlap (Langfelder and Horvath, 2008). Using the cutreeeHybrid function (Miller et al., 2010; Oldham et al., 2006) with a minimum cluster size of 50, deepsplit = TRUE, the transcriptome was assigned into modules (Morgan et al., 2013). A module dissimilarity threshold was set at 0.3 to merge significantly correlated modules. Module eigengenes were correlated (spearman) to the IHC count rank. Correlations that were not significant were coloured white on the heatmap, with the red or blue being used to signify the direction of the significant correlation. Module preservation was calculated using the Zsummary.pres statistic using the Tumour/NL CD103$^+$ dataset as a reference point for comparison to the others as previously described (Dillies et al., 2013; Miller et al., 2010; Oldham et al., 2006). Values greater than 10 were considered highly preserved, and greater than 5 moderately conserved. A gold module of 100 random genes and the grey module of unassigned genes was used as a negative control, 30 permutations was used (Langfelder et al., 2011). Visualisations were in ggplot2 and heatmap2 using custom scripts in R. Networks were visualised in Gephi (0.9.2) using Force Altas2 and Noverlap for visualization purposes. Nodes were coloured by the WGCNA assigned correlation to clinical statuses.

Chapter 3: Results 1: Inter- and intra patient immune heterogeneity in solid tumours.

3.1 Investigation into intra-tumoural and intra-patient bulk transcriptome heterogeneity.

3.1.1 Introduction.

This program of work requires isolating a tumour and completing high resolution analysis. However, morphological heterogeneity in cancer has long been appreciated by pathologists (Alizadeh et al., 2015; Hensley et al., 2016; Jamal-Hanjani et al., 2015; Junttila and de Sauvage, 2013; Kim et al., 2014). Recent data show that within a given tumour, the tumour cell clones may undergo marked diversification over space and time as a result of mutational divergence on the one hand and immunological selection on the other (Alizadeh et al., 2015). It has further become evident that immune attack may vary according to its location in the cancer tissue when quantified at the protein or gene expression level using immunohistochemistry or transcriptome (Satija and Shalek, 2014). Therefore, this suggests that when quantifying the density or quality of tumour infiltrating immune cells, the spatial distribution of immune cells in relation to tumour cells and the other components of the tumour microenvironment, need to be taken into account. In turn, this suggests that small biopsy samples may not accurately represent the tumour microenvironment or immune status of a particular patient's tumour. In practice however, it is difficult to sample multiple tumour areas in most primary human malignancies at any time other than at surgical resection.

This entire thesis is reliant upon the collection of fresh patient tumour infiltrating lymphocytes. This therefore raises the possibility that tumour heterogeneity at the genetic and pathology level may be a confounding factor. In lung disease, the ability to collect multiple biopsy and resection samples is limited as patients rarely undergo both a biopsy and a full tumour resection. Additionally, where a second invasive procedure is completed, it is rare for there to be sufficient tissue to provide excess material for research purposes. Consequently, this section focused on HNSCC where multiple sampling is clinically feasible as the tumour is located superficially in the oropharynx or oral cavity.

As described in multiple other solid tumours and following the seminal findings of Galon, et al,. (Galon et al., 2006), valuable prognostic information can be gained from simple enumeration of the T-cell infiltrates in both HPV(+)and HPV(-) head and neck squamous cell carcinoma (HNSCC) (Ward et al., 2014; Wood et al., 2016). Following simpler analysis of IHC, comprehensive transcriptome analysis of primary HNSCC and lung cancer have allowed further characterization of molecular signatures globally (Hammerman et al., 2012; Imielinski et al., 2012; Lawrence et al., 2015; Zhang et al., 2014). These transcriptome analyses have used RNA-seq to evaluate gene expression and offer unprecedented insight into biological processes that occur in tissue. To make sense of any changes detected in longitudinal assessments, understanding of natural variability is critical, as is

the understanding of how results from different assays (e.g. histological and transcriptome assessment) correlate with each other.

Consequently, this chapter comprises two separate sections.

1) In the first section the intention was for transcriptome analysis of whole tumour ("bulk transcriptomics") to assess the level of intrinsic variability in whole biopsies of HNSCC tumours, across temporal and spatial replicate samples, in the absence of any clinical intervention. Therefore, to allow analysis of spatial variability within a tumour, a resection sample was subdivided into multiple "biopsies" referred to as spatial replicates. To analyse possible variation over time, in 14 patients, an initial biopsy was taken followed by resection of the tumour within a four-week period (standard HNSCC clinical management). These samples of the same tumour, taken over a 30-day interval, are referred to as "temporal replicates". The tumours included both HPV(+)and HPV(-) HNSCC, in case of any aetiology related differences.

   For patients with multiple repeats at each time point, these samples were included in both the spatial and temporal analysis.

2) The next section of this chapter aimed to investigate whether this 'bulk' transcriptome interrogation was masking a degree of heterogeneity that could only be detected at the single cell level. Owing to recent advances in molecular and computational biology, the aim was to see if it was possible to take primary human malignancies and isolate T cells from them for high-resolution single cell immunophenotypic investigation. This preliminary experiment in the context of checkpoint blockade was completed to demonstrate the power of the technology and provide a baseline from which to work for the rest of the thesis.

3.1.2 Homogeneity and heterogeneity revealed by bulk tumour analysis.

3.1.3 Cohort characteristics.

To appropriately assess the variability in transcriptome phenotype, multiple biopsies were taken from discrete areas, at least 1cm apart. These multiple biopsies were taken at the time of clinical diagnosis and during Tumour resection from 14 patients (final number for sequencing is 44). RNA was isolated and subjected to RNA sequencing for each replicate tumour sample. Samples with transcript bias (Faherty et al., 2015; Gallego Romero et al., 2014) or low RNA integrity number (RIN) score were excluded (n=5), with their concordant sample (n=2); excluding 2 patients for a final cohort size of n=37 biopsies from 12 patients (Table 9). This cohort contains a combination of locations, differentiation status and smoking history.

Please note that the sample processing was completed by Mr Oliver Wood and Ms Emma King

Number	Analysis	Age	Smoking Status	Location	T	N	M	Differentation
1	Spatial	68	Non Smoker	Base of Tongue	3	2	0	Poor
2	Spatial	60	N/A	Larynx	4	0	0	Poor
4	Spatial	72	Ex-Smoker	Base of Tongue	2	2	0	Poor
8	Spatial	64	Smoker	Pyriform Fossa	4	2	0	Moderate
11	Spatial	61	Smoker	Base of Tongue	1	0	0	Moderate
12	Spatial	63	Smoker	Pyriform Fossa	4	2	0	Poor
15	Spatial	47	Smoker	Tonsil	3	2	0	Poor
16	Spatial	62	Smoker	Floor of Mouth	2	1	0	Moderate
3	Spatial and Temporal	N/A	N/A	Tonsil	3	1	0	Poor
13	Spatial and Temporal	57	Smoker	Tonsil	1	2	0	Poor
14	Spatial and Temporal	59	Smoker	Larynx	4	1	0	Poor
5	Temporal	61	N/A	Base of Tongue	1	2	0	Moderate
6	Temporal	76	N	Larynx	2	0	0	Well
7	Temporal	50	Non Smoker	Tonsil	2	1	0	Moderate
9	Temporal	48	Smoker	Tonsil	2	2	0	Poor
10	Temporal	63	Smoker	Larynx	4	2	0	Well

Table 9. Cohort information for the HNSCC tumour heterogeneity RNA-seq analysis.

Each patient was given a numerical identifier in this cohort, with the sample being annotated as either Spatial (red), Spatial and temporal (blue) or just temporal (yellow). The patient's age, smoking history and location of tumour, alongside the clinical staging is included. Samples are ordered by sample ID, inside the demarcation for the type of analysis (spatial or temporal) completed.

3.1.4 Consistent Immune signature between biopsy and resection specimen when analysed with Immunohistochemistry (IHC).

In order to investigate the variability in expression of CD8 expression at the protein level, the analysis used the more traditional IHC technique (Figure 9). This compared the initial diagnostic biopsy (INC) and the resection (EXC) sample taken up to 30 days later. This suggested that for this marker, expression was relatively stable over time. However, to investigate multiple markers, the use of systems, high resolution techniques which provide a more comprehensive analysis was required.

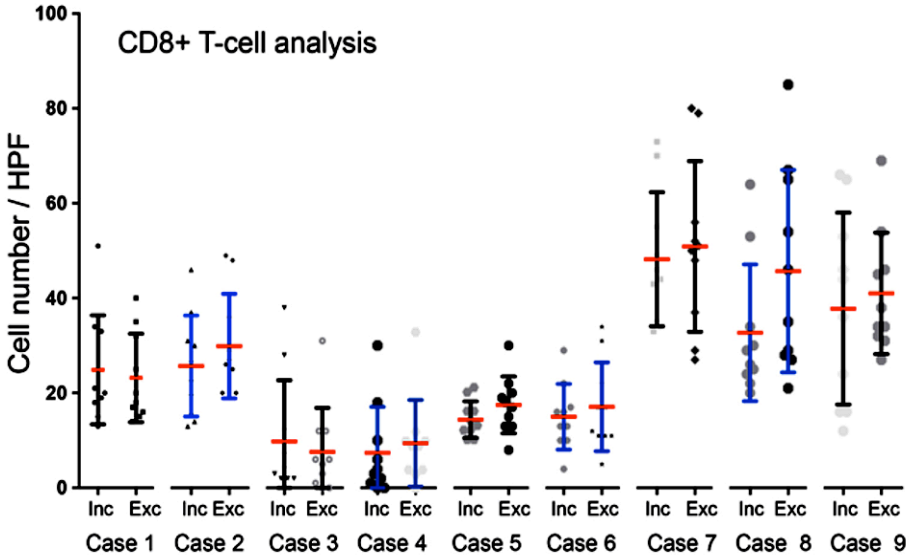


Figure 9. Temporal heterogeneity of $CD8^+$ protein expression in incisional (INC) and excisional (EXC) biopsies.

Immunohistochemistry staining on temporal samples was completed and 10 high power fields for each individual sample was counted from each slide. This gave a relatively consistent profile from which could be continued into the more complex transcriptome analysis. No samples were statistically different between each time point using a paired non-parametric test (Wilcoxon signed-rank test, $p > 0.05$).

Note Professor Thomas completed this set of CD8 IHC analysis on an additional cohort (A).

3.1.5 Variance filtering does not alter a sample's location in gene space.

Variance filtering has become standard in many transcriptome studies; this technique removes genes that represent subtle fluctuations in technical noise and are not variable across samples (Breen et al., 2015a, 2015b). This removal is important, as these genes artificially increase the correlation between all samples, which reduces the statistical power of any further analysis. In order to investigate the spatial and temporal dynamics of RNA-seq data, the data was to be visualised in reduced dimension space using principle component analysis (PCA) which allows a 'human' analysis by reducing the data down to axis of the most variation in the dataset.

However, how this filtering process would alter this data in 3D principle component gene space was unproven. It could be hypothesised that the filtering should make limited difference, as PCA uses the most significantly variable genes to construct the analysis. Therefore, the HNSCC tumour heterogeneity dataset was titrated at various intensities to the genes that were the most variable at the noted threshold, in Figure 10. As variance filtering removes the lowly expressed genes it did not appear to alter the 3D location of gene in the reduced dimension space (Figure 10). Consequently, the most variable genes were selected for the remainder of the analysis. This is important as the variance filtering maximises the chance of there being significant differences with downstream methods that are going to be applied in this chapter.

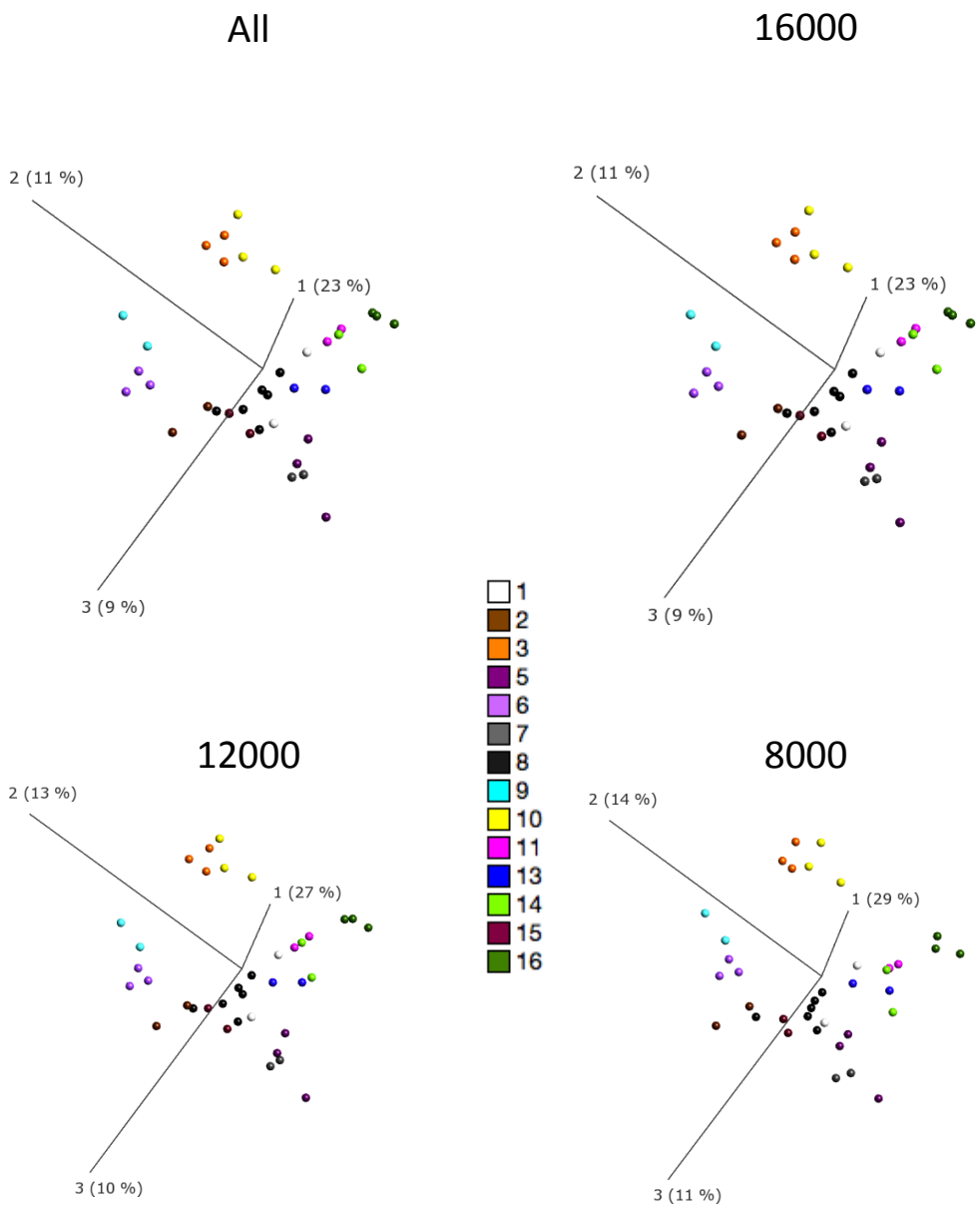


Figure 10. Impact of variance filtering on PCA sample location.

The HNSCC tumour heterogeneity dataset was $\log_2$ transformed and variance normalised as per Engel et al (Engel et al., 2016). The dataset was then filtered by the levels noted above the figure. The data was then visualised in 3D gene space using principle component analysis. This showed limited impacts on the location of the individual samples, therefore, for the next steps in this section, the analysis maintained the use of investigating the genes with highest variance to increase the chance of finding significant differences between each replicate by removing the low-level noise that would inflate the baseline of any correlations, reducing the chance of detecting significant changes. The numbers on each axis represent the first three principle components followed by the percentage of total variation in the data contained in that axis in brackets.

3.1.6 Spatial and temporal replicate samples cluster hierarchically and are significantly related.

Variance filtering of the RNA-seq data was carried out to display the most variable genes across the cohort, as optimized in the previous figure. These were hierarchically clustered using average-linkage dendrograms and displayed in a heat map (Figure 11). In most cases, the spatial replicate samples from the same patient clustered directly with the other samples from the same tumour. The exceptions were cases 2 and 14 and which clustered within the next branch of the dendrogram but were not immediately adjacent to their matched sample, suggesting a level of variability. The same genes and samples were visualized with principle component analysis (PCA). This showed clear grouping of replicate samples, complementing the hierarchical clustering.

The same approach was used to assess the relationship between temporally separated replicates (Figure 12). Hierarchical clustering showed these samples to locate with their replicate samples in all cases. PCA analysis also displayed close groupings of the temporally separated replicates in each case. However, this analysis, although a visual demonstration, fails to quantify the variation; therefore, the next stage was to quantify this variation using two methods that have been used extensively in the literature inter-sample correlation and the Euclidean distance between samples (D'haeseleer, 2005; Langfelder and Horvath, 2008).

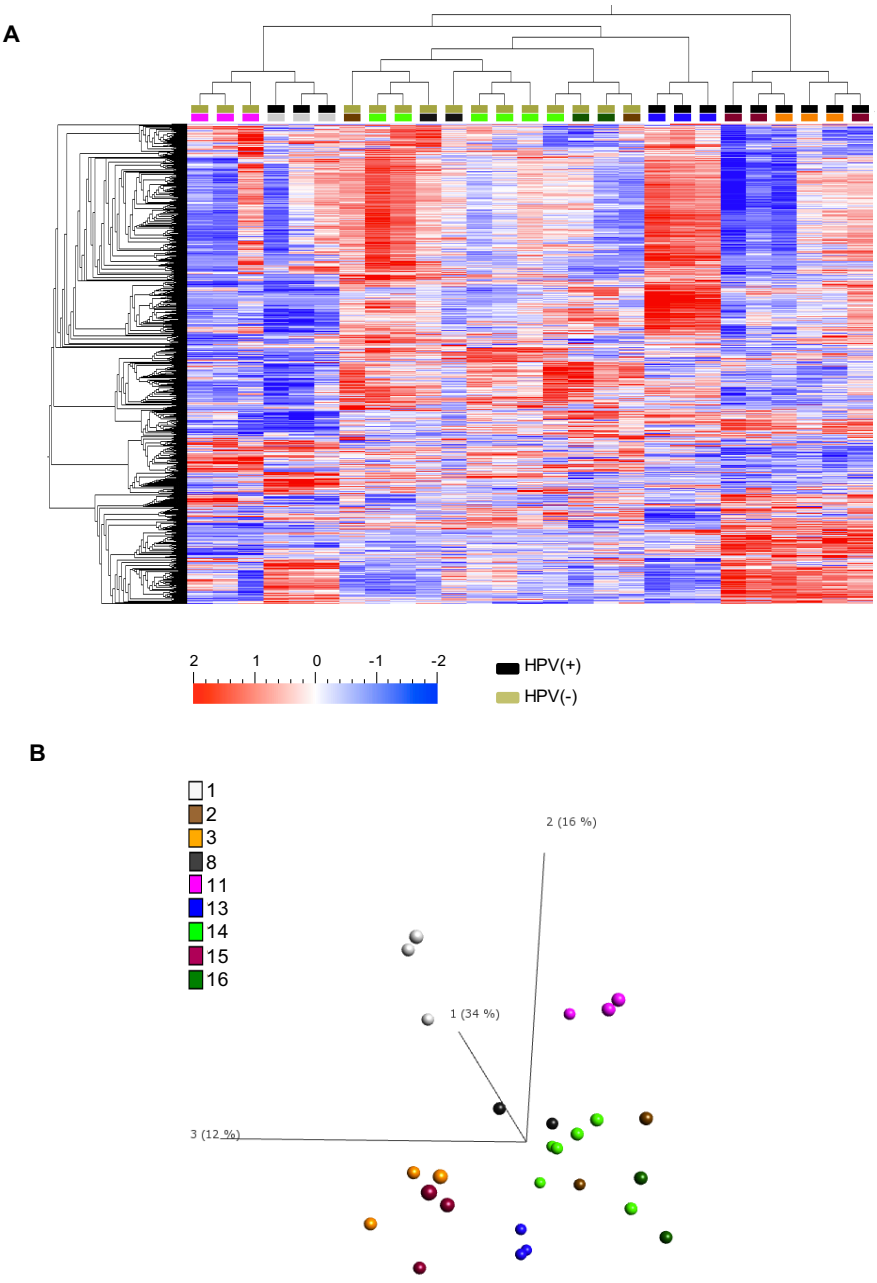


Figure 11. Variation in transcriptomes from spatially separated replicates.

RNA-seq data was variance filtered to display the most variable genes across the tumour replicates. Patient tumour replicates are colour coded and displayed on the heatmap and Principle Component Analysis (PCA)plot; HPV(+)= black and HPV (-) = beige. (A) Hierarchical clustering of single time point tumour replicates displayed as a heatmap shows close clustering of related samples. (B) PCA was used to visualize the sample to sample distances and also highlights the similarities between the tumour replicates. The numbers on each axis represent the first three principle components followed by the percentage of total variation in the data contained in that axis in brackets.

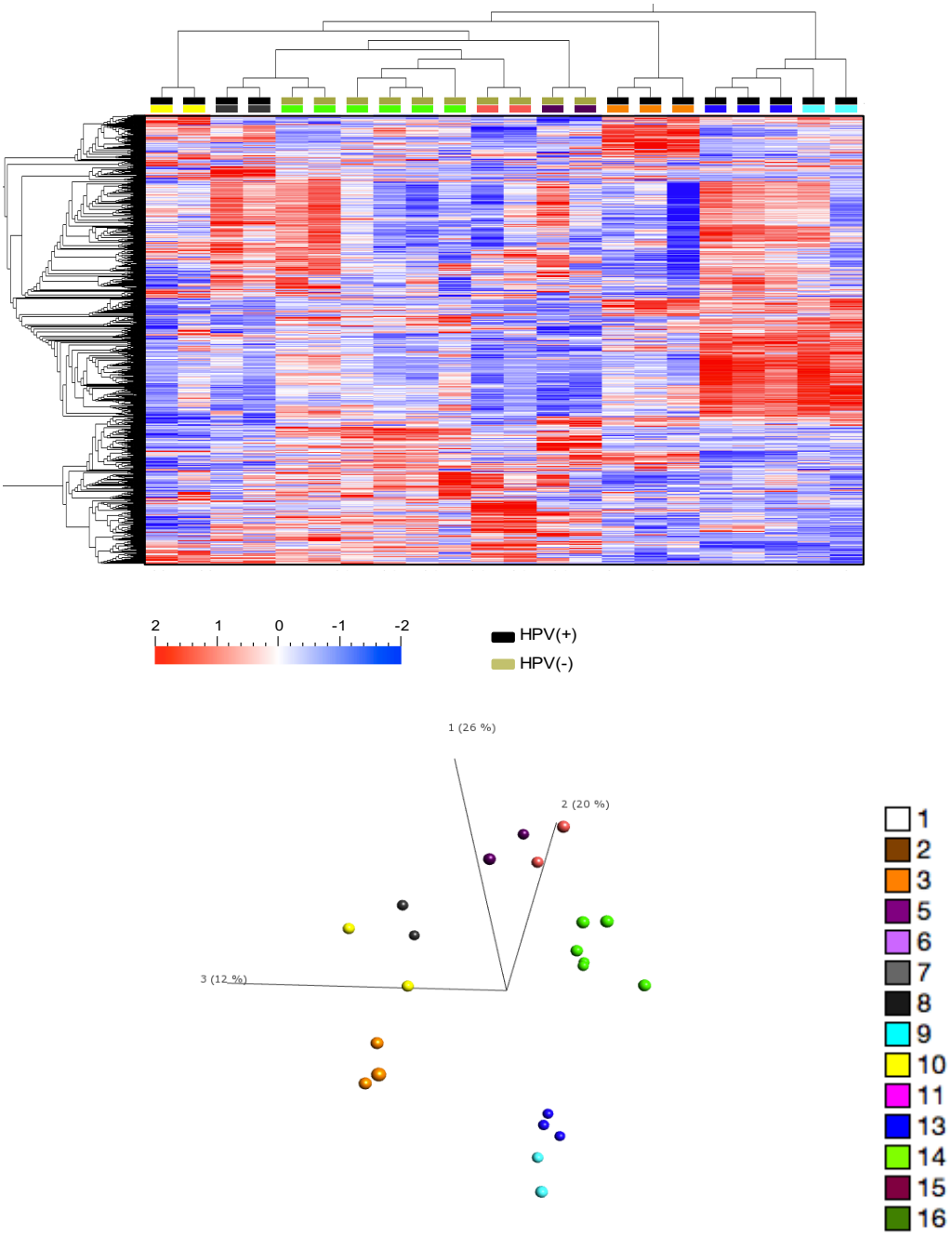


Figure 12. Variation in transcriptomes from Temporally separated replicates.

(A)Tumour replicates from patients at two time points (biopsy and resection) are displayed as a heatmap and hierarchically clustered the tumour replicates by case number in all instances. PCA also shows the clustering of the tumour replicates in low dimension gene space. (B) The numbers on each axis represent the first three principle

components followed by the percentage of total variation in the data contained in that axis in brackets.

3.1.7 Quantifying the inter-patient and intra-patient differences using Euclidian distance metric.

In order to quantify the level of variation between individual patients and within replicate samples, the Euclidean distance between samples was assessed between the related and unrelated replicates (Figure 13). Following which the distances were exported to Prism6 for visualisation.

By use of the Euclidean distance measure, the data demonstrated that replicate samples from any given patient were significantly closer to each other than to the average of the unrelated samples (*P* =<0·0001). No significant differences in the variability between spatially and temporally separated replicates was observed (p = 0·62). In addition, Euclidean distances between HPV^+ and HPV^- Tumours were calculated. The mean distance between related samples were not significantly different for HPV^+ as opposed to HPV^- Tumours (Mann-Whitney test; 0·32), although perhaps due to the small sample size of this cohort, assessment of mean distance between related samples might be underpowered.

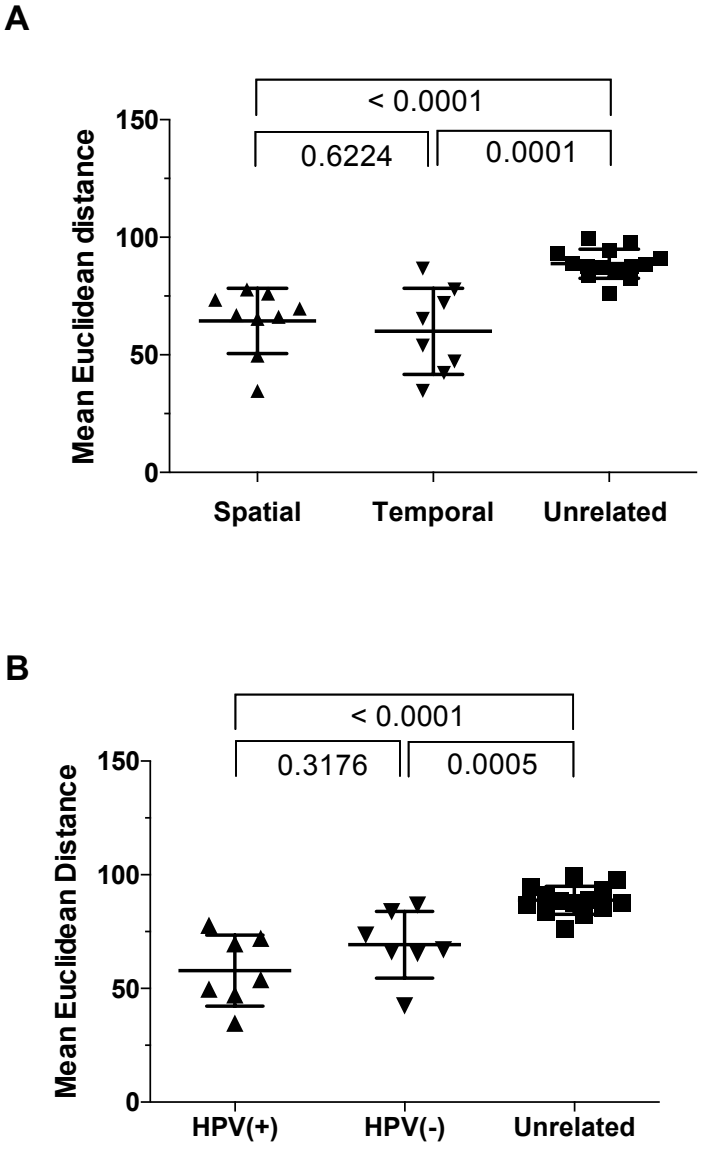


Figure 13. Demonstrating sample similarity with euclidian distance.

The distance between each sample in the hierarchical tree was calculated and used to assess how close autologous related samples were compared to heterologous ones. The mean Euclidean distance for related samples was plotted alongside the mean Euclidean distance for unrelated samples. (A) This represents spatial and temporal replicate distances for related and unrelated samples displaying a significant difference ($p<0.0001$, Mann-Whitney test) between samples that are related to those that are unrelated. Samples that are matched (autologous) cluster closer in the hierarchical tree than they do to unrelated (heterologous) samples (B) Assessment of Euclidean distance between samples from HPV^+ and HPV^- tumours displayed no significant difference with both being more similar to autologous samples than to unrelated ones. A trend towards HPV^- samples being closer to unrelated samples was observed.

3.1.8 Temporally and spatially separated replicate samples are highly correlated.

Correlation analysis of the top variance-filtered genes was carried out across the spatially and temporally separated replicate samples (Figure 14). The displayed correlation matrix had observed R^2 values ranging from 1 (highly correlated, yellow) to R^2= 0·55 (less correlated, blue). The spatially separated replicates show a high level of correlation to their related samples (Figure 14). Case 14 from whom six replicates had been taken, shows a variable level of correlation across the replicates but nonetheless a closer correlation between its own replicates than to other cases. The temporal replicates were also more closely correlated to each other than to unrelated samples. Interestingly, the correlation matrix identified that some cases (i.e. case #13 and 9) also correlated with each other, reflecting similarities in the tumour at the transcriptome level. Through quantifying the average correlation between related and unrelated replicate samples it was identified that in the spatially separated replicates the average R^2 for related samples was 0·8299 compared to the unrelated samples; R^2=0·6577. The dataset for temporally separated replicates yielded similar data with an R^2=0·8357 for related samples and R^2=0·6411 for the unrelated samples. These data suggest that each patient's tumour has a distinct transcriptome landscape, distinct from that of other cases. When using traditional differential expression profiling the analysis was unable to find any major significant differences in the cohort (DESeq2: corrected P <0.05, Log_2 Fold change >2).

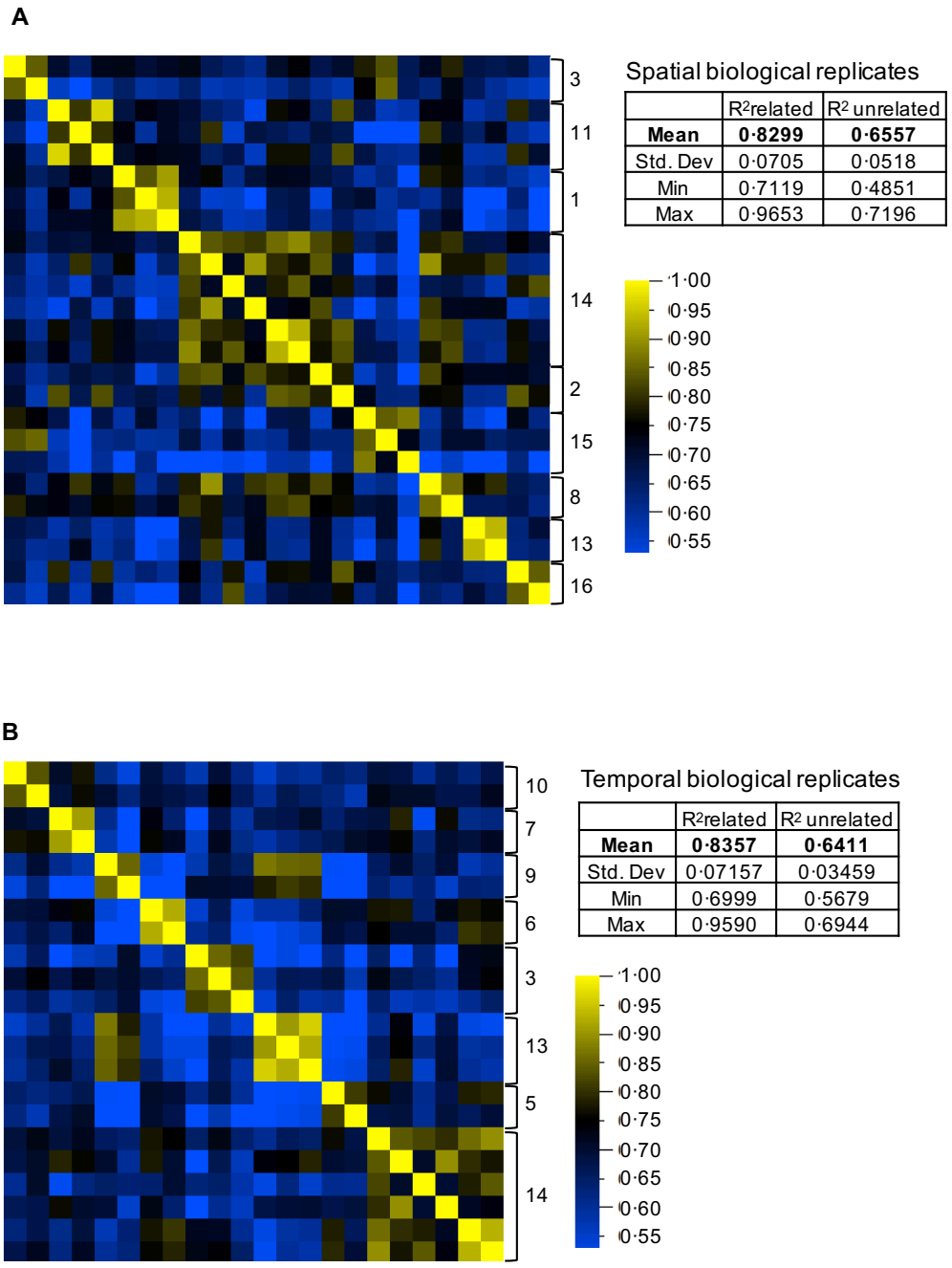


Figure 14. Correlation matrix analysis of HNSCC transcriptomes.

(A) Spatially separated replicates were assessed using a correlation matrix showing that autologous samples were more correlated than unrelated samples (Spearman correlation of top variance filtered genes). (B) Correlation analysis was also carried out on temporally separated replicates with related samples being more correlated. This final figure was done in collaboration with Oliver Wood for the manuscript.

3.1.9 Strong correlation of IHC and gene expression for the T-cell marker CD8 in this cohort.

In order to validate our findings on immune gene expression across the groups of replicates, a colleague undertook IHC for CD8 on the same frozen tissue used for the RNA-seq; this was not possible in 3 cases due to insufficient material. $CD8^+$ counts evaluated by immunohistochemistry using 10 HPF, mirrored that of the CD8A gene expression with a correlation of (R^2 =0·8424). Demonstrating that the gene expression data was reflected at the protein level.

Having now investigated the heterogeneity at the bulk level, the next stage was to begin investigating how each individual sub-population of immune cells varied. This following work primarily focused on the $CD8^+$ population of tumour infiltrating lymphocytes owing to their established role in mediating anti-tumour immunity.

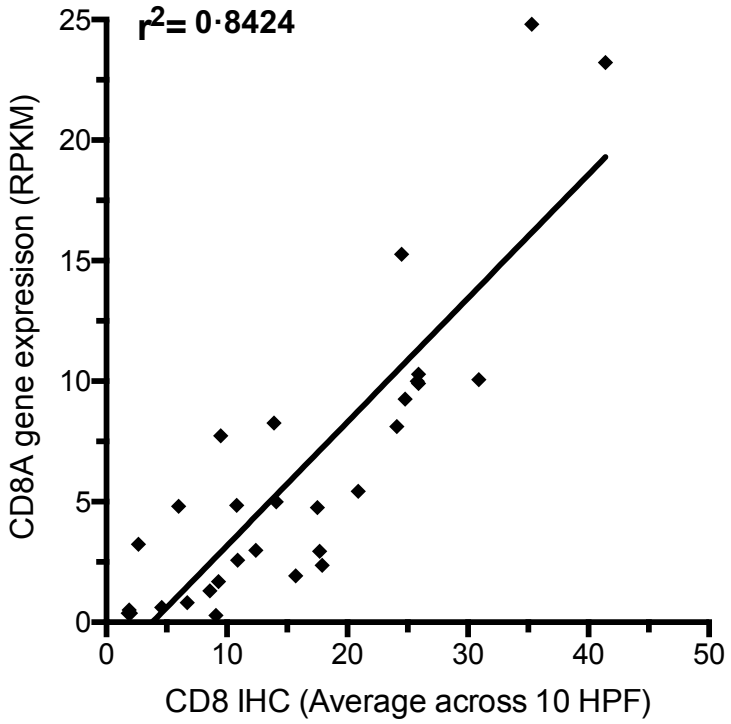


Figure 15. Correlation analysis of CD8 gene expression and CD8 IHC.

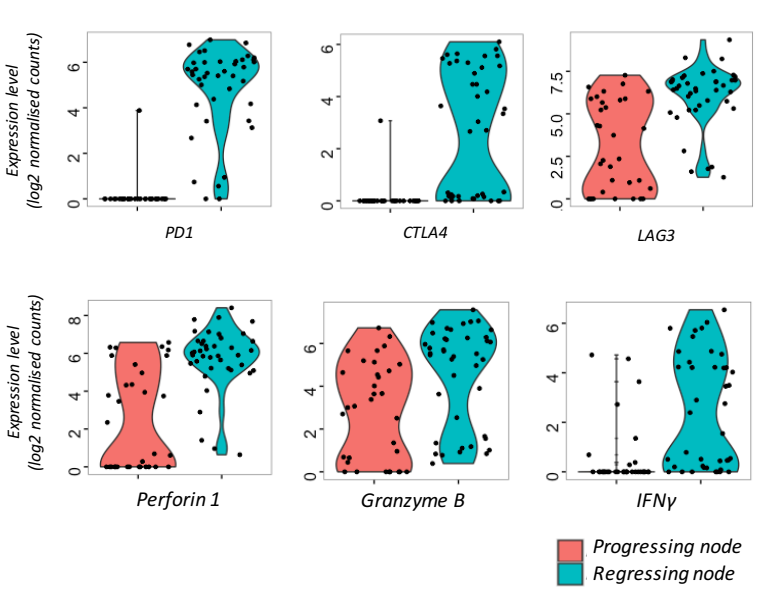
Gene expression for the samples cohort was compared to IHC CD8 counts across 10 high powered fields. Spearman correlation analysis shows an R^2 of 0.84 between IHC and RNA-seq for CD8A (A), showing strong correlation between the gene expression and protein quantification.

This data analysis was completed by Mr Oliver Wood.

3.2 Mapping of expanded TCR clonotypes in response to aPD1 therapy in an example case for method validation.

The T cell response that accompanies clonal divergence of tumour cells in human metastasis is unknown at the single cell transcriptomic level. We had the opportunity to study T cells in different metastases from a patient who was undergoing immunotherapy. This patient had previously received aCTLA4 therapy which was then followed by an aPD1 regime. With this treatment, the patient had two metastatic tumour masses; one that clinically responded to the treatment and one that did not. Through single cell sequencing on 48 $CD8^+$ cells isolated from a tumour mass that was responding to checkpoint blockade and one that was not responding. In order to understand the clonal variation and expression profiles of each T cell under the pressure of aPD1 therapy, this was accompanied by single cell T cell receptor reconstruction. We identified that the profile of a T cell in the responsive mass contained a much stronger cytotoxic gene expression profile (Figure 16A) with a clonal TCR (B). By contrast, in the unresponsive mass, the levels of PD1 were much lower than the responsive mass. Interestingly, we hypothesise that the CTLA4 treatment potentially could have left a long-term effect on the T cell profile owing to the lack of CTLA4 expression in the unresponsive mass. Furthermore, we noted the clustering of individual single cells from each mass locating in two major clusters in the tSNE visualisation. The TCRs identified that were repeated in the responsive mass did not maintain a completely distinct gene expression profile (Figure 17). This suggests there to be a degree of underlying heterogeneity that was lost with whole tissue transcriptomic analysis. However, single cell analysis has sensitivity issues owing to the nature of the non-normal distribution of RNA expression (Raj and van Oudenaarden, 2008). Therefore, it appears that it is vital to combine multiple techniques, with gene and protein expression profiling to understand the biology of a particular system.

(A)



(B)

Clonotype	Vβ–Jβ	Vα-Jβ	Clonally related cells/ total cells with rescued TCR Vβ/Vα
1	Vβ13-J2-3	Vα14 – Jα57	8/30
2	Vβ18-J1-3	Vα21 – Jα26	3/30
3	Vβ27-J2-7	Vα35 – Jα40	2/30

Figure 16. Single cell RNA-seq analysis of sorted CD8⁺ TIL from two lymph nodes with a divergent αPD1 response in a patient with metastatic melanoma.

CD8⁺ T cells were isolated (Wood, et al,. 2016), sorted into single 96 well plates and sequenced (Engel, et al,. 2016). Gene expression was calculated as a Log2, library size normalised and transcript normalised counts with expression visualised through a violin plot. Single CD8⁺ TILS isolated from the progressing node (N=32) in red and the regressing node turquoise (N=40) (A). The reconstruction of T cell receptor (TCR) sequences was completed using the TraCeR software. Transcripts were aligned using a reference genome (the international ImMunoGeneTics information system that contains nucleotide sequences for all possible human TCR V, D and J genes and constant regions. TraceR identified and flagged any non-productive TCR transcripts to call the correct productive TCR gene segments including the CDR3 motif. This information was combined with the digital gene expression matrix and visualised in an R statistical environment (v3.2). Gene segments for Vb and Va are displayed and the number of clonal repeats identified (B).

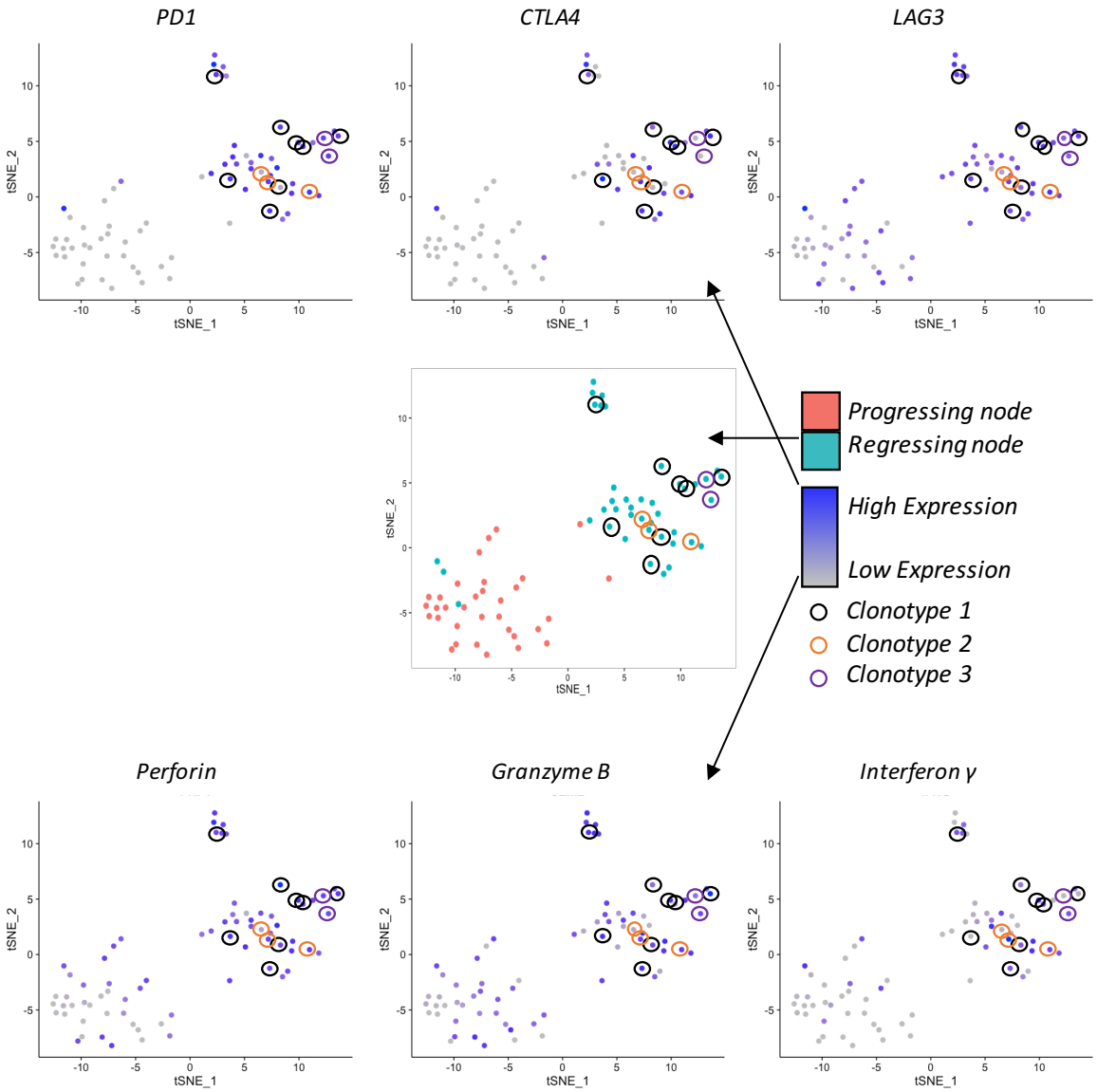


Figure 17. Mapping TCR clonality onto single cell gene expression profiles in response to checkpoint blockade.

$CD8^+$ T cells were isolated as per Wood, et al,. (2016); sorted into single 96 well plates and sequenced as per Engel, et al,. (2016). Cells that passed quality control were visualised in 2D gene space through the t-Distributed Stochastic Neighbour Embedding (t-SNE) algorithm. The TILS isolated from the 'progressive node' (N=32) are demonstrated in red and the 'regressing node' in turquoise (N=40). The expression level of key markers is represented through colour intensity with the identification of the gene of interest induced above. The repeat clonotypes 1-3 are ringed in black, orange and purple respectively.

3.3 Chapter 3 Discussion.

The results in this chapter contain several key findings. Initially, we demonstrated the ability to quantify the features in a biopsy to define a representative immune profile of an individual's tumour. This is owing to the conserved profile identified for each patient's tumour (Figure 9-Figure 15). When this variation was quantified using both Euclidean and Correlation methodologies, each patient's sample was significantly more similar to each other than to the rest of the cohort (Figure 14). Ideally, for a number of markers, the protein and gene expression would have been correlated together, however, owing to the lack of material this proved unachievable. This work suggested that despite the known genetic and some transcriptomic variation across a cohort and within a particular tumour type, the unique overall immune profile of each patient was maintained over time and space. This is vital information for any future clinical trials (which at the time of writing were currently recruiting), as understanding the intrinsic biological variability allows accurate interrogation of any changes induced by the clinical intervention. Future work on this cohort will also include exome analysis, to identify if changes in the genome and transcriptome are matched or divergent. This represents work being completed by another member of the group.

It is well established at this point that whole tumour profiling, although powerful, has reduced resolution as it takes the average of all cellular populations. The first paper to complete single cell profiling on human tumours was recently published in Science (Tirosh et al., 2016), which identified a large degree of variation that was masked at the whole tumour level. A concern was that since only ~50 cells were sequenced per location, it is possible that the full distribution of cells (and thus TCR clonotypes) within the spatially distinct cell sub populations may not have been captured. This is reliant on sequencing more cells per patient, which will be done in future work. From a technical perspective, we had demonstrated the ability to isolate and phenotype TILS from patients.

We were particular interested in the shift in expression profiles between the responsive and unresponsive tumour masses in terms of transcriptome profiling (Figure 16). The presence of clonal expansion in the TILS (Tumeh et al., 2014) in checkpoint blockade-responsive T cells was previously demonstrated using bulk TCR sequencing in response to aCTLA4 and aPD1 (Robert et al., 2015b; Roh et al., 2017). It was interesting to identify these phenotypic changes between the different conditions; however, it must be noted that this represents a small cohort, with low statistical power. A larger cohort is undergoing collection currently as part of future work.

Recent papers have demonstrated the throughput of 10X and droplet based technologies for analysis of thousands of cells at much higher throughput and lower cost, which represents the most suitable technique for large-scale analysis of unknown populations. However, for more focused

analysis of populations, there are several issues particularly for lymphocytes. A primary issue is that these methods which employ a 3' unique molecular identifier, fail to gain full coverage of an individual transcript. Consequently, these are unable to fully recover the TCR (Klein et al., 2015; Macosko et al., 2015). There are published reports of high throughput technologies to analyse the TCR, but then these fail comprehensively to characterise the cellular phenotype (Han et al., 2014; Simon et al., 2014; Turchaninova et al., 2013). Additionally, the Smartseq2 methods have been demonstrated to be much more technically robust with higher sensitivity (approximately two to three times), for T cells which have extremely low RNA amounts (~0.5pg), and this methodology could be suggested to be the most suitable unless high numbers of cells are analysed (Ziegenhain et al., 2017). The addition of unique molecular identifiers could also improve SMARTseq methodologies further or high numbers of Unique Molecular Identifier (UMI) based methods to compensate for the assay's reduced sensitivity.

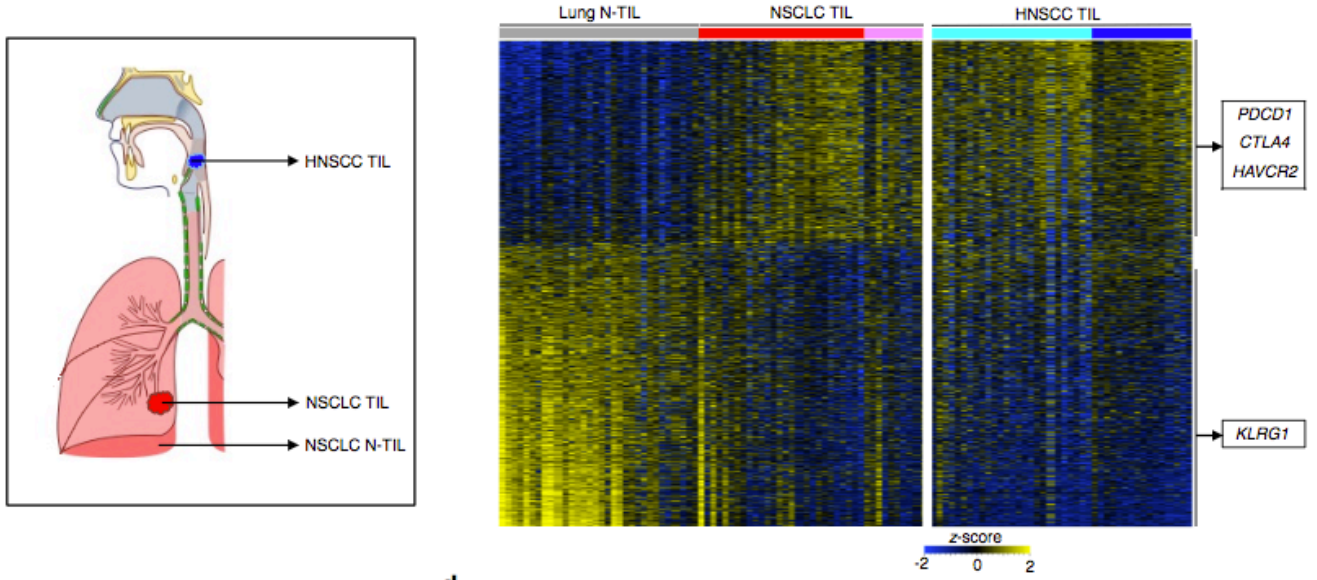
To conclude, in this chapter it has been demonstrated that biopsies or whole tumour resections, analysed by whole tumour transcriptome methodologies, reveal that within any given patient, the immune landscape is relatively constant in different locations and over time within different samples of the tumour; the immune signature is more consistent within-patient than between-patient. Although this allows analysis of a tumour biopsy, it does not interrogate individual cell populations. Therefore, as a preliminary cohort, a small number of $CD8^+$ T cells was analysed at the single cell level in both primary and post-treatment settings. We were able to identify clonal expansion and subtypes with only < 50 cells, suggesting that this was an avenue to explore for the cellular populations being hypothesised from concurrent projects.

Chapter 4: Results 2: Investigation of CD103$^+$ tumour infiltrating lymphocytes through flow cytometry.

4.1 Introduction to CD103$^+$ tumour infiltrating lymphocytes.

In the previous chapter the focus was to look at the entire CD8$^+$ population either through bulk whole tumour RNA-seq approaches or through single cell analysis of sorted populations. This has two issues; firstly, the lack of resolution of the CD8$^+$ in bulk tissue and secondly, the lack of sensitivity of single cell data (Engel et al., 2016). Consequently, a 'best of both' approach would be to isolate a cell population using flow cytometry and complete RNA-seq on this small population. This would allow high resolution analysis at a sensitivity similar to bulk analysis but on a particular population. Our group has applied this to the CD8$^+$ population in our recent publication (Ganesan et al., 2017). This analysis found that the transcriptome signature of a CD8$^+$ cell was maintained regardless of tissue site and disease aetiology, with highly variable expression of exhaustion molecules being found throughout the dataset, when compared to the background Lung tissue (Figure 18). However, the main differences in the transcriptomic profile appeared to occur when comparing the CD8$^+$ cells from a tumour with low number of immune infiltrates and those from tumours with high levels of immune infiltrate. Furthermore, there was a statistically significant correlation between CD8A and CD103 (*ITGAE*) found at the whole tumour level (Figure 19).

This analysis identified that a number of markers of tissue residency were differentially expressed between the tumours with differing TIL status. Importantly, when using a historical cohort with early stage disease, the differences in survival outcome between CD103 high and low infiltrates was clearly demonstrated. Although CD103 is a well-established marker of tissue resident CD8$^+$ T cells in mice (Mueller and Mackay, 2015; Wong et al., 2016) there was at the time, limited evidence in humans of these markers being relevant owing to the difficulties in obtaining human tissue (Cheuk et al., 2017; Schenkel and Masopust, 2014).


HNSCC TIL
NSCLC TIL
NSCLC N-TIL
Lung N-TIL
NSCLC TIL
HNSCC TIL
PDCD1
CTLA4
HAVCR2
KLRG1
z-score
-2
0
2

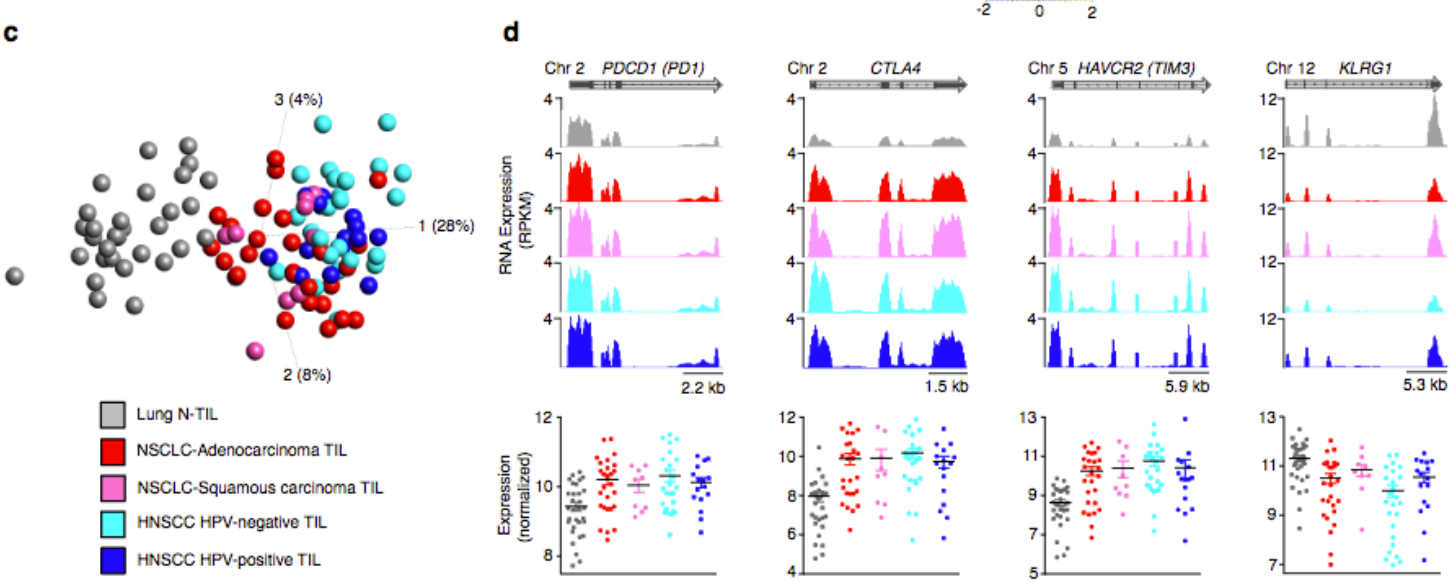

c
3 (4%)
1 (28%)
2 (8%)
Lung N-TIL
NSCLC-Adenocarcinoma TIL
NSCLC-Squamous carcinoma TIL
HNSCC HPV-negative TIL
HNSCC HPV-positive TIL
d
Chr 2 PDCD1 (PD1)
Chr 2 CTLA4
Chr 5 HAVCR2 (TIM3)
Chr 12 KLRG1
RNA Expression (RPKM)
2.2 kb
1.5 kb
5.9 kb
5.3 kb
Expression (normalized)

Figure 18. Transcriptomic analysis of tumour Infiltrating lymphocytes.

Schematic figure of samples used in this chapter and onwards, HNSCC, head and neck squamous cell cancer; NSCLC, non-small cell lung cancer; TIL, tumour-infiltrating lymphocyte; N-TIL, non-tumour-infiltrating lymphocyte (A). RNA-seq analysis showing row-wise z-scores of normalised read counts for each differentially expressed gene (rows) obtained by pairwise comparison of lung $CD8^+$N-TILS(n=32) versus NSCLC $CD8^+$TILS (n=36) (DESeq2 analysis, Benjamini-Hochberg adjusted $P < 0.05$ and 1.5-fold change) for various cancer subtypes, highlighted at the top: NSCLC adenocarcinoma (red) and squamous carcinoma (pink), HPV-negative (light blue) and HPV-positive (dark blue) HNSCC; n=41 for HNSCC $CD8^+$TILS. RNA-seq data from each independent sample is shown in columns (B). Principal component analysis of each $CD8^+$ T cell core transcriptome (symbol) from N-TILS and TILS of indicated cancer subtypes. Percentage of variance in each principal component (PC) is shown in parentheses next to the PC (C). RNA-seq analysis of the expression of exhaustion-associated genes, represented as University of California Santa Cruz (UCSC) genome browser tracks (RPKM, reads per kilo base per million mapped) (top) and dot plots (log2 normalised counts; error bars are mean ± SEM) (below); each dot represents data from an independent experiment (D). *HAVCR2* encodes TIM3.

Data taken from recent manuscript, Ganesan, et al., 2017.

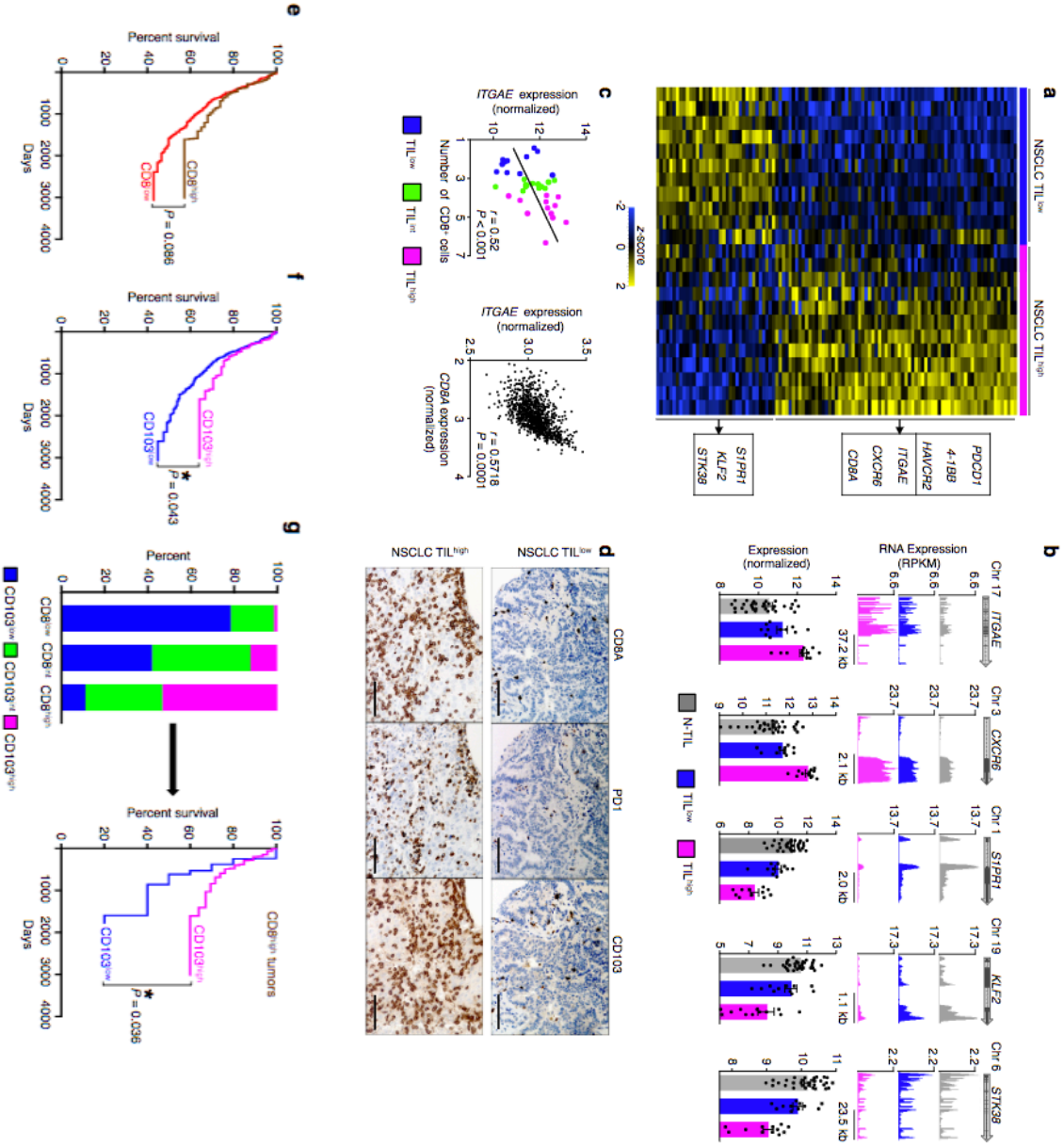


Figure 19. Tissue residency features in TIL^{high} tumours predict for survival.

RNA-seq analysis of NSCLC CD8^{+} TILS showing row-wise z-scores of normalised read counts for each differentially expressed genes (rows) obtained by comparison of TILlow versus TILhigh tumours (DESeq2 analysis, Benjamini-Hochberg adjusted P < 0.05). RNA-seq data from each independent sample is shown in columns (A). RNA-seq analysis of the expression of the indicated transcripts, represented as University of California Santa Cruz (UCSC) genome browser tracks (RPKM, reads per Kilobase per million mapped) (top) and bar graphs (log2 normalised counts; error bars are mean ± SEM) (below); each dot represents data from an independent sample (B).

Correlation of ITGAE (CD103) transcript levels (log2 normalised counts) in NSCLC $CD8^+$ TILS with the number of tumour-infiltrating cells quantified by immunohistochemistry (left); correlation of ITGAE (CD103) transcript levels with CD8A transcript levels in the TCGA lung cancer RNA-seq data set(n=1013) (right). The r value indicates the Spearman correlation coefficient (C). Immunohistochemistry staining for CD8a (left), PD1 (centre) and CD103 (right) in TILlow (top panel) and TIL^{high} NSCLC tumours (bottom panel) (D). Importantly when investigated in a cohort of 689 cases of Lung cancer the presence of CD103 levels was significant benefit to survival (E-G).

Data taken from our recent manuscript Ganesan et al., 2017.

4.2 Expression of differentiation and homing receptors in CD103$^+$ cells.

The first question was to confirm if these populations truly represented tissue resident cells. We therefore wished to rule out that the CD103$^+$ cells expressed markers of other known T cell memory or effector subtypes. For instance, we hypothesised there would be minimal expression of CCR7 or CD62L in the CD103$^+$ gate, but there may be a small level in the CD103$^-$ gate which could represent naïve or CM T cells that are captured during the tissue acquisition (Mueller and Mackay, 2015; Schenkel and Masopust, 2014). In keeping with our hypothesis, there was a significantly different percentage expressing CD62L between CD103$^+$ cells or CD103$^-$ cells which was not found for CCR7. However, when the Median Fluorescence Intensity (MFI) was quantified for both CCR7 and CD62L, there was a notable, statistically significant decrease in expression of both markers in the CD103$^+$ gate compared to the CD103$^-$ CD8$^+$ gate. This suggested to us that the majority of TILS were not central or naïve cells (as expected, given these cells should have limited trafficking to tissue sites), with a small level of CCR7 and CD62L positive cells in the CD103- gate. This lack of expression was not unexpected and ruled out the presence of central memory or naive T cells in the CD103$^+$ gate (Mueller and Mackay, 2015), but this alone did not help validate if these were effector CD8$^+$ populations or a T_{RM} population.

KLRG1 is an established marker of short-lived effector cells, which is down regulated on longer lived populations. It has previously been shown in the seminal work of Laura Mackay, that in murine settings, T_{RM} cells do not express KLRG1 (Mackay et al., 2013; Mueller and Mackay, 2015). Therefore, a further 7 primary lung tumours were analysed and the levels of KLRG1 compared between the CD8$^+$ CD103$^+$ cells or CD103$^-$ TILS (Figure 21). There were notably limited KLRG1$^+$ CD103$^+$ cells compared to the CD103$^-$ TILS, of which the population were on average greater than 50% KLRG1$^+$. When the percentage and MFI were quantified, the difference between the CD103$^+$ cells or CD103$^-$ TILS was significant (Figure 21). This therefore suggests that the majority of the CD103$^+$ TILS were longer lived memory cells, while the majority of the CD8$^+$ CD103$^-$ TILS were short lived effector cells.

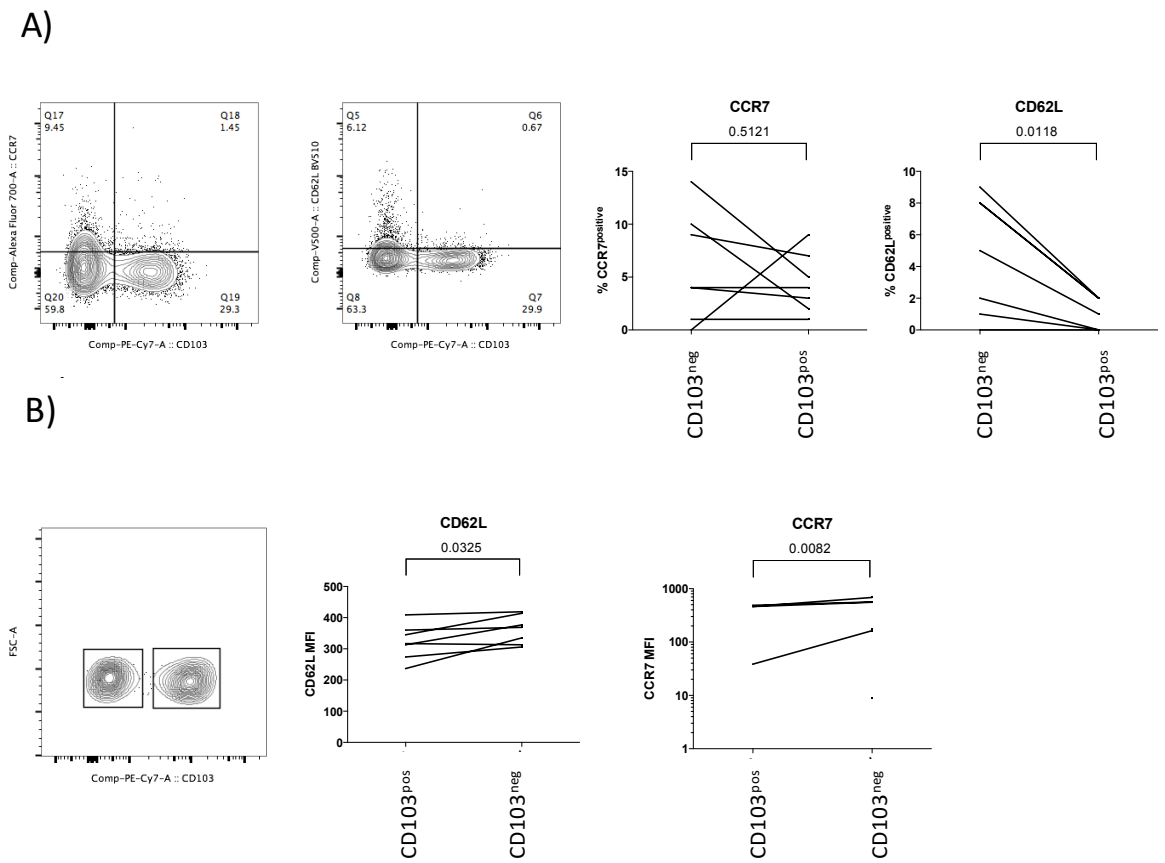


Figure 20. CD103$^+$ TILS do not express CD62L or CCR7.

CCR7 and CD62L are established markers allowing separation of the peripheral memory T cell population into Central and Effector memory CD8 T cells. CCR7 and CD62L are up-regulated on central memory cells, with a down-regulation on effector memory cells, mediating the effector memory cell's migration through the non-lymphoid tissue. Live, singlet, CD3$^+$ & CD8$^+$ TILS were isolated. Expression of CCR7 (Alexa700) and CD62L (BV510) was quantified as a percent of CCR7 or CD62L positive cells in the CD103$^-$ or CD103$^+$ gates (A). Interestingly while the percent of CCR7 cells was not significant, however the average (median intensity of expression of CD62L and CCR7 was significantly different between the CD103$^-$ or CD103$^+$ cells (B). Statistics were completed in Graphpad Prism6 using a Wilcoxon matched-pairs signed rank test.

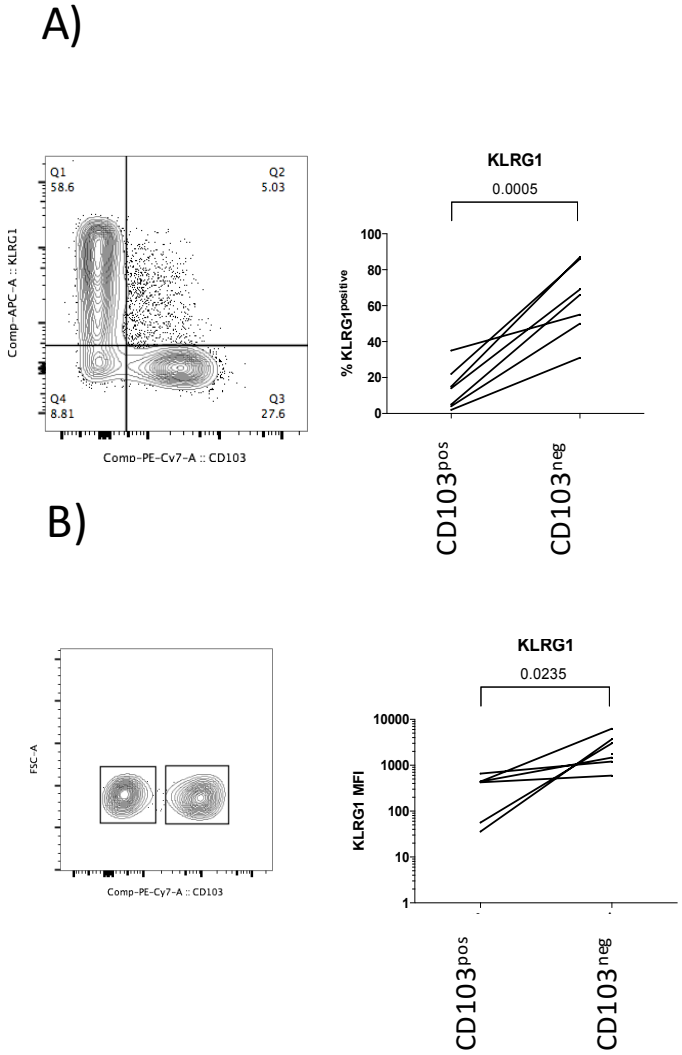


Figure 21. CD103$^+$ CD8$^+$ TILS have limited KLRG1 expression.

Use of expression of KLRG1 is used a representative marker of terminal differentiation on CD8$^+$ T cells and is a receptor for E-cadherin. KLRG1 expression is lost on more long-lived T cells, with expression gained on the more terminally differentiated subtypes. Live, singlet, CD3$^+$ & CD8$^+$ TILS were isolated from 7 primary lung cancer samples. Expression of KLRG1 (APC) was quantified as a percent of KLRG1 positive cells in the CD103$^-$ or CD103$^+$ gates. Furthermore, the average expression as median intensity was significantly different between the CD103$^-$ or CD103$^+$ gates. Statistics were completed in Graphpad Prism6 using a Wilcoxon matched-pairs signed rank test.

4.3 Co-expression of alternative T_{RM} markers, CD49A and CD69 is enriched in $CD103^+$ TILS.

Expression of CD103, as discussed above, is a notable marker of T_{RM} cells in murine settings and has recently been demonstrated in the non-malignant state in analysis of human organ donors (Wong et al., 2016). However, other research has investigated CD69 and CD49A as markers of tissue residency (Mueller and Mackay, 2015; Raaschou-Nielsen et al., 2013; Wong et al., 2016). CD69 is also expressed as a marker of T cell stimulation, while CD49A has only recently been suggested as a unique marker of subtypes of tissue resident cells (Cheuk et al., 2017). Therefore, it could be hypothesised that CD69 and CD49A should be enriched for CD69 and CD49A positive cells, while the $CD103^-$ gate may contain a mixture of both positive and negative populations. Previous literature has used CyTOF to show similar trends in the non-malignant state (Wong et al., 2016). To investigate this, the next step was to analyse co-expression of CD49A and CD69 in 7 cases of primary Lung cancer. It was immediately noted that $CD103^+$ $CD8^+$ TILS rarely failed to express CD69 and CD49A unlike the $CD103^-$ $CD8^+$, when quantified this difference in percentage was significant. Therefore we concluded, as hypothesised, that $CD103^+$ $CD8^+$ TILS were co-expressing CD49A and CD69, which was not observed in all the $CD103^-$ TILS which had mixed pattern of expression for these other markers.

As there were both positive and negative expression levels of CD49A and CD69 in the $CD103^-$ gate, it was unclear if the CD49A/$CD69^+$ cells were $CD103^-$ T_{RM} cells, which have been suggested in some murine settings (Bergsbaken et al., 2017; Khan et al., 2016). We therefore hypothesised that if these were $CD103^-$ T_{RM}, they should not express KLRG1 (Mackay et al., 2013). Therefore, we investigated the $CD69^+$ and $CD49A^+$ cells in the context of the expression of CD103. We hypothesised that the $CD103^-$ CD49A/$CD69^+$ cells would have higher KLRG1 expression suggesting that these are non-T_{RM} populations. For the $CD69^+$ cells, it was noted the expression of KLRG1 was significantly lower in the $CD103^+$ gate (blue) compared to the $CD103^-$ gate (orange). This comparison did not quite reach statistical significance in the CD49A comparison (blue and red). This suggested to us that these $CD49A^+$ or $CD69^+$ cells had elevated KLRG1 expression, which would imply that these are more differentiated cells.

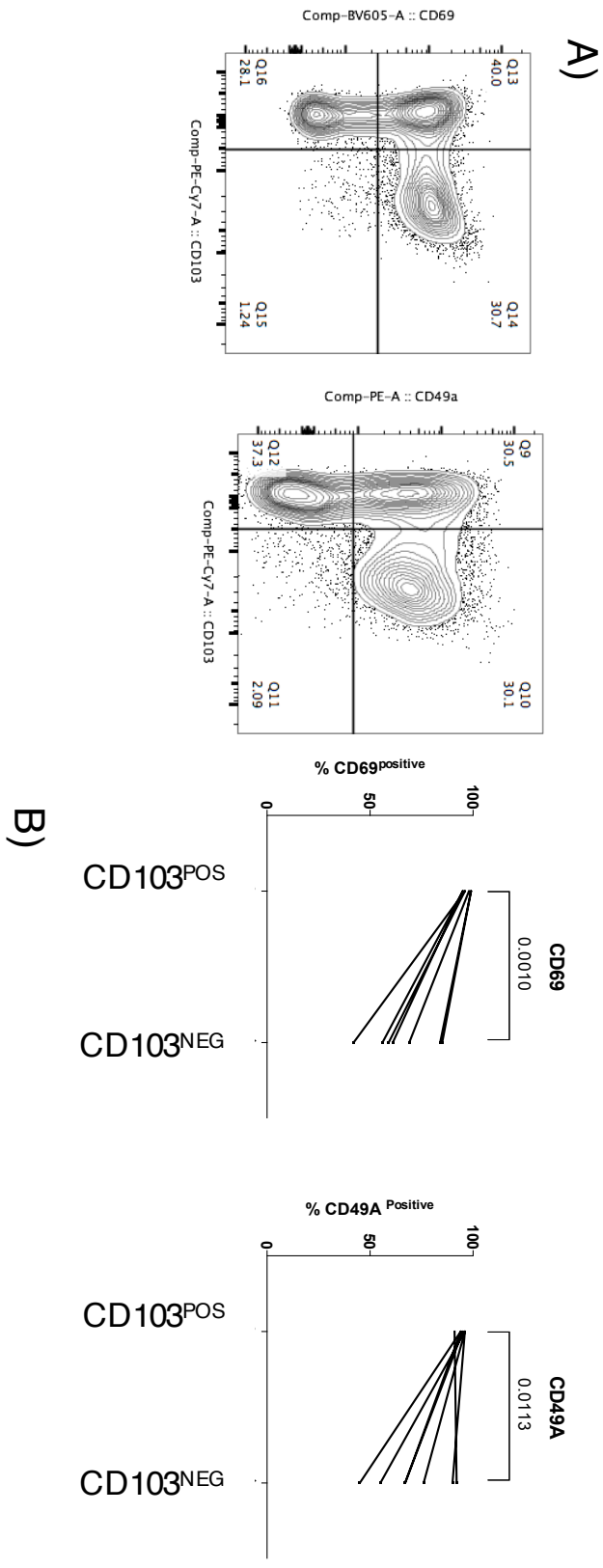
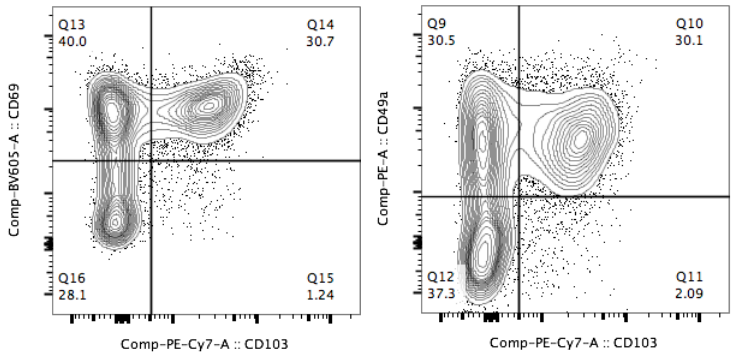

A)
Comp-BV605-A :: CD69
Comp-PE-Cy7-A :: CD103
Q16
28.1
Q13
40.0
Q15
1.24
Q14
30.7
Comp-PE-A :: CD49a
Comp-PE-Cy7-A :: CD103
Q12
37.3
Q9
30.5
Q11
2.09
Q10
30.1
B)
% CD69positive
0
50
100
CD103POS
CD103NEG
0.0010
CD69
% CD49A Positive
0
50
100
CD103POS
CD103NEG
0.0113
CD49A

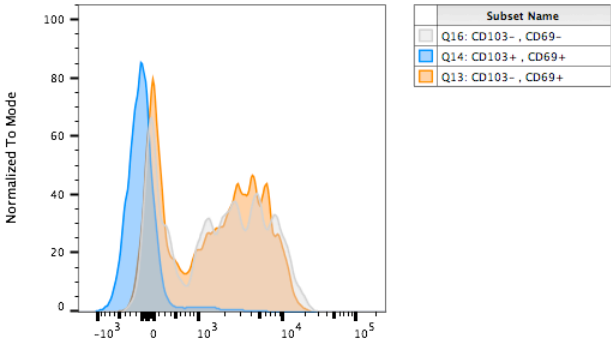
Figure 22. Expression of CD49A and CD69 on $CD8^+$ TILS.

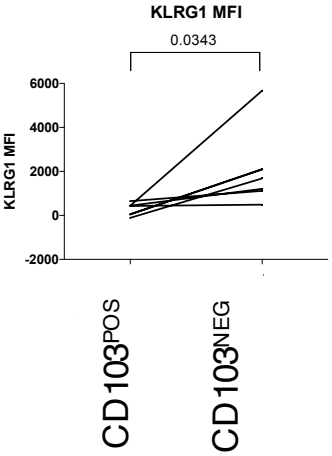
CD69 and CD49A have been reported in other literature to be markers of tissue resident $CD8^+$ cells, however CD69 is also expressed following antigen stimulation and CD49A has previously been demonstrated in a subtype of human skin T_{RM}. Expression of CD69 (BV605) and CD49A (PE) was quantified in the $CD103^-$ or $CD103^+$ gates with significantly higher percentages of the $CD103^+$ cells expressing these markers (A). This expression intensity was significantly higher in the $CD103^+$ gate than the $CD103^-$ using median intensity (B). Statistics were completed in Graphpad Prism6 using a Wilcoxon matched-pairs signed rank test.

A)

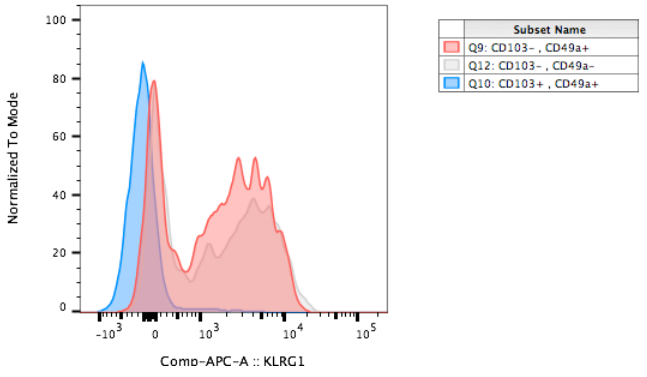


B) CD69





C) CD49A



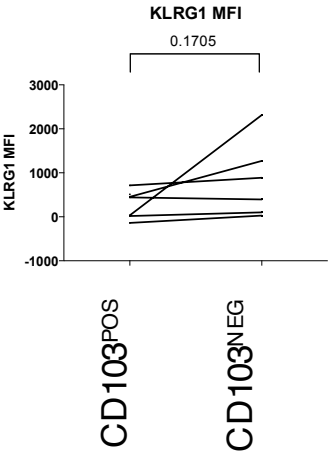


Figure 23. Expression of KLRG1 in CD49A/CD69$^+$ CD103$^-$ cells.

Given the importance of CD69 and CD49A in T_{RM} cells, we hypothesised that the CD103$^+$ cells would have significantly lower expression of KLRG1, while the CD103$^-$ cells would have increased KLRG1 expression as these represented the non-tissue residential cells. CD8$^+$ TILS were isolated from 7 patients and the CD49A and CD69 expressing CD103$^-$ or CD103$^+$ cells (A) were quantified for expression of KLRG1. The median intensity of KLRG1 was significantly lower on the CD103$^+$ CD69$^+$ cells than the CD103-CD69$^+$ cells (B). This trend was also found in the CD49A$^+$ CD103$^-$ or CD49A$^+$ CD103$^+$ cells although it did not quite reach statistical significance (C). Statistics were completed in Graphpad Prism6 using a Wilcoxon matched-pairs signed rank test owing to the paired, non-parametric nature of the dataset.

4.4 Expression of CD45-RA isoform in Lung CD8$^+$ TILS is limited to the CD103$^-$ population.

T cells emerging from the bone marrow express the isoform CD45RA, which, following stimulation by antigen, converts to the isoform CD45RO. However, there are reports of long-term memory cells expressing CD45RA which are found in pathological situations such as chronic viral infections (D'Asaro et al., 2006). We were therefore interested to see if these T_{RM} cells expressed CD45RA, or if there were low levels of CD45RA in the CD103$^-$ gate which may represent peripheral blood naive cells (Mueller and Mackay, 2015) that are acquired during the surgical procedure, which we hypothesised may be a reason for the small populations of CCR7 and CD62L positive cells which we had observed previously. It would have been preferable to investigate co-expression of CD45RA, CCR7 and CD62L to truly quantify if these were peripheral naïve cells that happened to have been traveling through the blood stream. However, the small amount of material available and limited nature of human primary tumours, meant that this analysis could not be completed to date. CD8$^+$ TILS were isolated from 5 patients. Representative gating for CD4$^+$, CD8$^+$, CD3$^+$-Double negative (DN) and CD3$^-$ cells is shown (A). This was to allow us to compare not just in the CD8$^+$ population, but also in the others as a control, although it is worth noting that the majority (80%+ of the CD103+ TILs) fall in the CD8+ gate. In the CD8$^+$ and DN populations we found limited numbers of CD103$^+$ CD45RA$^+$ cells (B). In the CD3$^-$ populations we found a large majority expressed CD45RA, but the CD103$^{+/high}$ cells did have a notable statistically significant decrease (D) which was not seen in the CD4 cells. The lack of NK or B cell markers limits any large interpretation into the CD3$^-$ population but is an interesting point for future work.

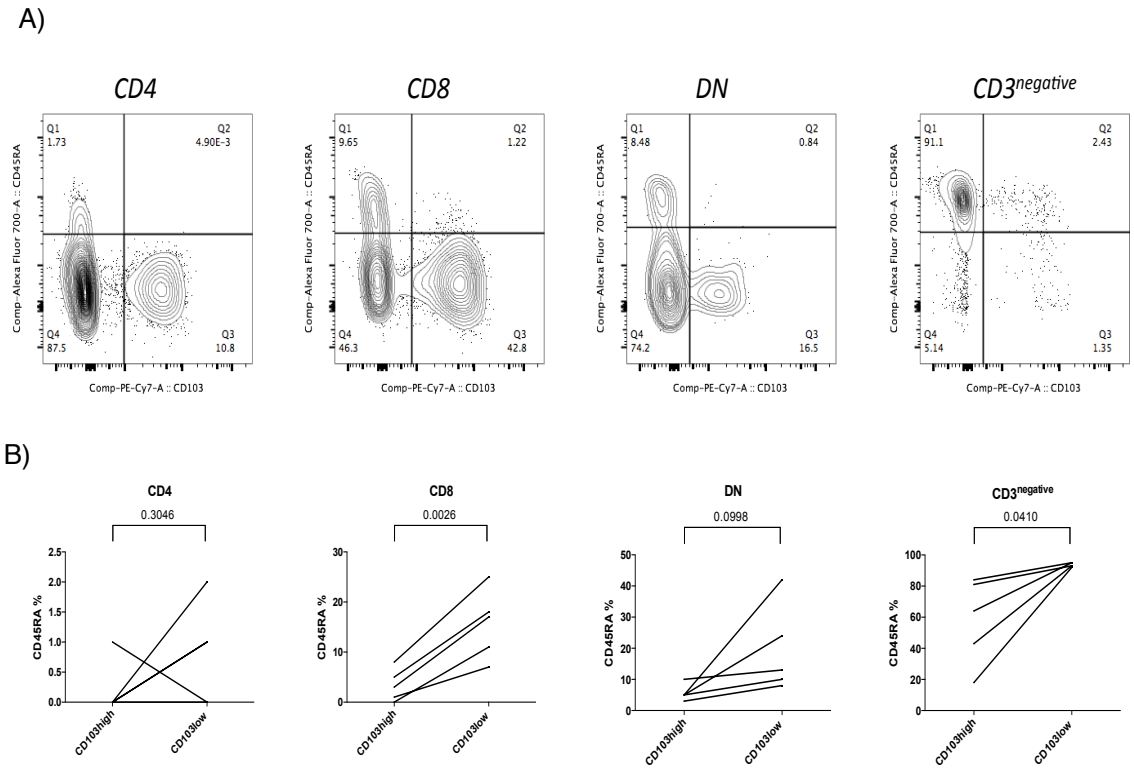


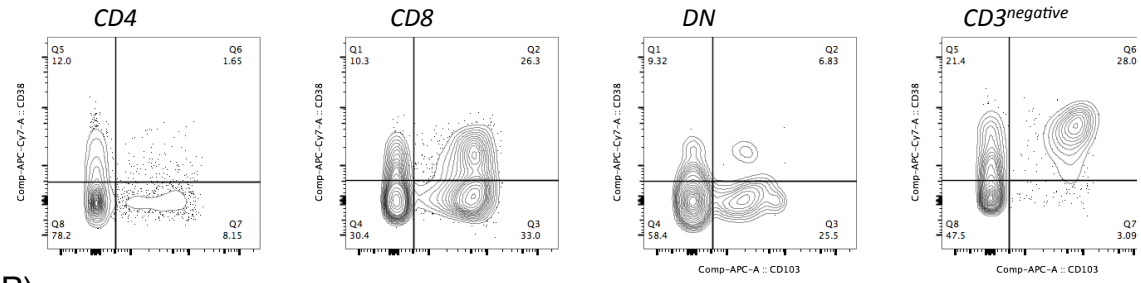
Figure 24. Expression of CD45RA on CD103^{+} compared to CD103^{-} TILS.

Expression of CD45RA is limited to either naïve T cells or the lesser characterised TEMRA cells. Owing to this, we hypothesised that we should identify limited expression of CD45RA in the lymphocytes extracted from human tissue and minimal expression in the CD103^{+} cells. Therefore, the protein levels of CD45RA (AlexaFluor700) were quantified in the CD4, CD8, Double Negative and CD3^{-} gates in the CD103^{+} and CD103^{-} cells. We found in the CD103^{+} population there was noticeably little expression of CD45RA (A). This percentage was found to be statistically significant in the CD8 and CD3^{-} cells, while not quite reaching statistical significance in the double negative population (B). The CD4 population appeared more inconsistent being not significantly different. Statistics were completed in Graphpad Prism6 using a Wilcoxon matched-pairs signed rank test.

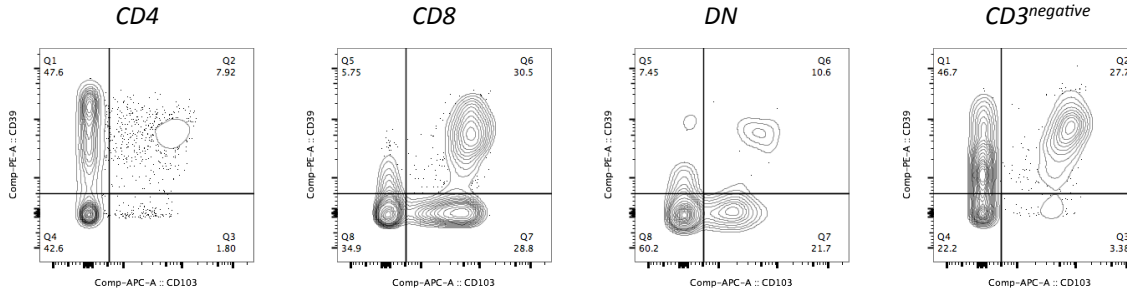
4.5 Expression of CD38 and CD39 is enriched on $CD8^+CD103^+$ compared to $CD8^+CD103^-$ cells.

CD39 is well described in CD4 regulatory cells (Gupta et al., 2015; Sim et al., 2014) and CD38 is established to be expressed on B cells (Chevrier et al., 2017; Liu et al., 2005). CD39 is generally thought to be an inhibitory molecule for T-cell activity while CD38 can be used as a surrogate for activation. We were therefore intrigued to investigate these surface markers in the context of tissue residency in 10 primary lung cancer cases, as our RNA-seq analysis had suggested a positive correlation in $CD8^+$ TILS between CD103 gene expression and both CD39 and CD38, suggesting these may encapsulate human tumour T_{RM} specific features. We therefore hypothesised that $CD103^+$ cells could have more cells expressing CD38 and CD39 than the $CD103^-$ populations. As before we compared these to the $CD4^+$, $CD8^+$, $CD3^+$-Double negative (DN) and $CD3^-$ cells to provide internal controls. Especially as while CD103 is a more established marker in the context of CD8 lymphocytes, it is currently unknown how relevant this marker is for other populations, so although the primary focus is the CD8 compartment it is worth investigating other populations. CD38 and CD39 were notably increased in the $CD103^+$ compared to $CD103^-$ $CD8^+$ TILS as predicted from our RNA-seq analysis; interestingly this was also apparent in the $CD3^+$ DN cells (A/B). The $CD4^+$ cells contained a large proportion of $CD39^+$ cells as would be expected with the large Treg population found in many primary tumours (Chevrier et al., 2017). When these differences were analysed statistically, comparing $CD103^+$ to $CD103^-$ cells, there was a statistically significant difference in the CD8 and DN cells for both CD38 and CD39 (C/D), but the differences were not significant in the CD4 and $CD3^-$ populations. Again, in the absence of a NK or B cell markers, interrogating the $CD3^-$ remains limited.

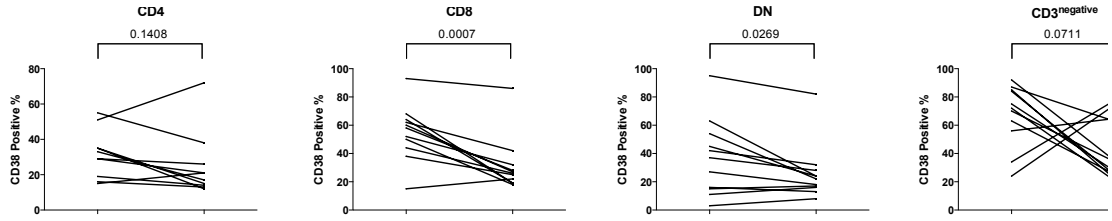
A)



B)



C) CD38%



D) CD39%

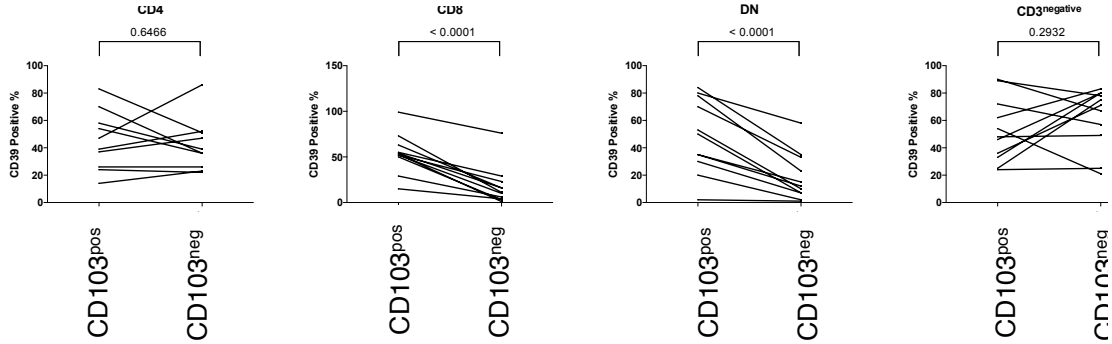
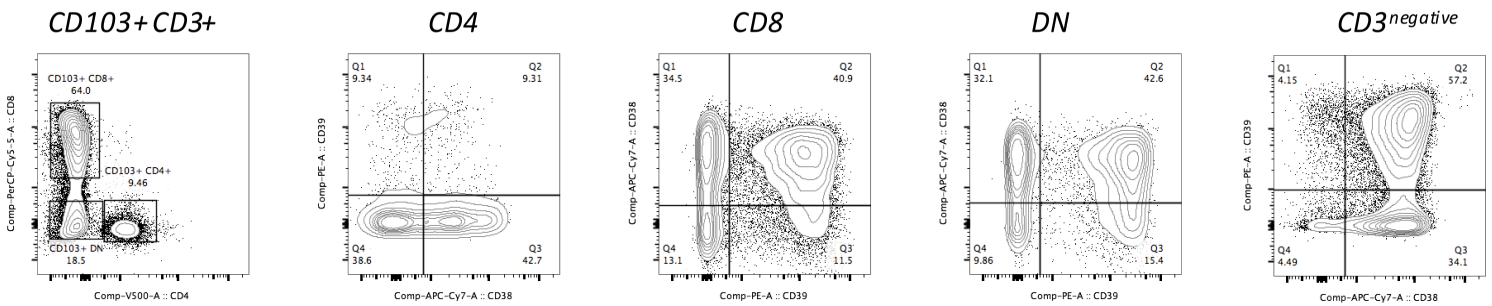


Figure 25. Identification of CD38 and CD39 on $CD8^+CD103^+$ TILS.

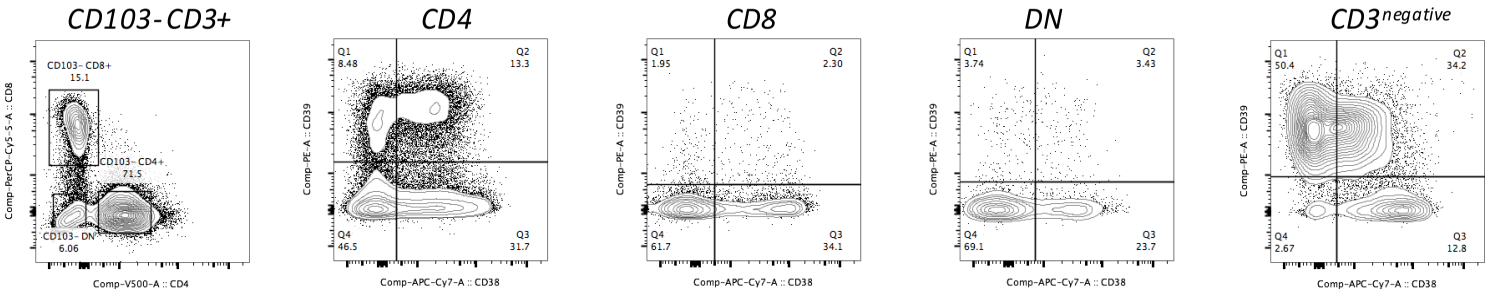
Expression of CD38 and CD39 has not previously been discussed in the context of T_{RM} cells but was suggested from the RNA-seq dataset. Therefore, percentages of CD38 and CD39 was analysed in CD4, CD8, Double Negative and $CD3^-$ cells from 10 primary Lung cancers. This analysis noted high levels of CD38 in the $CD103^+$ compared to $CD103^-$ cells in the CD8, Double negative and a subset of the $CD3^-$ patient samples but not the $CD4^+$ TILS (A). We noted a similar trend in CD39 between the $CD103^+$ and $CD103^-$ cells (B). When quantified we found significantly higher percentages in the $CD103^+$ gate in the CD8 and Double negative cells (C), but not the CD4 or $CD3^-$, albeit with a higher level of heterogeneity. We found a similar trend in the CD39 percentage analysis (D) with an enrichment for positive CD38 and CD39 expression in the CD8 and Double negative $CD103^+$ cells. Statistics were completed in Graphpad Prism6 using a Wilcoxon matched-pairs signed rank test.

Given the similar expression patterns for CD38 and CD39 in the $CD8^+CD103^+$ population it could be hypothesised that there may be a level of co-expression. To investigate this, for the samples analysed in the previous figure the $CD103^+$ and $CD103^-$ TILS were separated (A/B) and the $CD3^+$ populations were gated. Again, we compared this to the other lymphocyte populations to understand this phenotype in the context of general TIL biology. From this it was noted that, as expected, the $CD3^+$ $CD103^+$ cells were dominated by the $CD8^+$ population (A/B, panel 1). However, there was some degree of co-expression of CD38 and CD39 in the CD8 and DN $CD103^+$ cells compared to the $CD103^-$ counterpart. In contrast, the $CD4^+$ $CD103^+$ population had limited $CD39^+$ and $CD38^+$ cells in comparison to the $CD103^-$ $CD4^+$ TILS (A, panel 2). These two plots were overlaid to allow for direct comparison with the $CD103^-$ cells in grey (C). This demonstrated that the expression of CD38 *and* CD39 was enriched for by the $CD103^+$ cells in the $CD8^+$ and DN gate, but this shift between $CD103^+$ cells compared to the $CD103^-$ was not found for the $CD4^+$ and $CD3^-$ gate. This of course has the confounding factor that the $CD3^-$ gate contains NK and B cell populations, but is included for consistency. This suggested to us that these DN cells may have some similar properties to the $CD8^+$ compartment.

These co-expression percentages were quantified (Figure 27), for the CD4 (A), CD8 (B), DN (C) and $CD3^-$ (D) populations. These percentages were as a percent of the parent population to allow for direct comparisons of co-expression between each TIL $CD103^+$ and $CD103^-$ subpopulations. This quantified that in the DN and $CD8^+$ cells the $CD103^+$ gate was enriched for CD38 and CD39 co-expression suggesting that this phenotype may also be identified in this rare and un studied population.



B) CD103neg



C) Overlaid

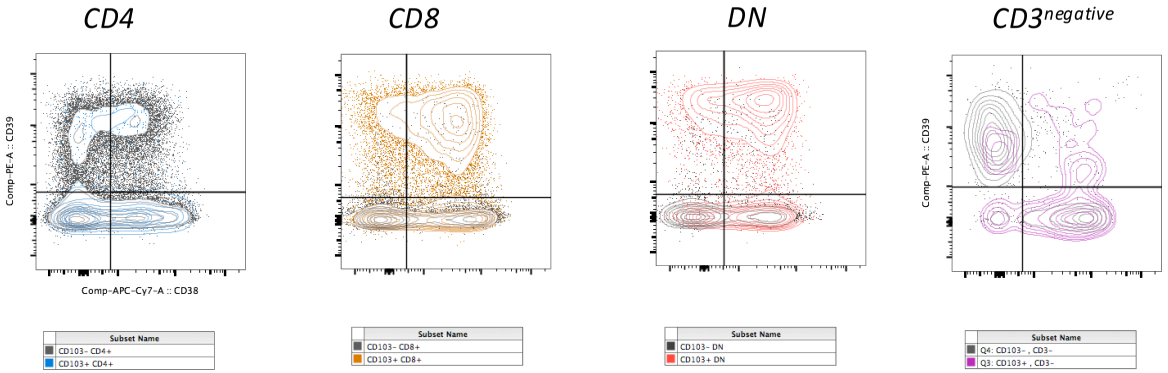
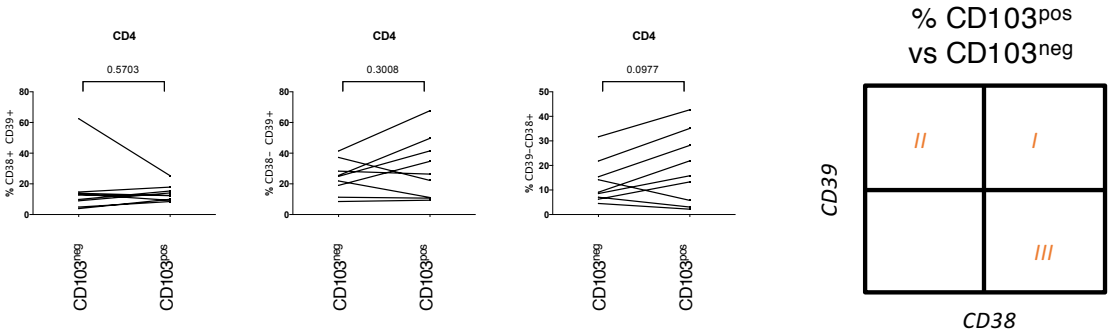


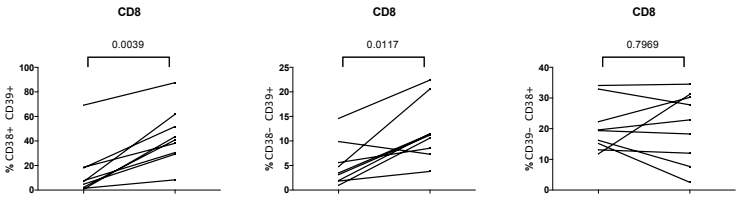
Figure 26. Visualisation of the CD38 and CD39 co-expression.

From the previous figure the analysis had noted a similar trend in expression between CD38 and CD39. We therefore hypothesised there may be a level of co-expression between these markers. Therefore, percentages of CD38 and CD39 was analysed in the $CD103^+$ and $CD103^-$ cells from CD4, CD8, Double Negative and $CD3^-$ gates from 10 primary Lung cancers in the $CD103^+$ population we found a number of CD38 and CD39 co-expressing $CD8^+$, $CD3^+$ and Double negative cells, but not in the CD4 gate (A). In contrast, in the $CD103^-$ population we found a number of CD38 and CD39 co-expressing CD4 cells with few CD8 and double negative cells (B). The $CD103^-$ $CD3^-$ rarely co-expressed CD38 and CD39 (D). These plots were overlaid demonstrating the divergence between $CD103^+$ and $CD103^-$ cell populations, particularly in the $CD8^+$ and Double negative cells (C). These percentages are quantified in the next figure.

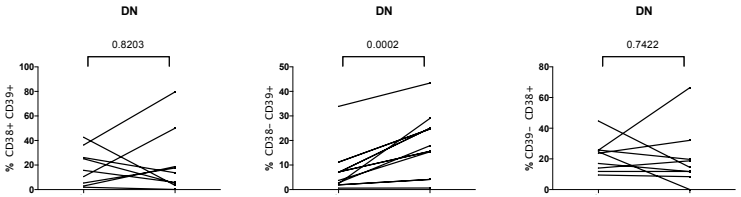
A) Live, Singlet, Lymphocyte, CD45+, CD3+, CD4+



B) Live, Singlet, Lymphocyte, CD45+, CD3+, CD8+



C) Live, Singlet, Lymphocyte, CD45+, CD3+, DN



D) Live, Singlet, Lymphocyte, CD45+, CD3-

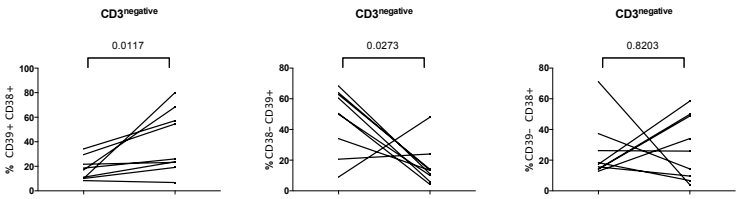


Figure 27. Quantification of co-expression of CD38 and CD39.

The next stage was to quantify the co-expression discussed in the previous figure, using the gates demonstrated in the previous figure, the percentage of CD38 and CD39 negative cells was quantified in the $CD4^+$ gate (A), $CD8^+$ (B), DN (C), $CD3^-$ (D) when comparing the $CD103^+$ and $CD103^-$ cells. The first column represents CD39/CD38 double positive cells, the second $CD39^+$ $CD38^-$ cells, and the third $CD39^-$ $CD38^+$ cells as denoted in the schematic in the top right.

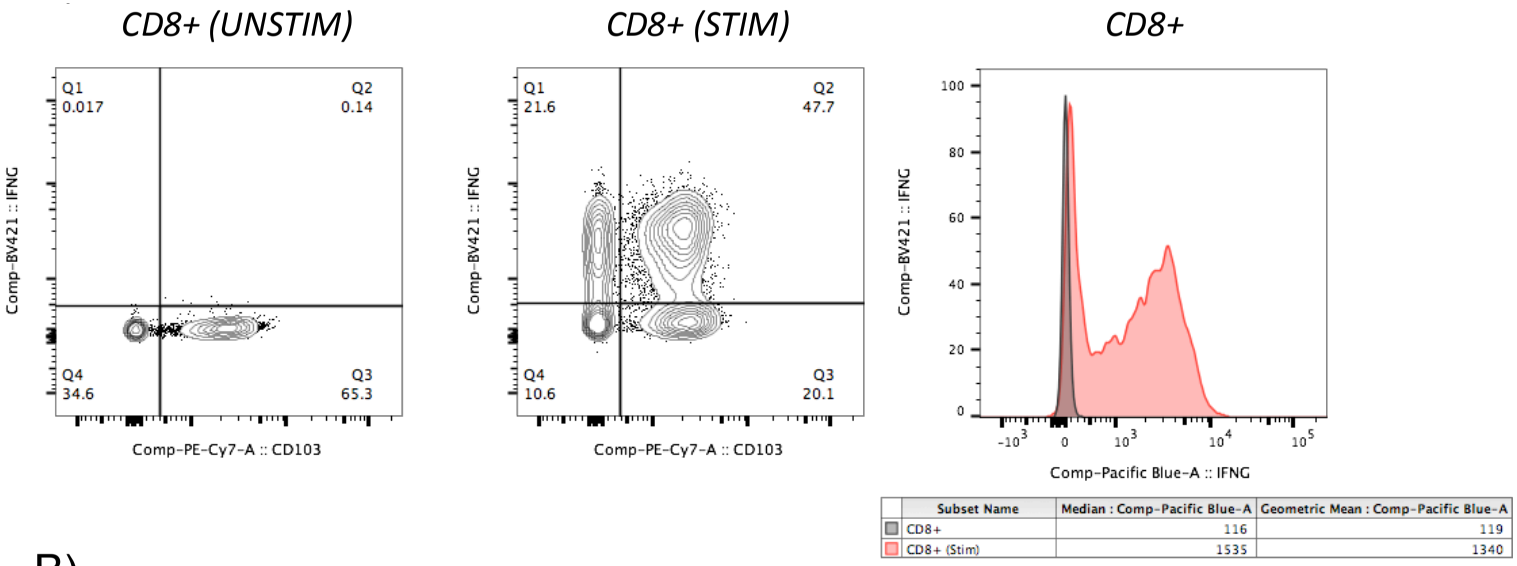
4.6 Expression of proteins linked to effector function are not enriched in CD103$^+$ TILS.

By comparing the transcriptomes of isolated CD8$^+$TILS from TIL^{high} compared to TIL^{low} lung tumours, we showed in TIL^{high} Tumours increased expression of transcripts linked to effector function (Interferon γ, *Granzyme A* and *Granzyme B*). There are several possibilities to explain the survival benefit for patients of CD103^{high} tumours. A hypothesis is that tumours with CD103$^+$-rich infiltrates will have an increased infiltrate of CD8$^+$ TILS thus improving survival. However, even in patients with identical numbers of CD8$^+$ TILS, we identified a large statistical difference in survival based on CD103 status (Figure 19).

Therefore, we had two potential possibilities, which were not mutually exclusive. The first is that CD103$^+$ CD8$^+$ TILS exhibit greater effector potential than CD103$^-$CD8$^+$TILS; thus, the TIL^{hi} Tumours may have T cells with greater effector potential than TIL^{lo} Tumours by virtue of having a greater proportion of CD103$^+$CD8$^+$ TILS. An alternative was that all CD8$^+$ cells from TIL^{high} tumours may have greater effector potential owing to features of the CD103+ cells. To understand these changes, we wished to compare the levels of these cytotoxic molecules between the CD103$^+$ and CD103$^-$TILS.

In order to investigate differences in effector functions of T cells, we applied standardised PMA/ionomycin stimulation which provides non-antigen-specific stimulation to all the TILS extracted from the tumour as we were unable to detect significant staining directly ex-vivo. We noted a significant increase in intracellular expression of Interferon γ (A), expressing the data as a percentage and as a histogram (red). Interestingly, when we compared the CD103$^+$ and CD103$^-$ TILS, we identified no significant difference in the percentage of Interferon γ positive cells (B). This therefore suggested that perhaps the increased Interferon γ expression in the tumours high in TILs was not linked to the expression on T_{RM} cells compared to the non-T_{RM} population but rather to a global change. We next completed this for CD107A which was not a differentially expressed gene but is a marker of cells undergoing degranulation (Figure 29A/B). Again, we identified limited expression directly ex-vivo (A), however following stimulation, this quantity increased greatly. This population was difficult to gate during the flow cytometry analysis, therefore the expression in populations that were negative for CD107A expression were used (example shown). Again, we identified no significant difference between the two CD8 populations (B) in the context of strong stimulation with PMA/Ionomycin.

!



Subset Name	Median : Comp-Pacific Blue-A	Geometric Mean : Comp-Pacific Blue-A
CD8+	116	119
CD8+ (Stim)	1535	1340

B)

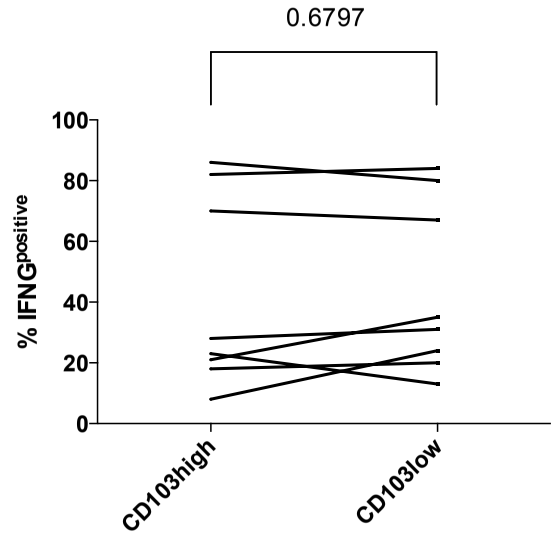
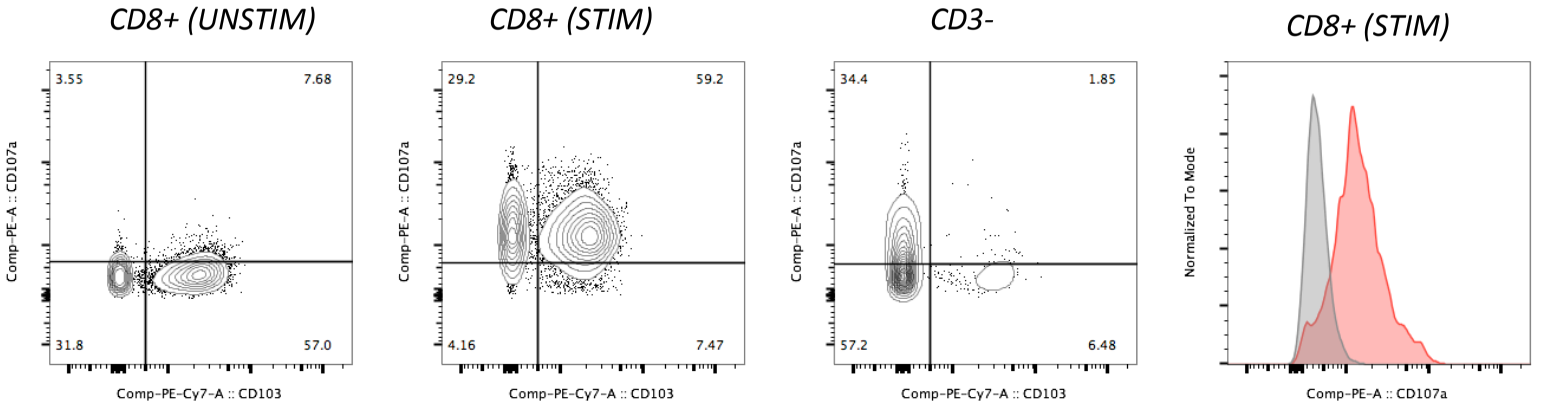


Figure 28. Expression of IFNγ in $CD8^+$ TILS.

From our manuscript (Ganesan, et al 2017., Nature Immunology), we hypothesised a number of molecules were downstream of IFNγ (BV421), which suggested there should be expression of IFNγ in the CD8+ TILS. Upon isolation of $CD8^+$ TILS from primary lung cancer (n=3; A, B), we failed to note significant IFNγ expression directly ex-vivo. However, upon the classical PMA & Ionomycin stimulation (n = 8), we were able to detect significant expression (representative A). Interestingly there was no noticeable variation between the $CD103^{+/high}$ and $CD103^{-/low}$ TILS (B). There was a large degree of variability between the patient samples which suggested there to be further factors other than expression of CD103, that lead to IFNγ expression. Statistics were completed in Graphpad Prism6 using a Wilcoxon matched-pairs signed rank test.

A)



Subset Name	Median : Comp-PE-A	Geometric Mean : Comp-PE-A
CD8+	360	383
CD8+ (Stim)	1340	1394

B)

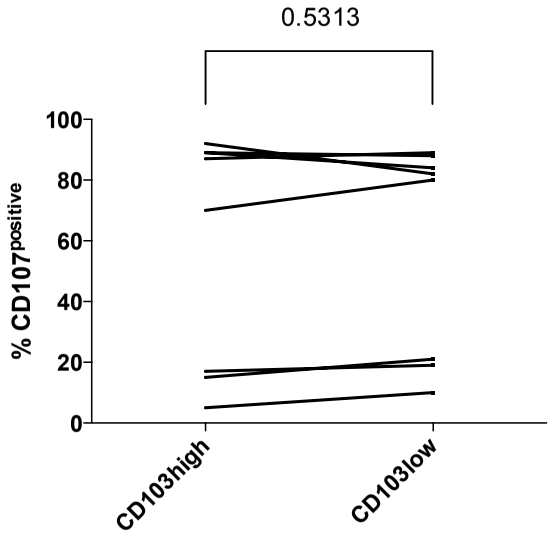


Figure 29. Expression of CD107A in $CD8^+$ TILS following stimulation.

We hypothesised that there may be a difference in CD107A expression between the $CD103^+$ and $CD103^-$ cells and/or there may be a difference between TIL^{high} or TIL^{low} tumours. Upon isolation of $CD8^+$ TILS from primary lung cancer (n=3; A, B), we failed to note significant reliable CD107A (PE) expression. However, upon the classical PMA & Ionomycin stimulation (n = 7) we were able to detect significant expression. Gating was completed on a the $CD3^-$ population as these should have populations that do not express CD107A. Interestingly there was no noticeable variation between the $CD103^{+/high}$ and $CD103^{-/low}$ TILS (D). Statistics were completed in Graphpad Prism6 using a Wilcoxon matched-pairs signed rank test.

4.7 Expression of cytotoxicity proteins in $CD8^+$ TILS are not enriched in the $CD103^+$ $CD8^+$ TILS.

The next stage was to assess the expression of cytotoxic molecules Granzyme A, Granzyme B and Perforin which was analysed in 9 primary tumours; it is worth noting that Perforin was not one of the differentially expressed genes between $CD8^+$ cells isolated from a TIL^{high} and TIL^{low} tumour but Granzyme A and B were. We expected that cells that expressed Granzyme B^+ should have higher expression of Granzyme A and Perforin and confirmation of this would validate the correctness of the FACS antibody panel. We confirmed that Granzyme B^+ cells indeed showed greater expression of Granzyme A and Perforin, which was then completed for every sample (Figure 30; representative A) to confirm the intracellular staining had been completed correctly. In both $CD103^+$ and $CD103^-$ gates, it was noted there were populations of Granzyme A, Granzyme B and Perforin positive cells (B). When completing paired analysis, there appeared to be no significant difference for any of the three markers between the $CD8^+$ TIL subtype when discriminated on CD103 expression (C). It could be that the difference at the gene level is not maintained at the protein level, or that the $CD103^+$ do not have any greater effector functions than the negative population. Alternatively, it could be that there are subpopulations that are affecting the average signature.

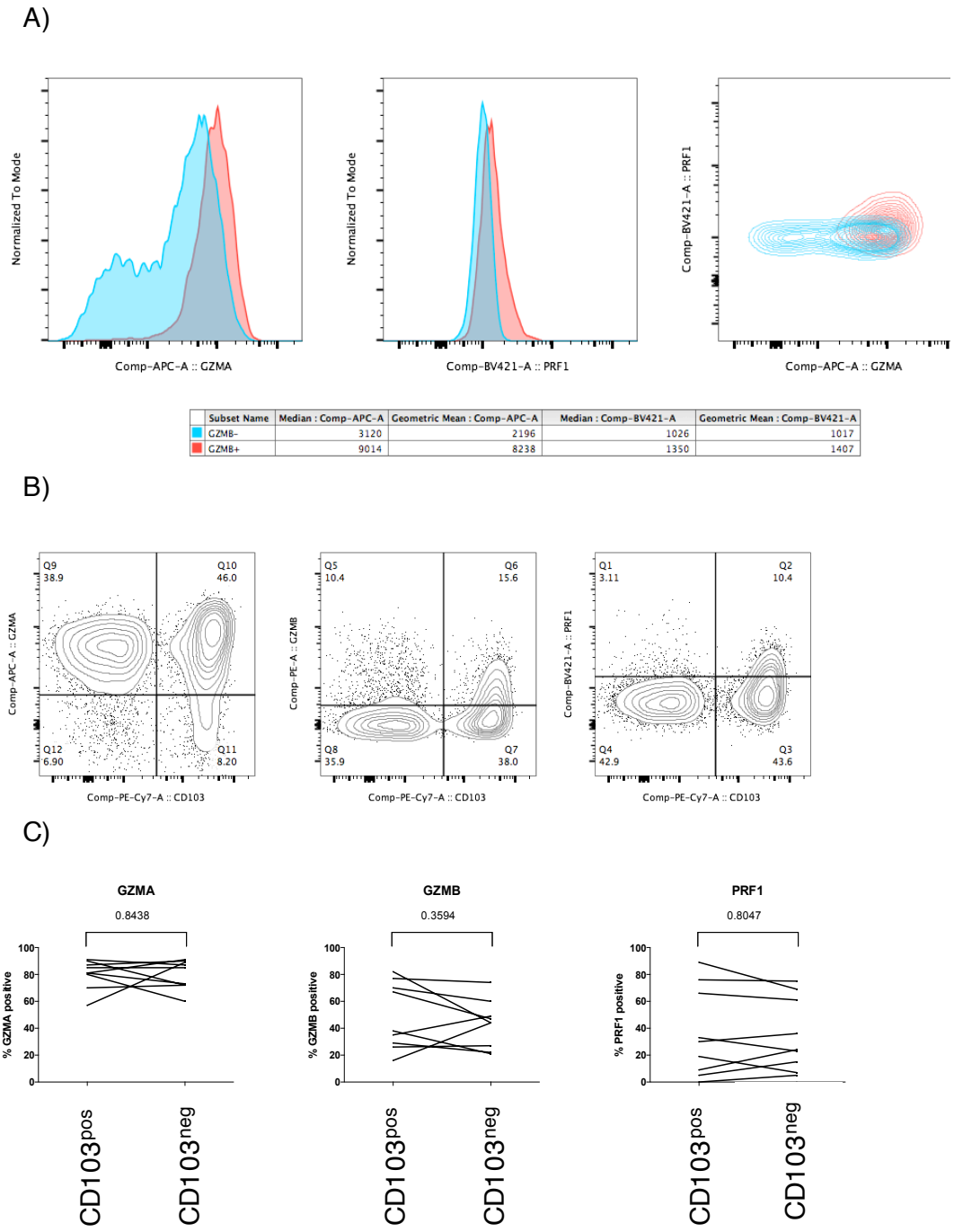


Figure 30. Expression of cytotoxic proteins in CD103^{+} and CD103^{-} TILS.

In our manuscript, Ganesan, et al. 2017, we noted a significant difference in gene expression of Granzyme A and Granzyme B between CD8^{+} cells isolated form TIL^{high} and TIL^{low} tumours. As the expression of CD103 was also differentially expressed, we wished to confirm if this gene expression pattern of GZMA and GZMB was owing to the enrichment of CD103^{+} cells. Consequently, CD8^{+} TILS were isolated from 9 patients and expression of Granzyme A (APC), Granzyme B (PE) and Perforin (BV421) was quantified in the CD103^{+} and CD103^{-} cells. To confirm that the staining had been completed successfully we confirmed that the Granzyme^{+} CD8^{+} TILS

had significantly higher intensity staining for Perforin and Granzyme A, suggesting the expected co-expression (A). The percentage of Granzyme A, Granzyme B and Perforin expressing cells was quantified in the $CD103^+$ and $CD103^-$ gates (B). This percentage was found to not be significantly different between the $CD103^+$ and $CD103^-$ cells (C). Statistics were completed in Graphpad Prism6 using a Wilcoxon matched-pairs signed rank test.

4.8 Chapter 4. Discussion.

In this chapter, we have identified three core findings. The first is that CD103 appears to reflect and discriminate a population of T_{RM} cells in human cancer. Second, we could confirm expression of surface markers predicted from the RNA-seq data and these were readily identified on the $CD8^+$ $CD103^+$ TILS. Additionally, it appeared this protein expression pattern was also found on $CD3^+$ $CD4^-$ $CD8^-$ TILs. Thirdly, we detected no significant differences in effector functions between the global $CD103^+$ and $CD103^-$ $CD8^+$ TILS.

The work in this chapter focused on validating transcriptomic observations at the protein level. These three objectives were to validate 1) that $CD103^+$ cells are candidate T_{RM} cells; 2) the protein expression novel markers in T_{RM} biology; 3) to identify if these cells have increased cytotoxicity. We correlated transcriptomic and protein expression data with clinical or pathological features (TIL status and T_{RM} levels).

We first aimed to validate that CD103+ cells are T_{RM} cells based on expression of appropriate markers. We focused on CD103 as it was a significant hit both at the transcriptomic level and in terms of its prognostic importance. CD103 is the alpha chain of the integrin αEβ7 (human mucosal lymphocyte-1 antigen), which binds the adhesion molecule E-cadherin in epithelial tissue. T_{RM} cells down regulate S1PR1 and KLF2 and we could observe that these genes were part of the gene set down-regulated in the TIL^{high} tumours (Mackay et al., 2016; Mueller and Mackay, 2015). Down-regulation of S1PR1 is necessary for the egress of T cells from the lymph nodes and leads to the down-regulation of the KLF2/KLF3 transcription factor that normally promotes tissue egress. This downregulation is mediated by CD69. If the CD103+ cells are T_{RM} cells, they should have limited expression of CCR7 and CD62L and we were able to confirmed this at the protein level. CCR7 date were analysed after Alexa Fluor700 labelling. The weaker fluorescence of this colour may explain why CCR7 analysis did not quite reach significance at the percentage level; but the data support our interpretation when considering signal intensity (Figure 20). The presence of a small number of $CCR7^+$ and $CD62L^+$ cells likely results from a few central memory or naïve cells being acquired in the tumour resection. Although we routinely wash these tumours and manually dissect any blood

vessels, a small level of contamination is impossible to avoid. Alternatively, these may be central memory cells localised to lymphoid-like structures that have been increasingly recognised in the context of tumour immunology (Finkin et al., 2015). The lack of CD62L and CCR7, although reassuring, was expected. As the effector populations express KLRG1, the expression should be relatively minimal in any $CD103^+$ cells, if they do represent memory populations. This was seen with minimal expression in the $CD103^+$ gate (Figure 21) and was anticipated from the RNA-seq data. This analysis did identify a $CD103^-$ $KLRG1^-$ population (~5-10% of $CD8^+$ TILS), although we were unsure what phenotype these cells are.

We anticipated the $CD103^+$ cells would have increased CD69 and CD49A expression, which are known to be enriched (but not perfect) markers of tissue residency. We gained confidence that these are T_{RM} cells based on the expression of these two markers being (Figure 22). When these markers were analysed using histograms it became apparent there were potentially a number of subpopulations of T_{RM} cells, which has encouraged us to complete further single cell analysis to interrogate the subtypes by using Smartseq2 approaches, which would allow analysis of the TCR clonality between any subpopulations (Macosko et al., 2015; Stubbington et al., 2016).

The second stage was to validate some of the markers identified from the TIL^{high} compared to TIL^{low} comparisons. Two of the top hits were CD39 and CD38 which have been well described on CD4 regulatory and B cells. We identified that the DN and $CD8^+$ cells that were $CD103^+$ had significantly higher expression of CD38 and CD39 (Figure 25). We noted the expected $CD39^+$ cells which likely represent the regulatory cells. Furthermore, the CD38 positive cells in the $CD3^-$ gate are potentially B cells, although NK cells have been demonstrated to express CD38 and CD39. CD39 is a cell-surface enzyme that dephosphorylates ATP to AMP. Interestingly, in a murine viral challenge setting, expression of CD39 has been shown to highlight the most exhausted T cells (Gupta et al., 2015). An experiment is planned but has not been completed yet, in which the expression of CD39 will be compared to PD1, 4-1BB and other exhaustion markers. CD38 is involved in NAD (nicotinamide adenine dinucleotide) metabolism and has been used as a marker of T cell (and B cell) activation.

The third hypothesis was that the $CD103^+$ lymphocytes could have higher expression of effector molecules as we identified a correlation between CD103 and these markers in the whole CD8 TIL transcriptomics. However, for each marker, we failed to find a statistically significant difference in expression. This therefore suggested that regardless of CD103 expression, there may be a difference in the expression of these effector molecules in the TIL^{high} compared to TIL^{low} $CD8^+$ TILS. Consequently, once the dataset presented in this chapter was combined with an additional dataset completed alongside Dr Ganesan, we were able to identify a difference in Granzyme A

expression between TIL^{high} and TIL^{low} tumours (Ganesan, et al 2017,. Figure 5F). Having given us confidence that these cells were T_{RM} cells, we had been able to identify some novel biology from the $CD8^+$ RNA-seq. The next stage, to be covered in the next chapter, was to investigate the transcriptome profile of the T_{RM} cells in the tumour and background lung. This analysis could potentially give better resolution on these T cell expression profiles that were lost in the whole $CD8^+$ transcriptome.

Chapter 5: Results 3: Investigation of CD103$^+$ lymphocytes through population transcriptomic techniques.

5.1 Introduction.

The previous chapter explored the phenotype of CD103 expressing cells within the tumour infiltrating population of CD8+ lymphocytes. First, it was shown that there was no expression of CCR7, CD62L or KLRG1 within the population indicating they are not central memory CD8$^+$ cells or short-lived effector cells. The strong expression of additional residency markers CD69 and CD49A by the CD103+ population led us to conclude that these cells were likely human T_{RM} cells. From the whole CD8$^+$ RNA-seq and flow cytometry data, it could have been hypothesised these cells would have a unique expression profile compared to the cells not expressing CD103. Therefore, the next stage was to apply the approaches of unbiased transcriptome analysis developed in Chapter 3 to the cells identified in Chapter 4. This would allow validation of a multitude of markers and a clean resolution of the differences between CD103$^+$ and CD103$^-$ cells in the tumour and background tissue.

A recently published paper investigating background lung CD103$^+$ and CD103$^-$ cells in a small cohort, had identified and validated the importance of a transcription factor RBPJ (Hombrink et al., 2016). However, the authors failed to identify differential expression of genes encoding known important proteins including, KLRG1, CD49A or HOBIT (*ZNF683*), among others which would have been expected from the known differences between tissue localised and peripheral cells in mice and humans (Mackay et al., 2016; Zheng et al., 2016). The authors had concluded that CD103$^+$ and CD103$^-$ were similar cells, however they had detected differential expression of *S1PR1* which is the main pathway driving the divergence in T_{RM} development. This suggested that perhaps the small sample size had failed to allow detection of differences owing to the heterogeneity in the patients, particularly given the cohort was a mixture of T cells isolated from patient's suffering from different respiratory diseases. An additional paper by Kumar (Kumar et al., 2017), et al had completed a small cohort (n = 3) using CD69$^+$ as a sorting marker. In their analysis, one of the top hits was *ITGAE* (CD103), suggesting this was likely reflective of some of the CD103 features with the addition of any markers of antigen specific engagement. At the same time, a paper by Malik, et al,. (2017) in Science Immunology had presented the first murine mechanistic paper for Tumour Infiltrating T_{RM} cells, demonstrating these cells were required for the long-term anti-tumour immunity to Melanoma (Malik et al., 2017).

Therefore, this chapter contains the data analysis of CD103$^+$ cells isolated from Lung cancer tumours and background lung. This is supplemented with the CD103$^-$ populations as a reference

point. The main aim consequently was to define the genome-wide transcriptomic profile of these cells. This data could then be overlaid with known prognostic information such as the level of immune infiltration to provide an understanding if there are features of these populations that are enriched in particular clinical subgroups.

5.2 Results.

5.2.1 $CD103^+$ and $CD103^-$ CTLS isolated from background lung are transcriptionally distinct and display divergence in known tissue residency features.

Recent studies have highlighted the importance of T_{RM} in several diseases including cancers of multiple origins (Cheuk et al., 2017; Ganesan et al., 2017; Iijima and Iwasaki, 2015; Schenkel and Masopust, 2014). To systematically define the transcriptomic differences between $CD103^+$ and $CD103^-$ $CD8^+$ cells, they were sorted into these populations from normal lung tissue and matched lung cancer respectively and discriminated on CD103 expression. Unbiased visualization of RNA-seq data, using 2D t-stochastic neighbour embedding (tSNE), revealed the distinct nature of these two populations (Figure 31A). Analysis of differential gene expression identified 696 differentially expressed genes (DEGs) between these $CD103^+$ and $CD103^-$ populations (Figure 31B). The DEGs included transcripts previously reported to be enriched in T_{RM} such as *S1PR1, S1PR5, ITGA1, RBPJ* (Cheuk et al., 2017; Piet et al., 2011), as well as inhibitory molecules *PDCD1, CTLA4* and *KLRC1* (Park and Mackay, 2017). The expression of key surface molecules CD49A and KLRG1 was confirmed at the protein level (Figure 32A). Gene set enrichment analysis (GSEA) of transcripts differentially regulated in $CD103^+$ cells showed a significant enrichment for a previously defined T_{RM}-specific gene set (Figure 32B) (Mackay et al., 2016). Using permutation testing, we identified that < 20% of the modules identified through weighted gene co-expression network analysis (WGCNA) (Oldham et al., 2006) in the NL-$CD103^+$ (Non-involved Lung) transcriptome were maintained in the NL-$CD103^-$ population (Figure 33), emphasizing the unique properties of human T_{RM} populations using individual transcript and modular analysis.

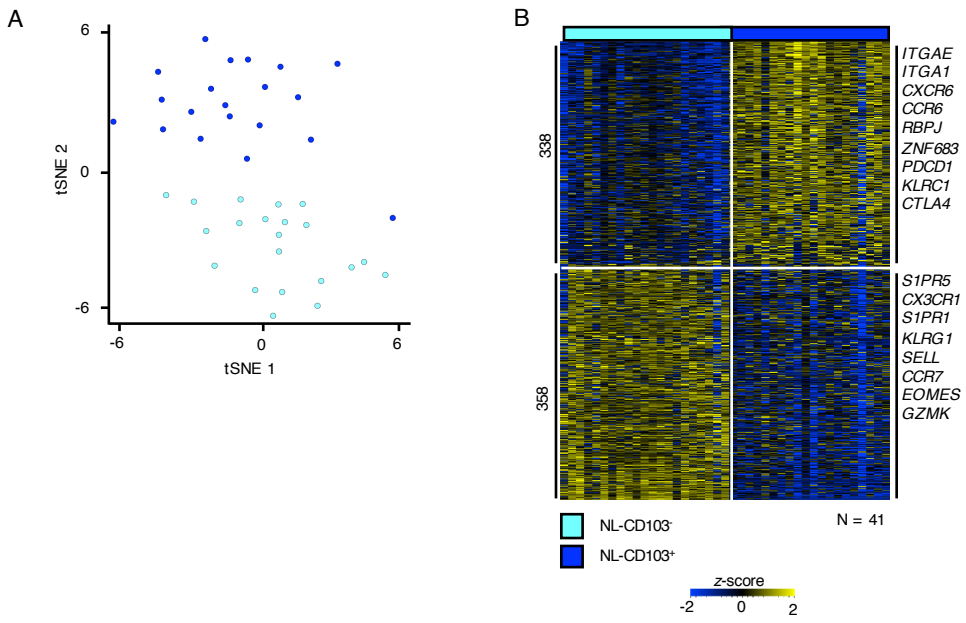


Figure 31. Transcriptomic analysis of background lung CD103$^+$ and CD103$^-$ CTLS using RNA-seq.

Visualization of RNA-seq data, using 2D t-stochastic neighbour embedding (tSNE) using 10 principle components and ~2500 most variable genes (A). RNA-seq analysis showing row-wise z-scores of transcripts per million for each differentially expressed gene (rows) obtained by pairwise comparison of lung CD8$^+$CD103$^+$ versus lung CD8$^+$CD103$^-$ (n= 41) (DESeq2 analysis, Benjamini-Hochberg adjusted P < 0.05 and 2-fold change (B).

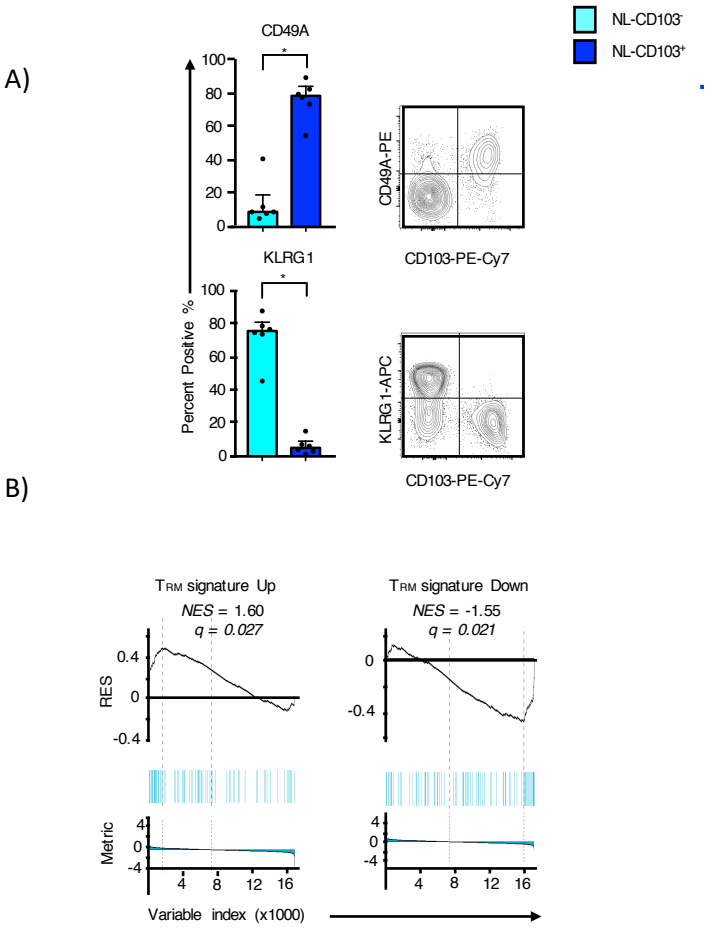


Figure 32. Known residency features are enriched in CD103^{+} cells.

Lymphocytes were quantified as Live, Singlet, CD3^{+}, CD8^{+}, CD19^{-}, CD20^{-}, CD14^{-}, CD4^{-} lymphocytes and expressed as a percentage positive in the CD103^{+} and CD103^{-} gate. Statistics were completed in Graphpad Prism7 using a Wilcoxon matched-pairs signed rank test owing to the paired, non-parametric nature of the dataset. We calculated the significant enrichment for a previously defined signature of Tissue Residency (Mackay et al., 2016) in the CD103^{+} and CD103^{-} NL-CTLS using Gene Set Enrichment Analysis.

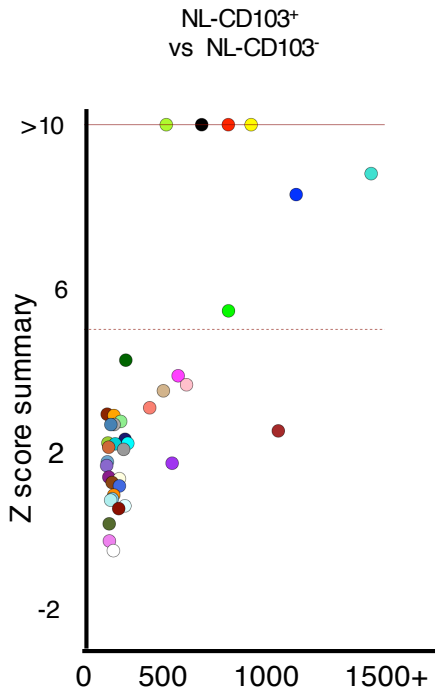


Figure 33. NL-CD103^{+} cells have distinct features at a modular level.

Signed WGCNA analysis using a beta value of 8, created from the NL-CD103^{+} transcriptome, divided the transcriptomic network into distinct modules. Each dot (and colour) represents a singular module. Module preservation was calculated using 30 permutations and a negative control of randomly assigned genes. Preservation was considered for values with a Z score above 5.

5.2.2 CD103$^+$ CTLS are equally responsive to ex-vivo stimulation, despite expression of inhibitory molecules.

Given the upregulation of inhibitory molecules in the NL-CD103$^+$population, we were interested to determine whether these cells were in a hypo-functional state, hence paired RNA-seq before and after stimulation was completed. Consistent with murine studies, we found that following stimulation of these human T cells with PMA/ionomycin, there was strong up-regulation in the expression of transcripts related to effector functions in both NL-CD103+ and CD103- cells with no significant difference between them, despite the expression of these inhibitory molecules (Figure 34; (Park and Mackay, 2017). This suggests that the expression of PD1 likely reflects an intrinsic feature of T_{RM} cells as has recently been described (Shwetank et al., 2017). In CD103$^+$ cells under stimulated conditions, we observed a highly significant increased expression of *IL1RL1* (*IL33R*) and the transcription factor *ZEB1*. IL33 is shown to be required in optimal responses of innate epithelial lymphocytes(Chan et al., 2014) and functions as a 'alarm' of tissue damage (Martin and Martin, 2016). While ZEB1 is a transcription factor that has not been studied in the context of T_{RM}.

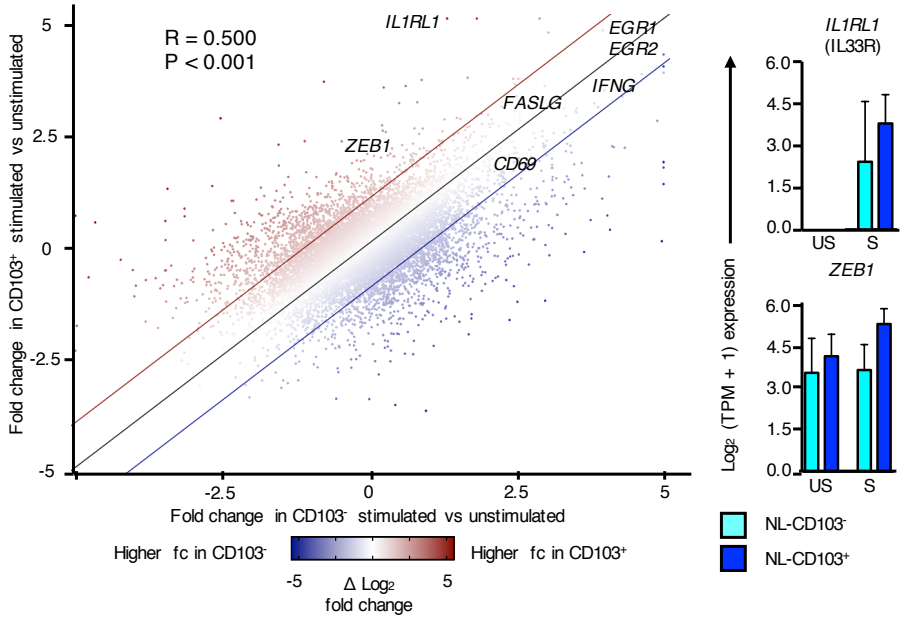


Figure 34. NL-T_{RM} are not hypofunctional despite PD1 expression.

Cells were stimulated with PMA/Ionomycin for 4 hours (S) compared to an unstimulated control (US). Fold Change is DESeq2 variance stabilised, correlation is spearman. Bars are mean +/- Standard Deviation.

5.2.3 The signature of NL CD103$^+$ CTL is distinct from profiles in murine acute lung infection and human skin.

To investigate the extent of conservation of T_{RM}-specific gene expression in different tissues, we compared the differentially expressed genes between CD103$^+$ and CD103$^-$ cells in lung (this study) with those reported for skin (Cheuk et al., 2017). This comparative analysis suggests that the majority of differentially expressed genes were specific to lung (Figure 35). Perhaps surprisingly, we found very little similarity between the T_{RM} of murine and human lung (this study) despite a core shared transcriptomic program that was reflective of the known T_{RM} pathways. Together these results demonstrate the tissue-specific features of CD103$^+$ T_{RM} in human lungs.

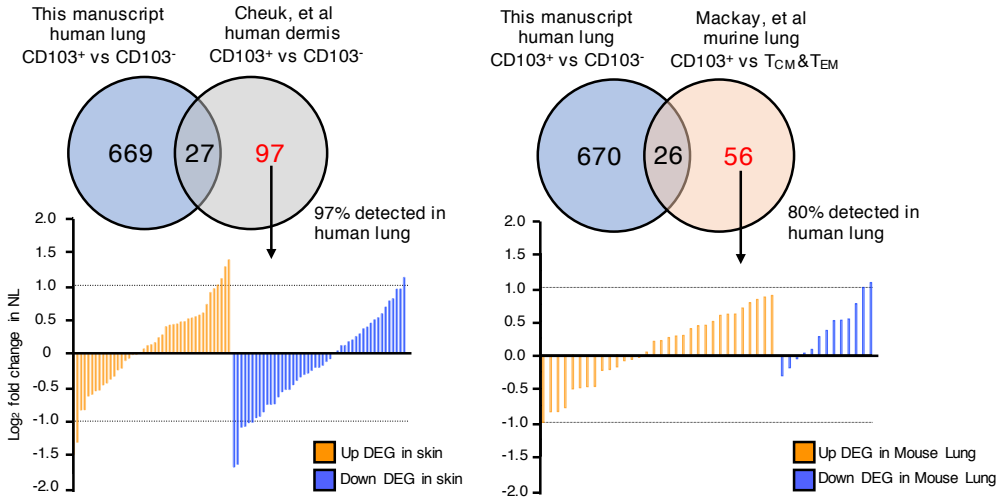


Figure 35. Distinct Program of human Lung T_{RM}.

Human Dermis data was analysed as per this study and the DESeq2 Stabilised Fold Change of genes not found to be significant in the NL are shown. For Mackay, et al., the data was analysed with Limma using FDR < 0.05 and a fold change of 2, to be consistent with the human data. This data emphasises the unique profile of T_{RM} isolated from each tissue.

5.2.4 Preserved transcriptomic elements in Tumour-infiltrating CTLs highlights enhanced clonal expansion of T-T_{RM}.

We next asked if there are common features in both Tumour (T`) and Normal lung (NL) $CD103^+$ CTLS. Unbiased visualization using 2D t-stochastic neighbour embedding (tSNE) revealed the distinct transcriptomic profile of the four populations (Figure 36A) defined by CD103 expression and Tumour status of the source tissue. We observed 90 genes which were differentially expressed both when comparing $CD103^+$ *versus* $CD103^-$ cells from normal lung as well as CD103+ and CD103- cells from Tumours (Figure 36B). Of these 90 genes, *TNFRSF9* encoding 4-1BB, was differentially upregulated in CD103+ cells from the Tumour but downregulated in CD103+ cells from the normal lung, suggesting that in the tumour, this population was enriched for antigen-specific activation (Bacher et al., 2016). 606 genes were uniquely differentially expressed in the NL. We evaluated the behaviour of these genes in Tumour-derived CTLS, and compared the expression levels between the Tumour $CD103^+$ and $CD103^-$, emphasizing these 606 genes were *lung specific features*. Further analysis of the shared 90 DEGS included transcripts encoding inhibitory molecules and other proteins that may represent 'human lung T_{RM} functional molecules' (Figure 37B). Of note we identified a divergence in transcripts for multiple receptors traditionally associated with NK function, such as *KLRC1* and *KLRF1*, encoding CD159A and NKp80 respectively, which we hypothesise may be involved in cell sub-type specific activation. Furthermore, we identified upregulation of *SPRY1,* (Collins et al., 2012) *MYO7A* (Zheng et al., 2017) and *CD226* (Chan et al., 2014)*,* which have all been suggested to be involved in controlling T-cell responses, demonstrating the exquisite control found in human T_{RMS}. We also identified *XCL1* which has been shown to be involved in the recruitment of highly potent dendritic cells, which may reflect some of the superior functional recruitment properties of $CD103^+$ cells (Ohta et al., 2016). *GPR25* has high homology to *GPR15*, which has been shown to be necessary for organ specific localization of T-regulatory cells (Kim et al., 2013b).

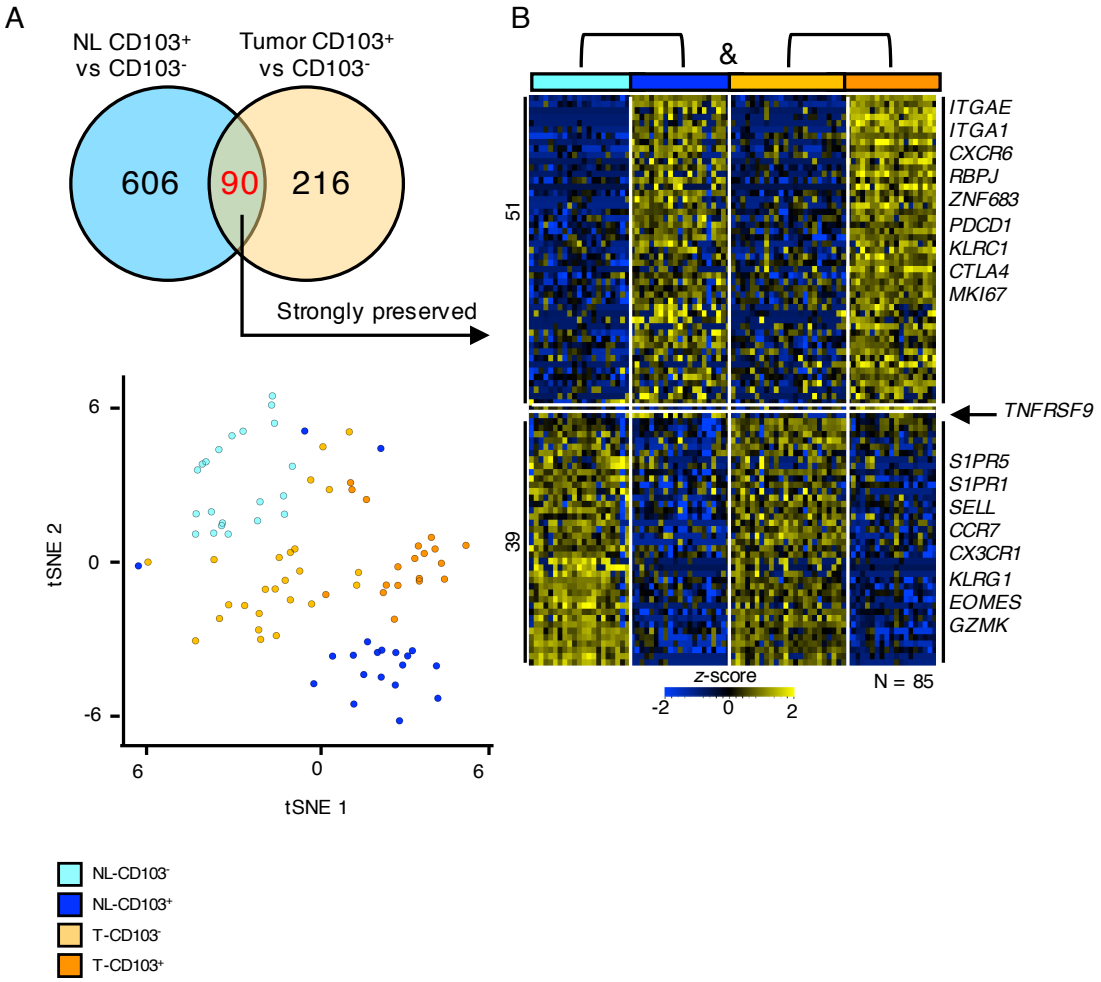
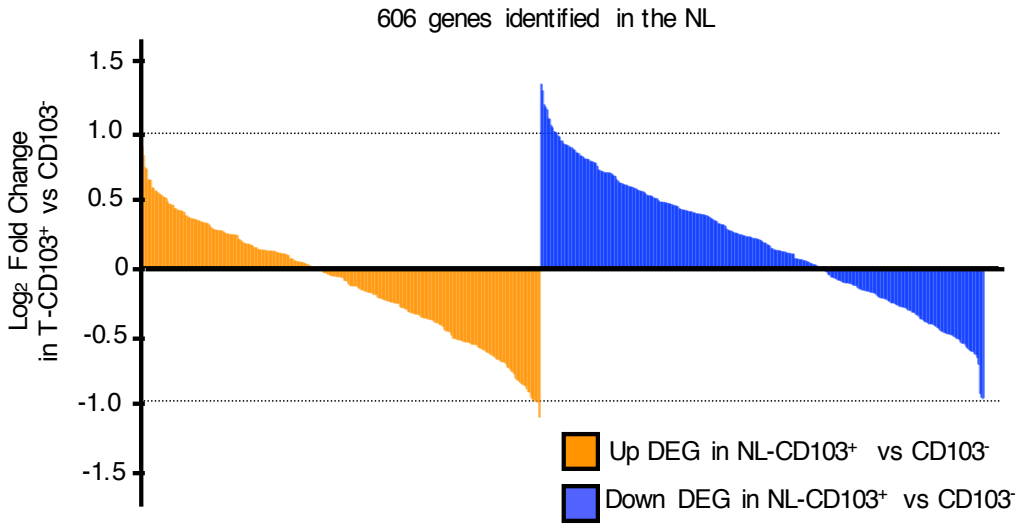


Figure 36. Preserved elements in NL and tumour T_{RM}.

Visualization of RNA-seq data, using 2D t-stochastic neighbour embedding (tSNE) using 10 principle components and ~2500 most variable genes, is shown (A). RNA-seq analysis showing row-wise z-scores of transcripts per million counts for each differentially expressed gene (rows) obtained by pairwise comparison of lung $CD8^+CD103^+$ versus lung $CD8^+CD103^-$ (n= 85) (DESeq2 analysis, Benjamini-Hochberg adjusted $P < 0.05$ and 2-fold change (B).

A)


606 genes identified in the NL
Log2 Fold Change in T-CD103+ vs CD103-
1.5
1.0
0.5
0
-0.5
-1.0
-1.5
Up DEG in NL-CD103+ vs CD103-
Down DEG in NL-CD103+ vs CD103-

B)

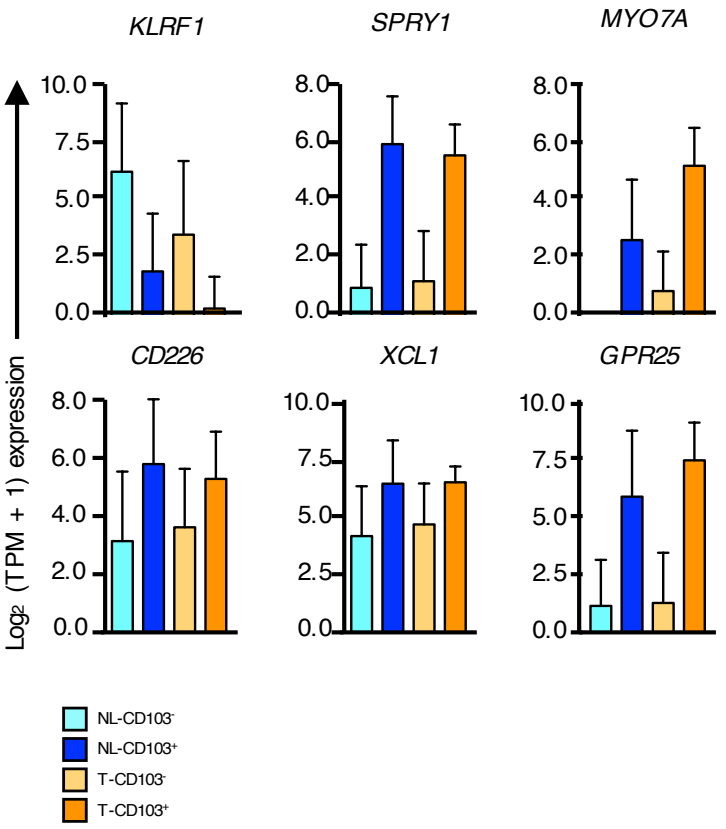
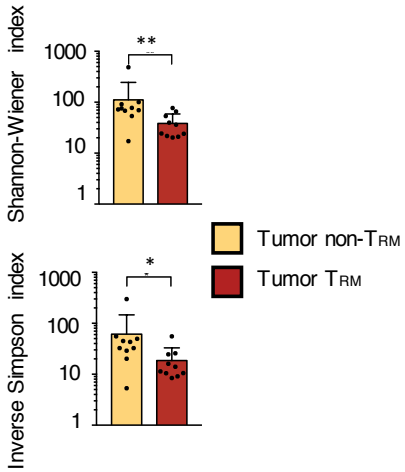

KLRF1
SPRY1
MYO7A
CD226
XCL1
GPR25
Log2 (TPM + 1) expression
NL-CD103-
NL-CD103+
T-CD103-
T-CD103+

Figure 37. Non-preserved and preserved elements in NL and Tumour T_{RM}.

The fold change of the 606 genes (previous figure) not identified in the Tumour were shown to have a random distribution in the Tumour. The DESeq2 Stabilised Fold Change in the tumour localised CTLS that were not found to be significant in the NL are shown. Further investigation on the 90 genes commonly shared between the two compartments identified a further set of transcripts encoding molecules of interest (B). Mean +/- SD is shown of genes significantly different between T_{RM} and non-T_{RM} in both the lung and tumour as per Figure 36.

5.2.5 TCR-seq identified enhanced clonal expansion in T-T_{RM}.

A)



B)

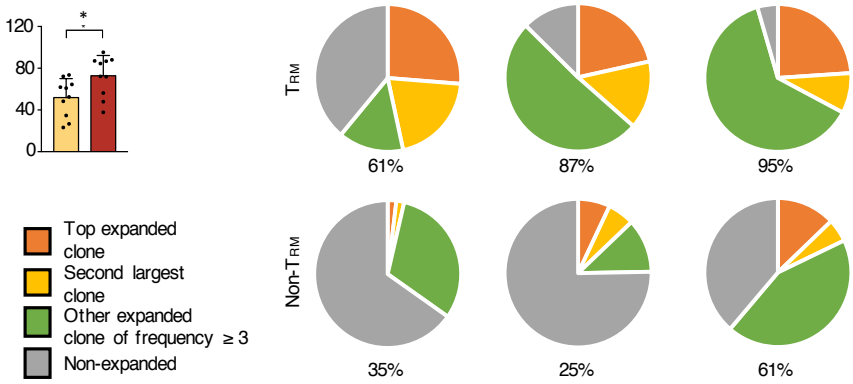


Figure 38. Clonal expansion in T-T_{RM} is identified via TCR-seq.

(A) Shannon-Wiener diversity and Inverse Simpson indices obtained using V(D)J tools following TCR-seq analysis of β chains in tumour TRM and tumour non-TRM populations. Bars represent the mean, t-line the s.e.m., and symbols represent individual data points (*P ≤ 0.05; **P ≤ 0.01; n=10 patients). (B) Left, bar graphs show the percentage of total TCRβ chains that were expanded (≥3 clonotypes). Bars represent the mean, t-line the s.e.m., and dots individual data points (* P ≤ 0.05; n=10 patients). Right, pie charts show the distribution of TCRβ clonotypes based on clonal frequency.

Furthermore, in-keeping with our common upregulation of *MKI67,* we identified increased clonal expansion in $CD103^+$ compared $CD103^-$ populations (Figure 38) using TCRβ sequencing, using both measures of statistical diversity and percentage expanded (Patil et al., 2018). Further, we also found shared TCRs between the tumour and NL, suggestive of local 'patrolling' by these resident cells (Ariotti et al., 2012). By using Ingenuity pathway analysis (IPA) for both $CD103^+$ and $CD103^-$ populations from the Tumour but not NL, we identified a common upregulation of pathways related to presumed activation by tumour antigen IFNγ and TNF, among others, (Figure 39); this is in-keeping with our preceding publication. We therefore next wished to understand the specific molecular characteristics of T-T_{RM} in more detail.

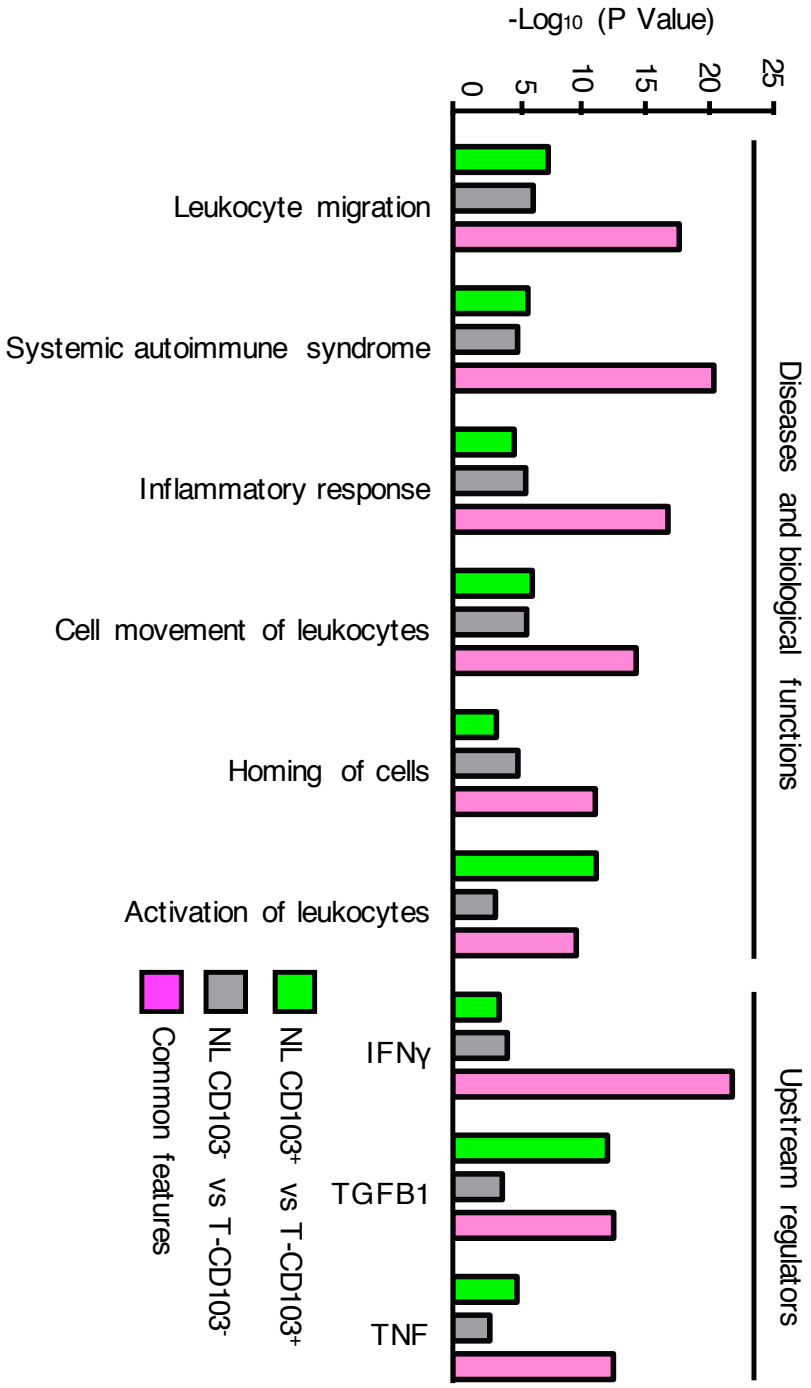


Figure 39. TILS are enriched for activation and cell cycle regardless of T_{RM} or non-T_{RM}.

Pathway analysis of genes common and uniquely differentially expressed using Ingenuity Pathway Analysis. This data confirms our interpretation that the dominant features of the transcriptomic changes between the cells isolated from tumour and lung are features of activation and proliferation.

5.2.6 Unique Features of T-T_{RM}.

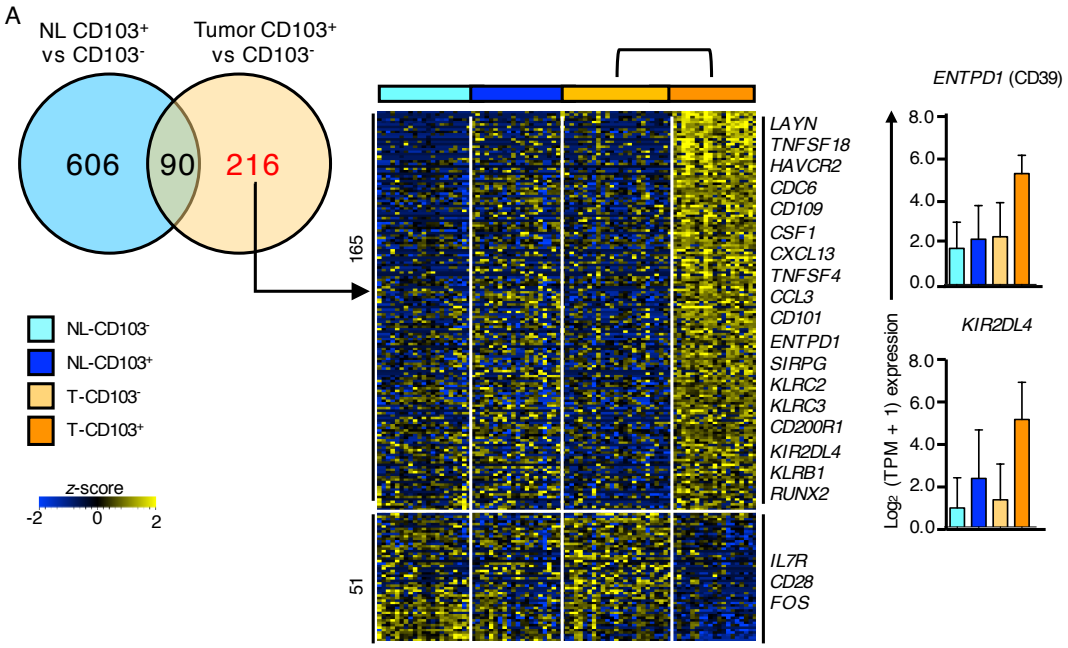
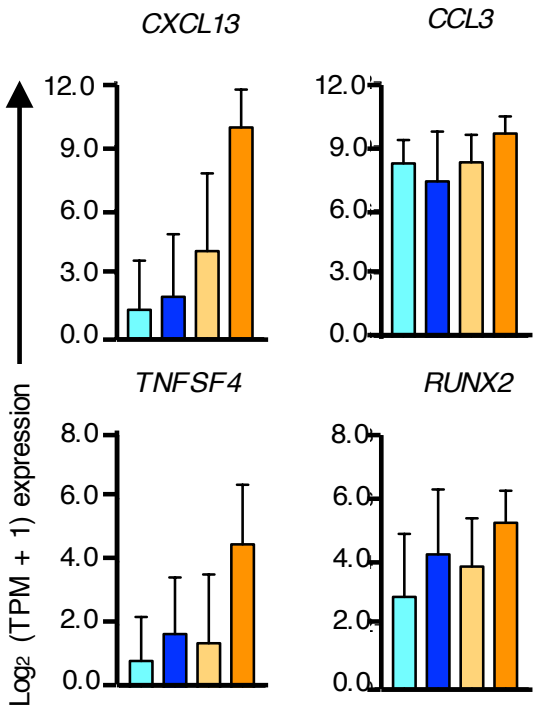


Figure 40. Unique properties of T-T_{RM}.

RNA-seq analysis showing row-wise z-scores of normalised read counts for each differentially expressed genes (rows) obtained by pairwise comparison of $CD8^+CD103^+$ versus lung $CD103^+$ (n= 85) (DESeq2 analysis, Benjamini-Hochberg adjusted $P < 0.05$ and 2-fold change). Example expression of previously predicted genes, CD39 and KIR2DL4 are shown.

A)



B

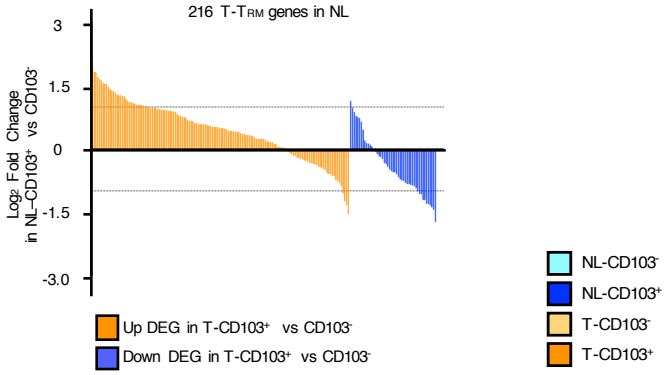


Figure 41. Unique properties of T-T_{RM} are not found in the NL.

Expression of selective genes identified from differential expression analysis as being significantly differentially expressed (Figure 40;mean +/- SD) (A). The DESeq2 Stabilised Fold Change of genes in the NL localised CTLS that were not found to be significant in the NL differential expression profiling, are shown (B), demonstrating the lack of conservation of the expression patterns between the two tissue locations (tumour and NL).

Uniquely detectable in the Tumour CD103^{+}/103^{-} comparison we had identified a series of 216 transcripts (Figure 40) including molecules we had previously validated, such as *ENTPD1*, encoding CD39 (Ganesan et al., 2017). As expected, given the unique upregulation of *TNFRSF9 (4-1BB),* we identified multiple transcripts likely reflective of enhanced activation by interaction with tumour antigens (ATA), such as *LAYN, HAVCR2* and *TNFRSF18* and related increased cell division (*CDC6, TOP2A).* Further inspection also identified transcripts with known recruitment and/or supportive properties, potentially reflecting a mechanism underpinning the improved survival of CD103^{hi} patients (Figure 41A) (Djenidi et al., 2015; Ganesan et al., 2017). For instance, CXCL13 is a chemokine established for the recruitment of B cells and T Follicular Helper cells, while CCL3 and CCRL2 are more studied in T cell (Honey, 2006) and neutrophil (Otero et al., 2010) recruitment, respectively. Heightened expression of *TNFSF4* (encoding OX40L) may reflect a supportive role of T_{RM} in furthering other T cell responses (Croft et al., 2009). Given the importance of *RUNX3* in murine T-T_{RM} (Milner et al., 2017), we speculate the human system may also use *RUNX2* which is a significant hit in our dataset. Finally, we noted upregulation of NK receptors *KIR2DL4, KLRB1, KLRC2* and *KLRC3,* which, given the downregulation of CD28 transcripts, may reflect a T-T_{RM} use of alternative molecules as co-receptors (Jabri and Abadie, 2015). We also noted a tumour-specific downregulation of *IL7R* (encoding CD127). A waterfall plot emphasized these 216 were lesser preserved in the NL (Figure 41B), demonstrating that these T-T_{RMS} features must reflected a distinct population.

5.2.7 T-T_{RM} have enhanced expression of GZMA following stimulation.

A)

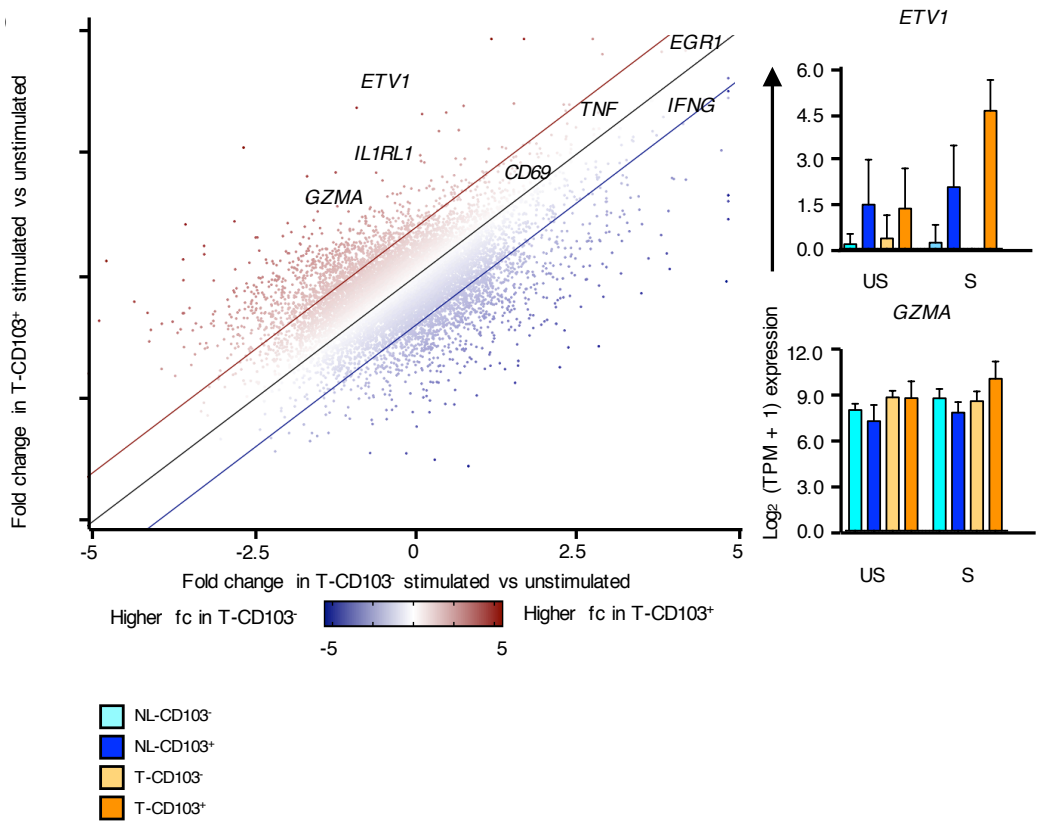


Figure 42. Unique properties of T-T_{RM} following stimulation.

Cells were stimulated with PMA/Ionomycin for 4 hours (S) compared to an unstimulated control (US). Fold Change is DESeq2 variance stabilised, correlation is spearman. Bars are mean +/- Standard Deviation.

5.2.8 T-T_{RM} have multiple unique modules.

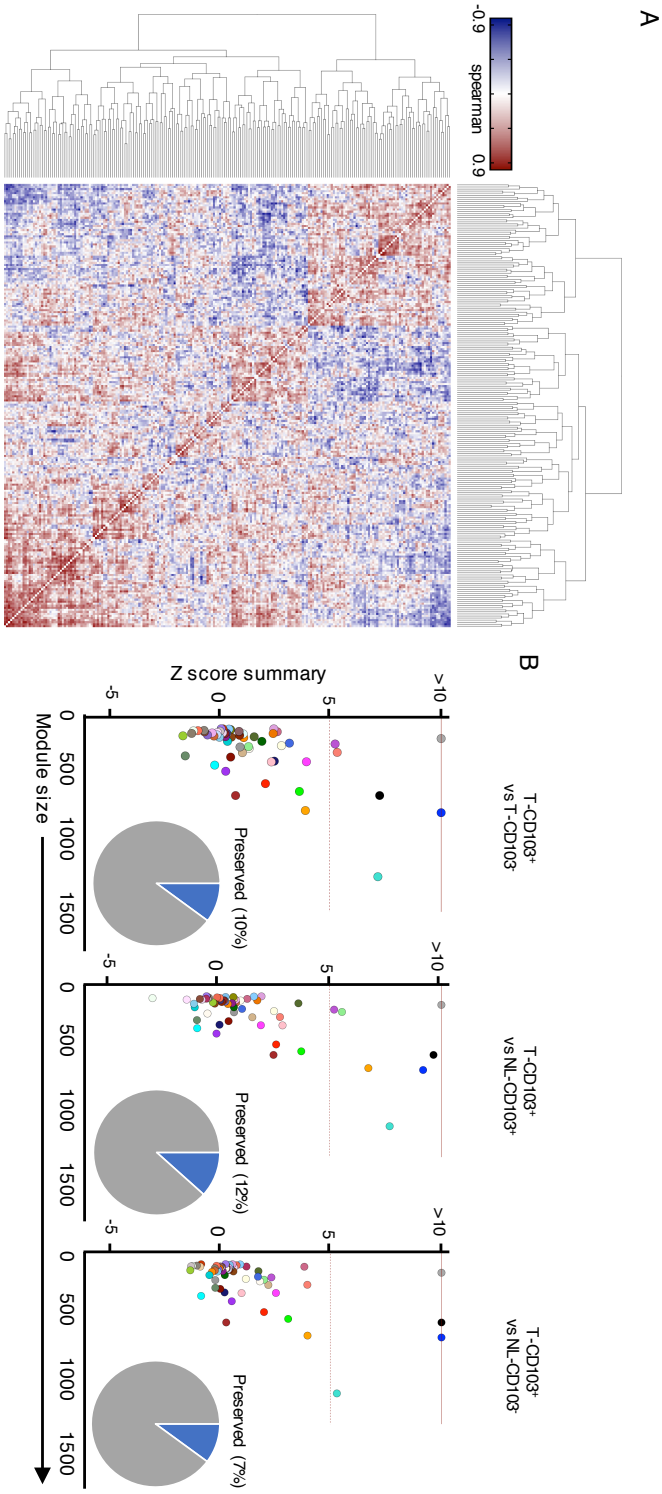


Figure 43. T-T_{RM} have unique modules.

The 216 genes uniquely differentially expressed in T-T_{RM} were clustered (spearman correlation; ward linkage) (A). Signed WGCNA analysis using a beta value of 8, created from the T-CD103$^+$ transcriptome, divided the transcriptomic network into distinct modules. Module preservation was calculated using 30 permutations and a negative control of randomly assigned genes. Preservation was considered for values with a Z score above 5.

To further probe the features of this population we completed RNA-seq following 4 hours stimulation, which identified a unique upregulation of the transcription factor *ETV* and cytotoxic molecule *GZMA,* with almost an 8 and 3-fold upregulation, respectively (Figure 42). This unique upregulation of *GZMA* suggested there may be some enhanced cytotoxic potential in these T-CD103$^+$ cells. Given we had previously identified several unique features of CTLs that were reflective of the level of immune infiltrate, we next wished to investigate if there were changes in the molecular profile of our purified T_{RM}.

To investigate these populations in more detail we next asked if the DEGs specific to these T-T_{RM} represented a singular program or multiple distinct elements. Unbiased co-expression analysis identified multiple sub-modules, suggestive of multiple biological processes (Figure 43A), although this co-expression matrix did include a large module (top left) which included genes such as *HAVCR2* and cell cycle associated genes (discussed later). To approach this at a genome-wide scale, WGCNA-based modular analysis found poor preservation from the T-CD103$^+$ compared to other populations (Figure 43B). Clustering of these modules (Figure 44) identified minimal correlation across the weighted network, implying that multiple unlinked biological processes were unique to the Tumour-CD103$^+$ population. Given the prognostic importance of immune infiltrate levels, we were interested to identify if any of these T-CD103$^+$-specific modules were linked to these biomarkers. Strikingly, we identified that multiple modules had strong correlations with infiltration levels, suggesting that it is a *qualitative* and *quantitative* change in these highly infiltrated tumours. Importantly, these were not in one branch of the module dendrogram, demonstrating that several processes were correlated to immune infiltrate levels. Two examples from the 216 T-CD103$^+$ specific genes are shown: *IL7R* & *CDC6*; (Figure 45A), of which 16%, were assigned into significant modules (Figure 45B). As we had identified two modules, assigned a random colour, in this case, mediumpurple3 and mediumorchid, which were correlated to the level of immune infiltrate, we hypothesized that these modules may themselves, be correlated. Further,

to understand the inter-module relationship, we looked to identify if there was any correlation between these modules which could suggest these are gene expression programs that are inter-linked. Strikingly, the quantification of module eigengenes (a construct representing a module) were statistically significantly correlated at a R = -0.69, emphasizing that these two modules were likely reflective of a phenotypic shift between TIL^{hi} and TIL^{lo} tumours (Figure 45C). We visualized the mediumorchid module, which was a tightly co-expressed module, of which we visualized the hub genes and others of known biological significance (Figure 45D), which shows the dominant presence of genes negatively correlated to immune infiltrate levels. Given we were identifying multiple processes, we needed to move to scRNA-seq to investigate if these changes reflected potential subpopulations or a more global phenotypic change with the hypothesis that genes in mediumorchid should be enriched in TIL^{lo} patients.

5.2.9 Unique modules in T-T_{RM} are correlated to immune infiltrate levels.

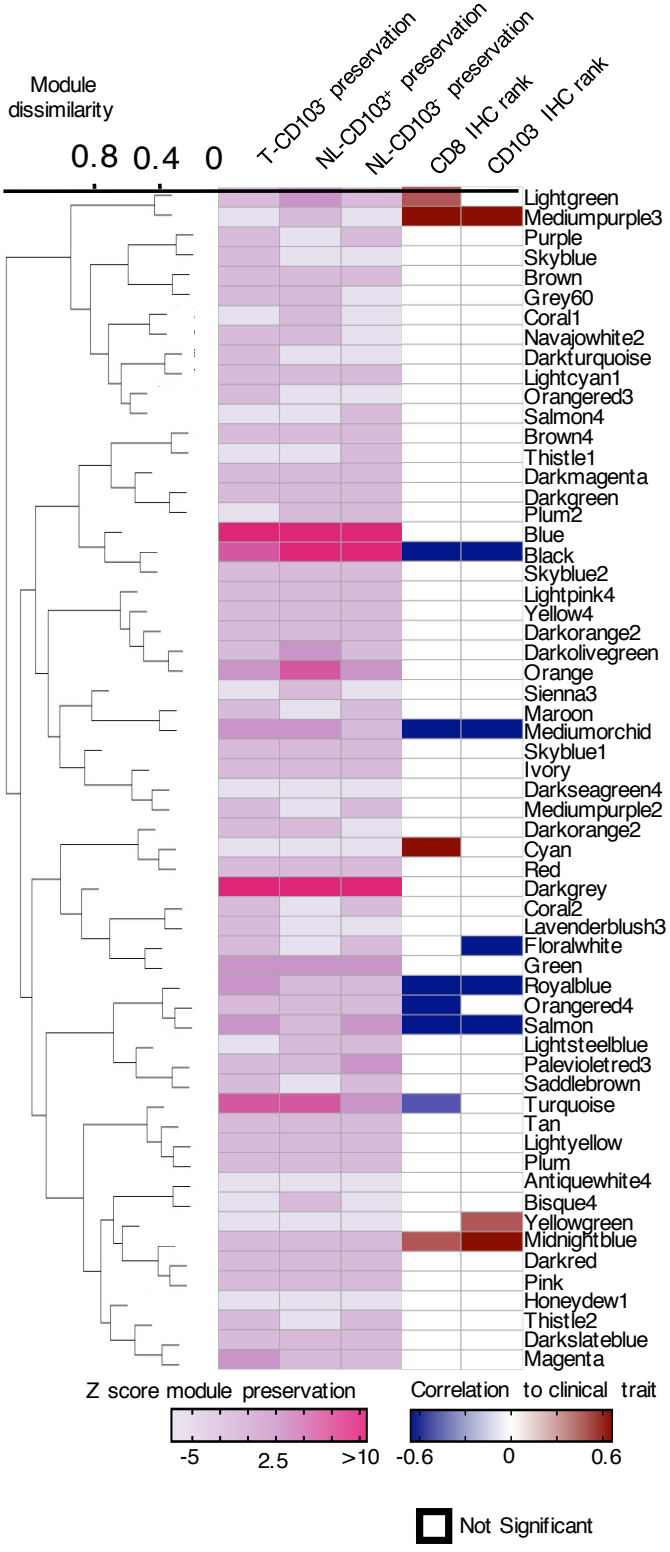
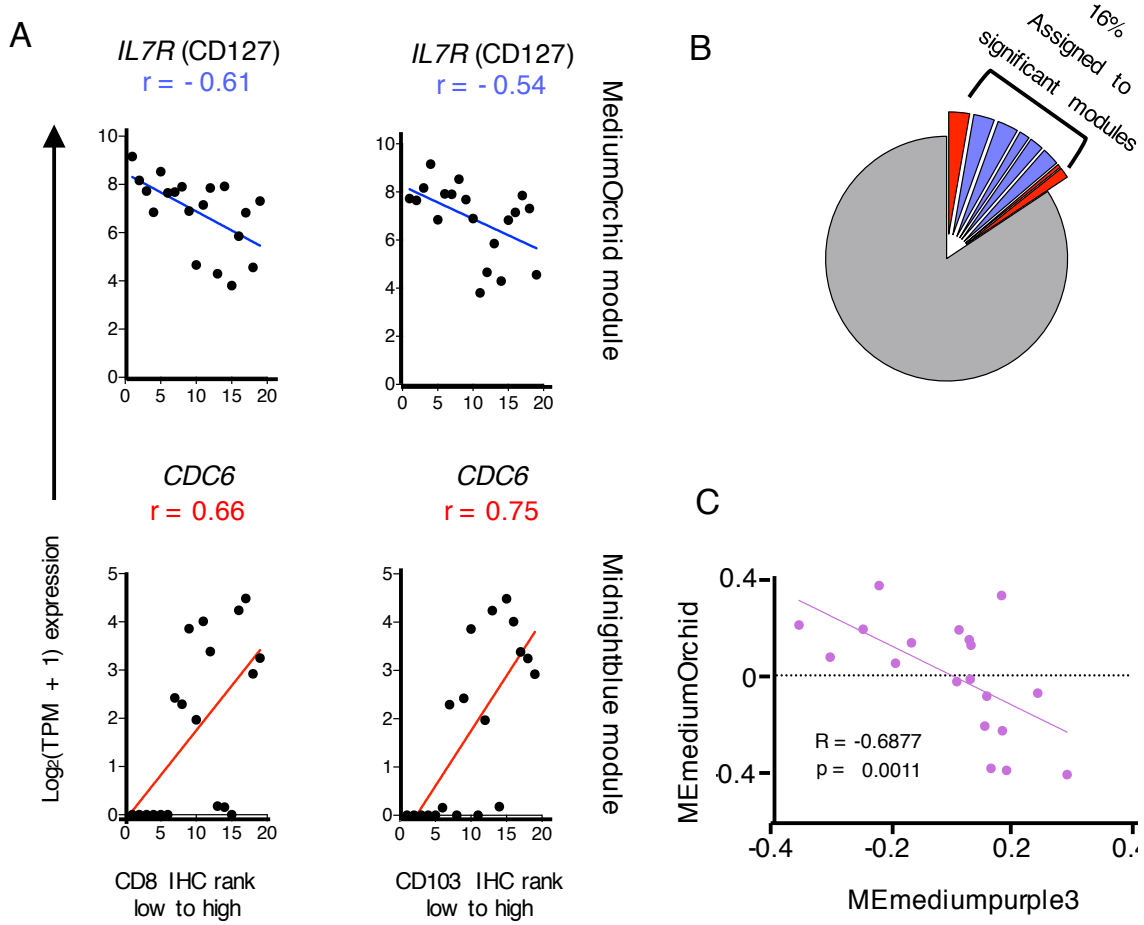


Figure 44. TI-T_{RM} unique modules correlate to known prognostic indicators.

(A) The eigengene modules identified in T-T_{RM} were clustered and the module preservation Z score was visualised, the significant correlation to levels of CD8 and CD103 (Spearman) are shown. On the left, this identifies how well a module is preserved, with values greater than 5 being represented of transcriptomic features that are found in multiple cell types studied. The right hand then studies a particular module and reflects if that process is linked to a known clinical parameter, in this case the amount of immune infiltrate. Each module is assigned a colour as a unique identifier which is written on the right side of the heatmap. Examples of some modules are in the following figure.


A
IL7R (CD127)
r = - 0.61
IL7R (CD127)
r = - 0.54
MediumOrchid module
CDC6
r = 0.66
CDC6
r = 0.75
Midnightblue module
Log2(TPM + 1) expression
CD8 IHC rank
low to high
CD103 IHC rank
low to high
B
16%
Assigned to
significant modules
C
MEmediumOrchid
R = -0.6877
p = 0.0011
MEmediumpurple3

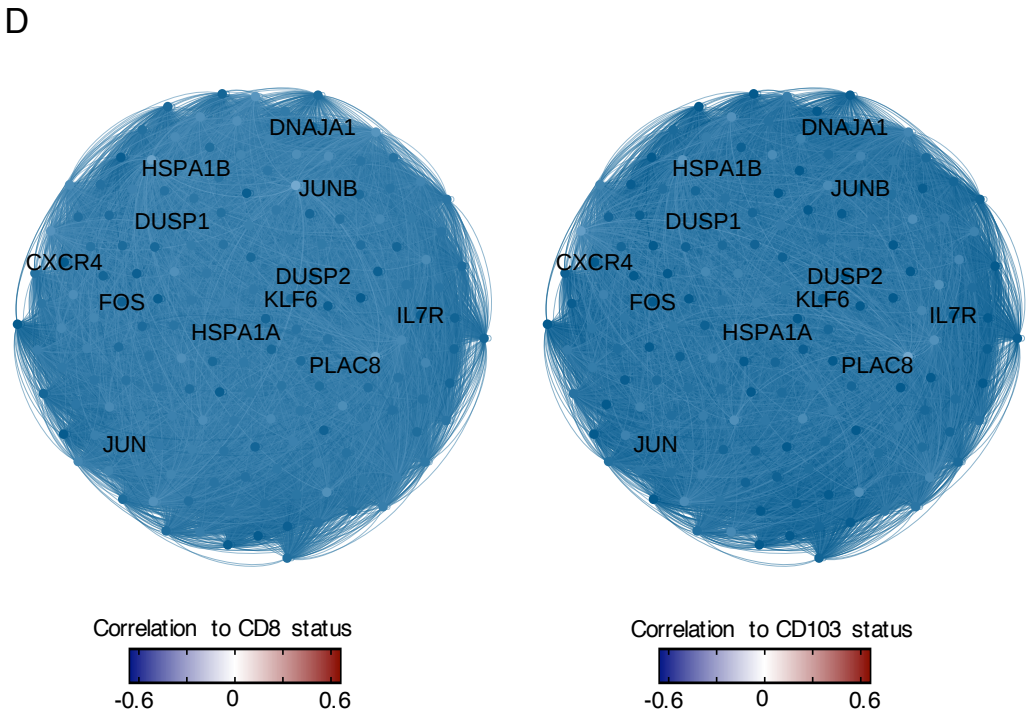

D
DNAJA1
HSPA1B
JUNB
DUSP1
CXCR4
DUSP2
FOS
KLF6
IL7R
HSPA1A
PLAC8
JUN
Correlation to CD8 status
-0.6
0
0.6
Correlation to CD103 status
-0.6
0
0.6

Figure 45. TI-T_{RM} unique modules correlate to known prognostic indicators B.

The spearman correlation between two example genes that were correlated to immune infiltrate level and assigned into significant correlated modules are shown (A) of a total of 16% of the uniquely differentially expressed genes (B). Given mediumorchid and mediumpurple both correlated to Immune infiltrate levels, the expression of the eigengene (ME) was shown (Spearman; C). Gephi Visualisation of mediumorchid module, with nodes coloured by a gene's specific correlation to CD8 and CD103 IHC counts (Forceatlas2; D). Note, given the network is generally a dark blue confirms these features are linked to a low level of immune infiltrate.

5.3 Chapter 5: Discussion.

Previous transcriptomic studies of lung T_{RM} have reported divergent conclusions (Hombrink et al., 2016; Kumar et al., 2017). While murine studies have focused on relatively acute infection models, these are perhaps not representative of long lived T_{RM} in the human lung which must be maintained for decades. Here, we initially set out to understand the transcriptomic landscape of human T_{RM}. We identified that sorting on $CD103^+$ CTLS enriched for previously defined 'gold standard' markers of resident cells (CD49A and KLRG1), which we validated at the protein level with flow cytometry We initially focused on the NL-T_{RM} as there was more literature to reference and validate our findings with. One striking observation was the enhanced expression of known and postulated inhibitory molecules in T_{RM} cells in both the normal lung and tumour. We hypothesise that these cells may have superior TCR affinities or other advantages, which necessitates the expression of inhibitory molecules to prevent activation-induced apoptosis during development. More recent murine studies have demonstrated that PD1 is expressed in an antigen-independent manner in T_{RM} (Shwetank et al., 2017). To interrogate whether these NL-T_{RM} cells were in an 'exhausted' state, we completed RNA-seq after 4 hours stimulation with PMA/ionomycin. We identified equal upregulation of known effector molecules in NL-CD103+ and CD103- cells demonstrating that they were not an 'exhausted' phenotype despite the enrichment for PD1 expression specifically in the CD103+ cells. We had hypothesized that perhaps this indicated the NL T_{RM} cells may have superior functional properties in the cells isolated from the NL compartment. However, it was still an interesting observation that there was similar effector potential with the significant expression of inhibitory molecules. It may be that in view of the known heterogeneity in human transcriptomic data, that 6 patients in this stimulated RNA-seq cohort is insufficient to detect subtle differences, particularly if there are a number of subsets inside each population (see next chapter). We next moved to the common elements of both tumour and NL T_{RM}

Finally, the tissue-specific phenotype of human lung T_{RMS} despite a core conserved signature is highlighted by the limited overlap with T_{RM} described previously in acute infected murine lung and human skin. Regardless, the expression of PD1 by T_{RM} cells in the lung tissue and possibly other organs (skin, gut and pituitary gland) raises the possibility that anti-PD1 therapy may non-specifically activate potentially self-reactive T_{RM} cells or xeno-reactive T_{RM} to cause adverse immune reactions such as pneumonitis, dermatitis and colitis (June et al., 2017). Comprehensive analysis of the T_{RM} phenotype and TCR repertoire of CTLs present in Tumour and organs affected by adverse reactions is likely to lend proof for these hypotheses.

A core 90 gene signature was conserved between the T_{RM} in tumour and normal lung; we found poor preservation of the un-shared genes emphasizing that despite a core of shared features, T-T_{RM} were a distinct phenotype. Further inspection of the 216 T-T_{RM} specific gene signature indicated this population was likely receiving enhanced tumour derived antigen-specific activation. The enhanced expression of markers of antigen-specific activation or 'exhaustion' on tissue resident cells may suggest that T_{RM} are the population responding to checkpoint blockade, given these cells are the majority of the $CD8^+$ population expressing the targets of these therapeutic agents (Couzin-Frankel, 2013). Akin to the NL, the Tumour T_{RM} were further enriched for PD1 compared to the non-T_{RM}, but following stimulation we identified equivalent or greater upregulation of IFNG, and other effector functions.

Further investigation using an unbiased genome-wide approach identified mediumpurple3, and mediumorchid, among other modules that correlated to the level of CD8 and CD103 infiltration in a particular patient's tumour. When focused upon these two modules in particular and identified that these two modules had a strong inverse correlation. We concluded that there must be mutually exclusive programs that are correlative to the level of immune infiltrate, which should include genes such as *IL7R, CXCR4, HSPA1A, HSPA1B* which may represent genes that should be significantly reflective of clusters of subpopulations. This was a reference point to now move to single cell analysis to understand the heterogeneity inside these populations which was apparent.

Chapter 6: Results 4: single cell analysis of T-T_{RM}.

6.1 Introduction.

In this era in which single cell immunology has come of age, we had originally planned to use Smart-seq2-based methodologies to increase the resolution. However, since we had noted a large degree of heterogeneity and were identifying changes in the populations between TIL^{hi} and TIL^{lo} patients, we felt that completing analysis of more patients and more cells may provide greater information than smaller cohorts with greater resolution. Therefore, we moved to the 10x droplet based single cell RNA-seq methodology and by pooling patient's together we could enrich for TIL^{hi} and TIL^{lo} features and then, in future work, use the SNP profile for deconvolution of each pool into patient specific information. Hence, by completing single cell analysis, we could understand a number of hypotheses from the population profiling. We then continued this analysis with Smart-seq2 to validate and improve the resolution of any findings.

We hypothesised that the ability to have single cell gene expression profiles would allow us to validate the genes identified from the bulk population analysis. Additionally, if the expression of CD103 dominated the variation in the dataset it would suggest that sorting on CD103 is indeed the best option for sorting T_{RM} as we have suggested from the previous chapters. Furthermore, as we have seen in previous chapters, there appears to be significant heterogeneity inside the T_{RM} population which should be reflected in genome-wide profiling with unique subtypes. As we showed increased clonality and markers of antigen-specific engagement, using unbiased genome-wide approaches we would be able to identify if this is found in all T_{RM} or in particular subtypes. We would also be able to identify if there was an enrichment for particular subsets in TIL^{hi} or TIL^{lo} patients, in keeping with the network analysis. As discussed previously, there is a major downside to this, in that the reduced resolution of the 10x methodology could result in reduced sensitivity and an inability to complete co-expression analysis at the single cell level, particularly if the different between subsets is subtle.

Consequently, this section had two broad aims, the initial analysis was to confirm the purified population transcriptomic profiles identified in the previous chapter. The second set was to confirm if the signature identified at the population level was reflective of a singular subset or was a broad feature of the entire dataset. Finally, we wished to validate any findings at the protein level using flow cytometry and an alternative single cell methodology (Smart-seq2, as per chapter 3).

6.2 Results

6.2.1 CD103$^+$ and CD103$^-$ CTLS are transcriptomically distinct at the single cell level.

As our WGCNA analysis had suggested there were multiple biological processes, we sorted 9000 CD8$^+$ cells from 12 donors (7 TIL^{hi}, 5 TIL^{lo}) for single cell analysis using cut-offs used in our previous paper. We sorted for both CD103$^+$ and CD103$^-$ CTLs from Tumour only, grouping one set of 6 TIL^{hi} patients into one pool, and the remaining 6 patients (5 low and one high) into another, henceforth called TIL^{hi} and TIL^{lo} pools. Unbiased visualization of these cells (Figure 46) independently demonstrated the distinctive genome-wide profile of sorted CD103$^+$ and CD103$^-$ single cell transcriptomes. Analysis of the 100 genes (Figure 47) most correlated to principle component 1, representing the most defining features of the data, included T_{RM}-associated genes *ITGAE* and *ZNF683*, alongside markers of presumed activation by tumour antigen. Importantly, these genes were co-expressed with T_{RM} genes, emphasizing that T-T_{RM} must represent the dominant population of human TILS responding to the tumour. Further clustering (Seurat (Klein et al., 2015)) generated 8 clusters, which we visualized onto the tSNE plot. The 4 left clusters were enriched for CD103$^+$ cells so we termed these the T_{RM} clusters, while we hypothesized that the remaining 4 likely represented more characterized populations, such as effector cells.

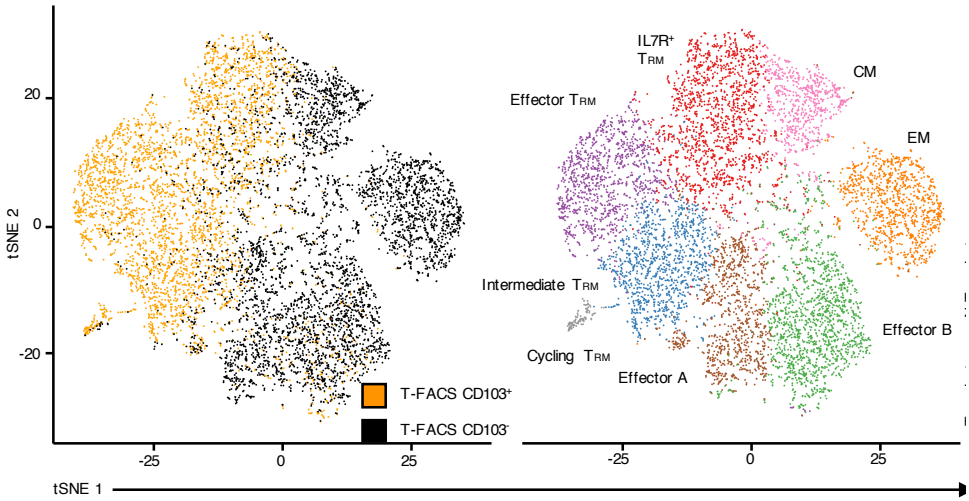


Figure 46. CD8^{+} TILS cluster into 8 unique subtypes.

tSNE visualisation of ~9,000 CD103^{-} and CD103^{+} cells from 12 donors (7 TIL^{hi}, 5 TIL^{lo}) from two pools of 6 patients. Data was processed in Seurat 2.1 and the expression of CD103 from the sort was overlaid, demonstrating a clear distinction based on CD103 protein expression. A principal components analysis was then run over the most variable genes, and the first eight principal components were used with resolution 0.6 to generate 8 clusters, a preliminary biological assignment was identified using the *findallmarkers* function in Seurat.

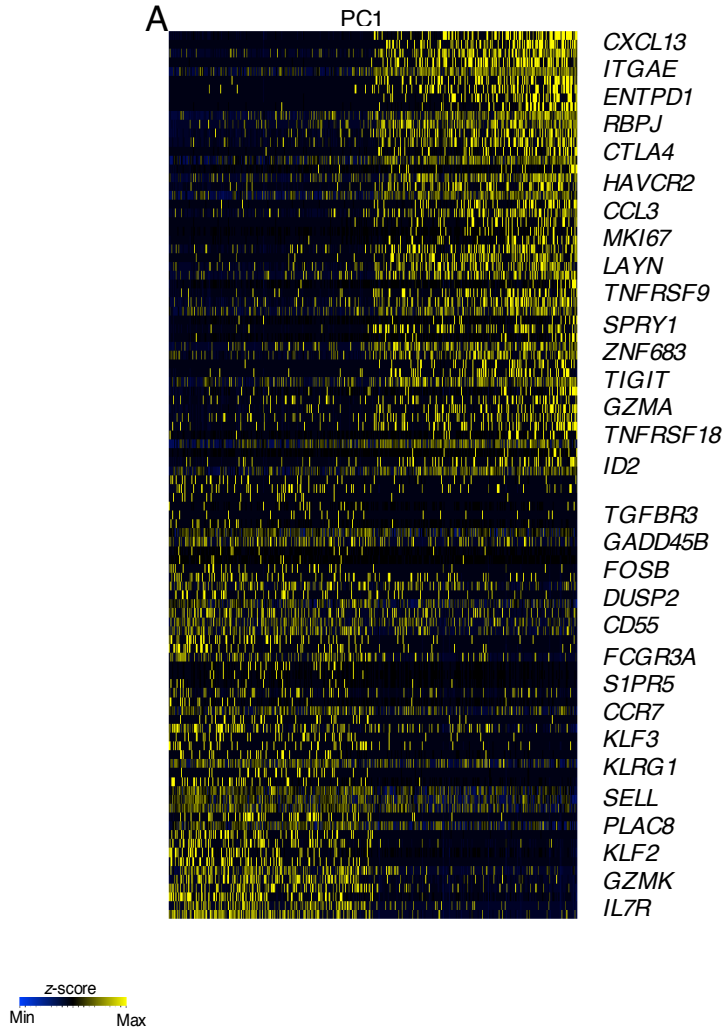


Figure 47. Principle component one is reflective of the majority of the variation and includes co-expressed activation and residency features.

The genes most correlated to principle component one both negative and positive are shown for every cell. Each column is a cell and each row is a Z score normalised expression value, normalised for the number of UMI and percent mitochondria recovered from each cell. PC1 stands for principle component 1. Highlighting the genes that dominate the variation in the $CD8^{+}$ transcriptome

6.2.2 Further investigation of cluster subtypes.

Pair-wise (cluster *versus.* every other cluster) differential expression profiling of the non-T_{RM} clusters (MAST) identified upregulation of known lineage markers and GZMK, identifying that these populations were enriched for features of Central and Effector Memory cells, alongside two clusters of effector cells (green and brown) (Figure 48). This biological assignment was owing to several clusters expressing *KLF2* and *KLF3*, which are known inhibitors of the tissue residency phenotype. Additionally, one cluster was enriched for transcripts expressing CCR7 and CD62L, likely reflective of central memory cells acquired during the tissue acquisition. Given we had detected known biologically distinct subsets, inside the $CD103^-$ clusters we next wished to study our bulk profiles and compare that to our single cell phenotype.

We next wished to understand and confirm our population-based analysis (previous chapter) inside these T_{RM} clusters. We generated the average of each T_{RM} cluster and clustered the 216 upregulated unique genes, we defined in the previous chapter of the tumour T_{RM} population. This analysis emphasized that these 216 genes were mostly reflective of the purple ("Effector T_{RM}") and grey ("Cycling T_{RM}") clusters (Figure 49A). To validate this observation, we overlaid the expression of the T-T_{RM}-specific genes onto the tSNE plot (Figure 49B); this emphasized that genes such as *CXCL13, ENTPD1, MKI67 and TNFRSF9* were enriched in the grey and purple cluster, while *IL7R* was only expressed in the red cluster ("$IL7R^+$"). As we had identified a cluster with shared features, we termed this blue cluster 'intermediate", although it is worth noting the cluster was also enriched for HSPA1A, HSPA1B, as directed by the *WGCNA* analysis in the previous chapter. Based on this analysis it was apparent the Cycling T_{RM} were related to the Effector T_{RM}, with the addition of undergoing cell cycle. This would bias differential expression profiling; hence they were not included in subsequent analysis. We now wished to understand each cluster in more detail.

Complementary methods of differential expression analysis identified a union of 1598 genes that were distinct in the effector cluster; this was visualized in a heatmap (Figure 50). Further, this analysis emphasized the highly significant upregulation of markers associated with increased presumed activation in the effector cluster. This includes transcripts encoding cytotoxic molecules, such as IFNG, Granzymes and perforin, surface molecules such as CD39 and PD1. However, it is well established that different clustering methods will provide divergent results. Using an alternative methodology, Sincell, we further demonstrated the distinct molecular profiles of these populations (Figure 51) confirming these cell subsets, in keeping with the large number of differentially expressed genes identified.

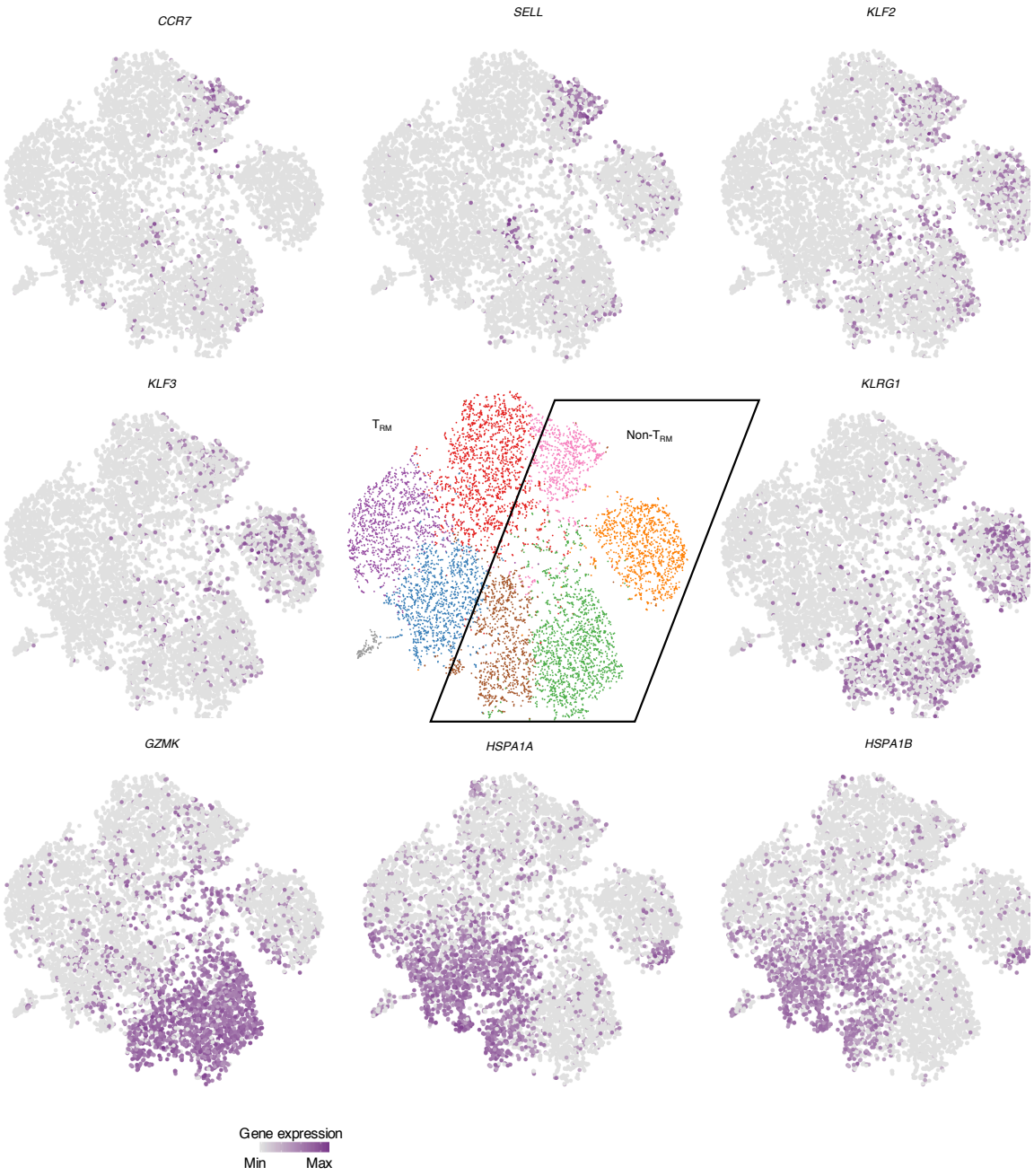
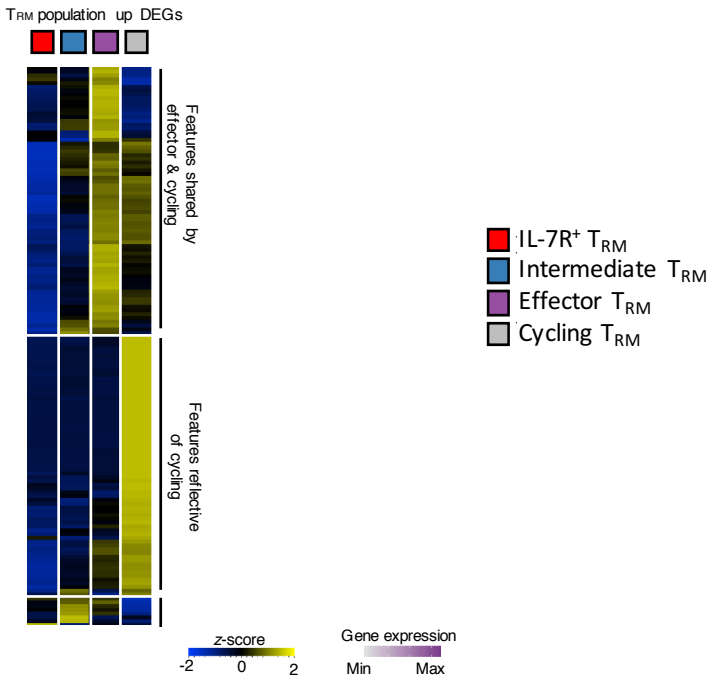


Figure 48. Non-T_{RM} clusters are reflective of effector cells, effector memory and central memory.

Pair-wise differential expression profiling using MAST identified features unique to each population. From these differential expression analysis, we selected differentially expressed genes which are known markers of particular lineages. The expression of lineage genes identified from this profiling are shown.

A



B)

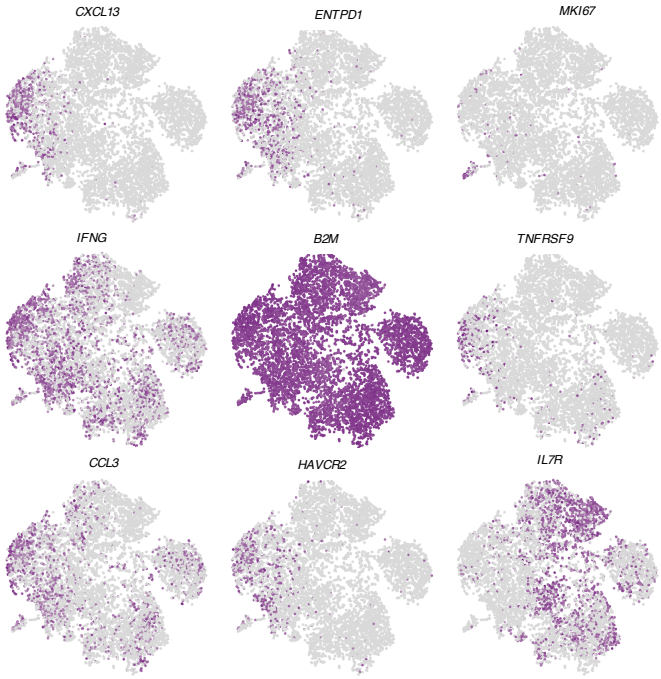


Figure 49. Population differential expression analysis is reflective of purple and grey clusters.

The average expression from the CD103^{+} cell enriched clusters was calculated using *Averageexpression* function in Seurat. Genes upregulated in the T-T$_{RM}$ population that was defined in the previous chapter are shown, variables are clustered using average linkage (A). Expression of these key population DEGs are shown over the tSNE plot (B).

A

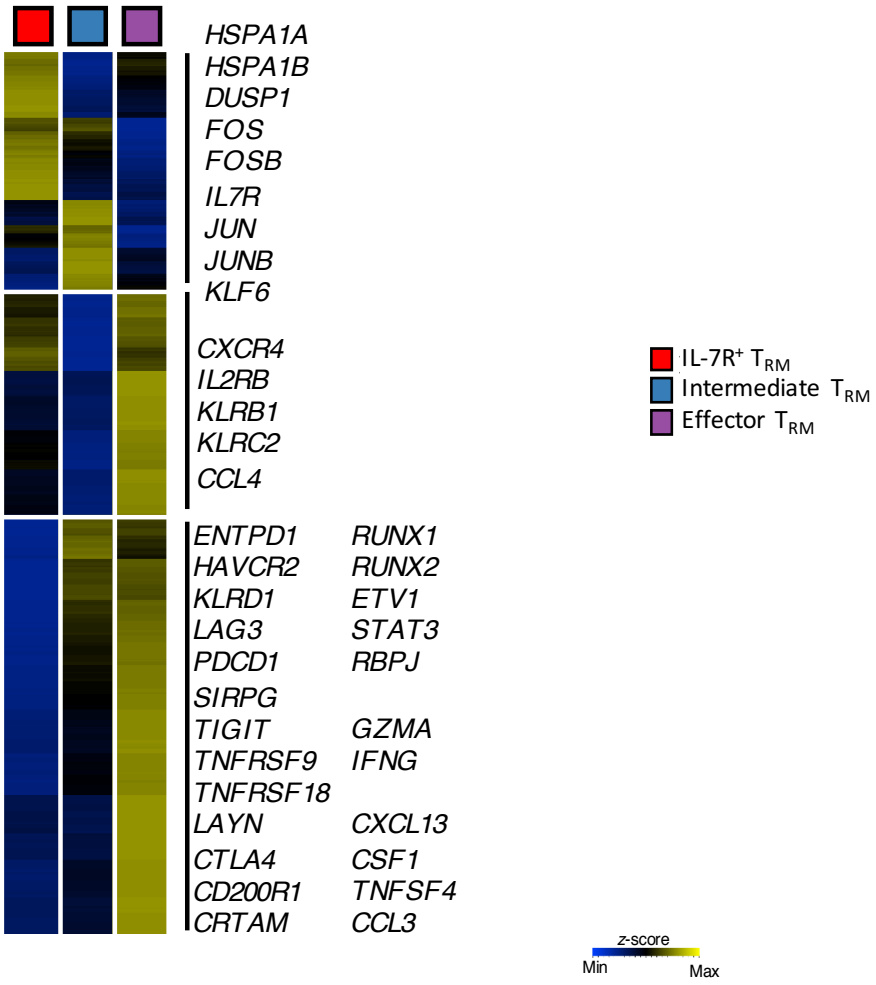


Figure 50. Single cell differential expression analysis highlights unique features of T-T_{RM} clusters.

The average expression from the $CD103^+$ cell enriched clusters was calculated using *Averageexpression* function in Seurat. The union of genes identified from differential expression profiling (MAST and SCDE) from the effector cluster are shown. This shows genes that are specific to a particular cluster. This data highlights the features of a subset that was enriched for features potentially reflective of superior functional properties.

A

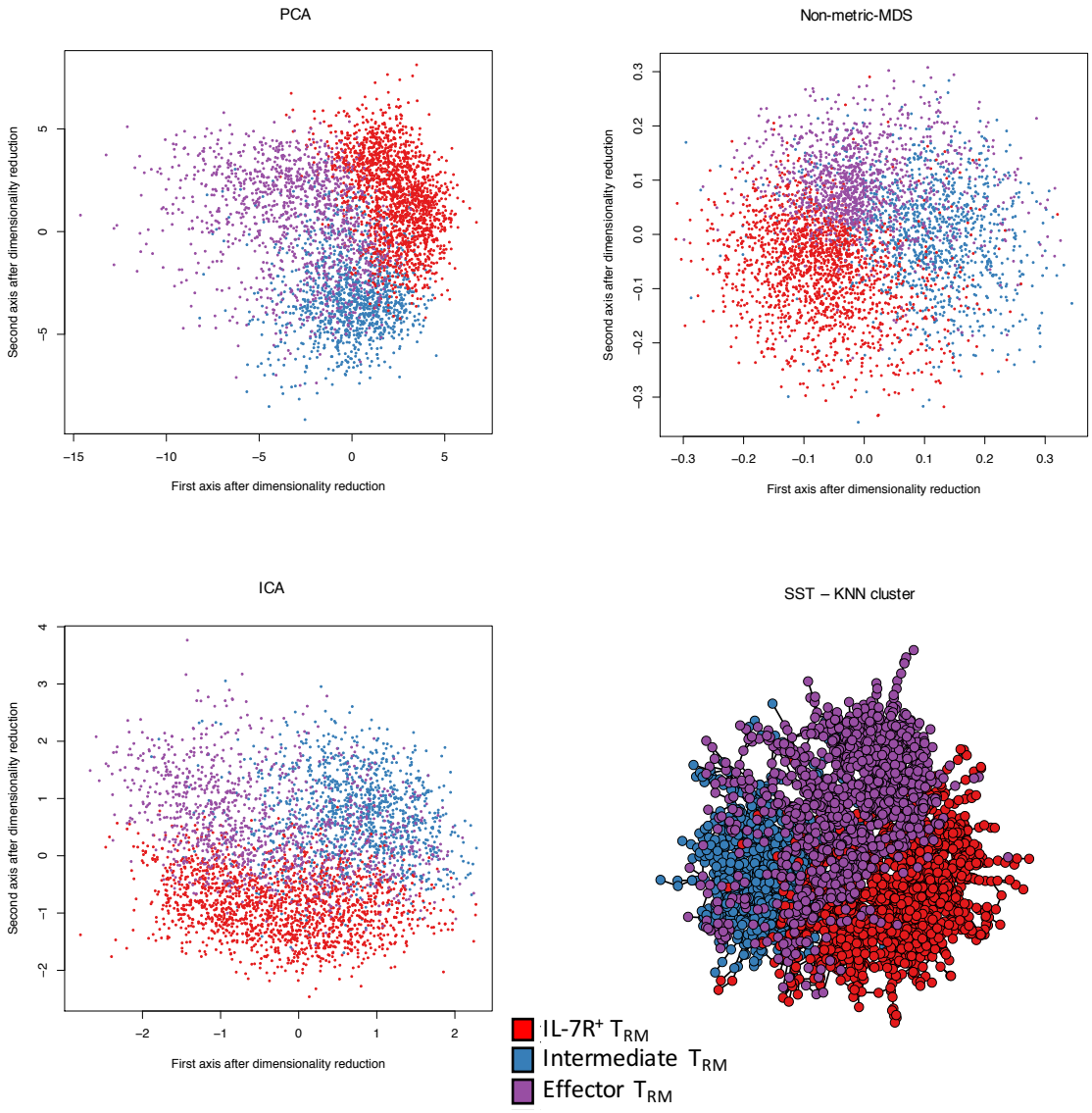


Figure 51. Sincell alternative clustering methodologies identify similar subsets of T_{RM}.

Sincell clustering was completed as per default using the variable genes identified in Seurat. Cells were overlaid with the cluster identity assigned from Seurat. Multi-dimensional scaling (MDS), Similarity Spanning Tree K-Mutual Nearest Neighbour (SST-KNN). This indeed confirmed that the clusters detected by Seurat were also identified in other dimensionality reduction methods.

6.2.3 CD127$^+$ T-T_{RM} are precursors which develop into intermediate or effector T-T_{RM}.

To understand the dynamics between clusters we generated a monocle2 hierarchy map which demonstrated that the IL7R$^+$ cluster likely formed a precursor-esque population, from which the intermediate and effector clusters diverged (Figure 52). Given we have recently shown in CD4-CTLS that IL7R$^+$ are enriched for long lived cells we hypothesise this may be a similar mechanism (Patil et al., 2018). Hence, we now wished to investigate these populations in more detail in the context of our network analysis and known clinically relevant features (i.e. TIL density), which suggested that the *IL7R* and *HSPA1A* genes would be enriched in TILlo patients.

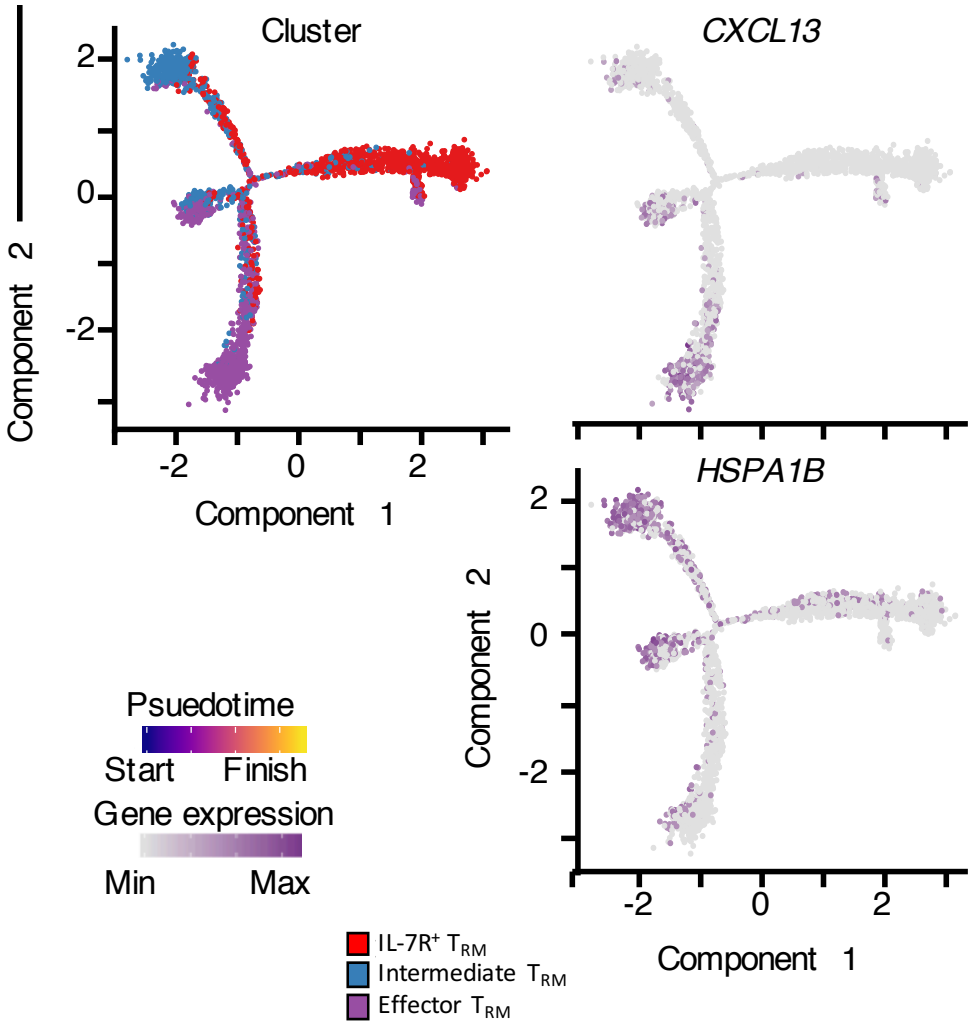


Figure 52. Monocle2 hierarchy modelling identifies precursor population expressing IL7R.

Monocle2 modelling using the most variable genes identified in Seurat identifies a cell state hierarchy. The pseudo time and Seurat assigned clusters are overlaid (Left Panels), and key known differentially expressed genes from the pairwise differential expression profiling are overlaid (Right Panels). This analysis suggests that the red cluster forms a precursor from which the cells diverge into the purple or blue cluster.

6.2.4 Effector T-T_{RM} are enriched in TIL^{hi} patients.

This analysis highlighted the genes differentially expressed, included *CXCR4, IL7R, FOS, KLF6, HSPA1A and HSPA1B*. Importantly these genes included those in the 'mediumorchid' module that was strongly inversely correlated with immune infiltrate levels. Given our WGCNA analysis had shown *CXCR4, HSPA1B,* among others, was correlated to TIL status at the population level, we hypothesized that there may be differences in representation between TIL^{hi} and TIL^{lo} patients. We found that in the TIL^{hi} patients the cycling and effector T_{RM} clusters were enriched, whilst the TIL^{lo} patients had enhanced representation in the $IL7R^{+}$ and intermediate clusters (Figure 53).

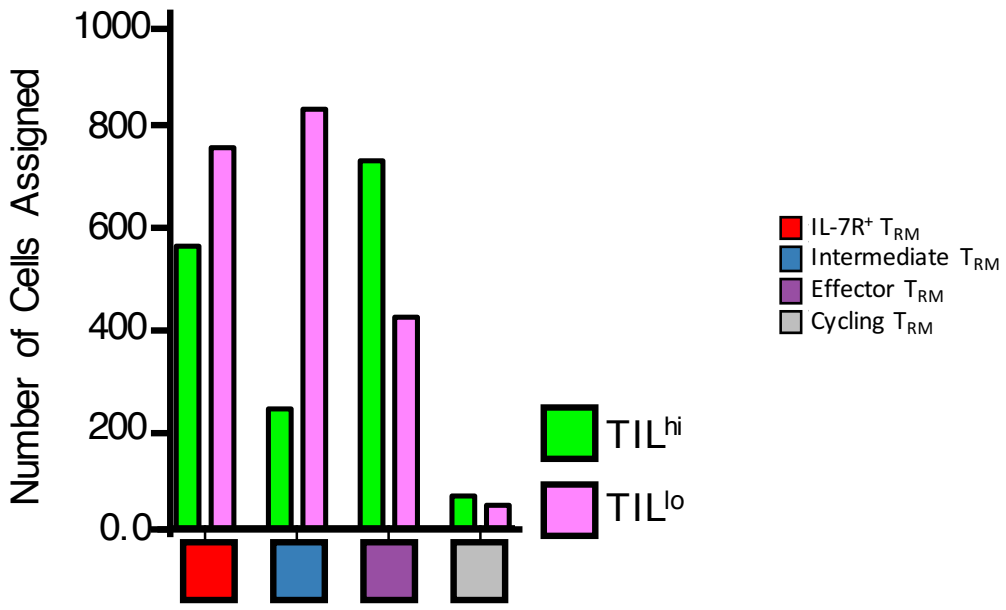


Figure 53. Effector and intermediate clusters (purple and grey) are enriched in TIL^{hi} and TIL^{lo} patients respectively.

The number of cells assigned into each cluster from the Seurat clustering from the TIL^{hi} and TIL^{lo} pools This highlights that these highly infiltrated tumours are enriched for a particular subtype of T_{RM}.

6.2.5 Effector T-T_{RM} are enriched for markers of antigen specific engagement in Lung and HNSCC cancer.

We next used flow cytometery to validate that the protein expression in these clusters was indeed reflective of transcriptomically defined populations (Figure 54). The effector population (purple) was unique to the Tumour, and significantly enriched for expression of PD1, 4-1BB and TIM3 at the protein level. This reinforced that there was a specific molecular program including CD39, TIM3 and PD1 that was associated with antigen specific engagement specifically in the T_{RM} cells isolated from the tumour. Further our transcriptomic profiling suggested this cluster was also enriched for markers of cytotoxicity, suggesting these cells to be potentially hyper functional. Having identified this novel population in lung cancer, we next wished to demonstrate it was also present in HNSCC (Figure 55). However, we did not complete statistics as this was a small cohort of only three patient samples.

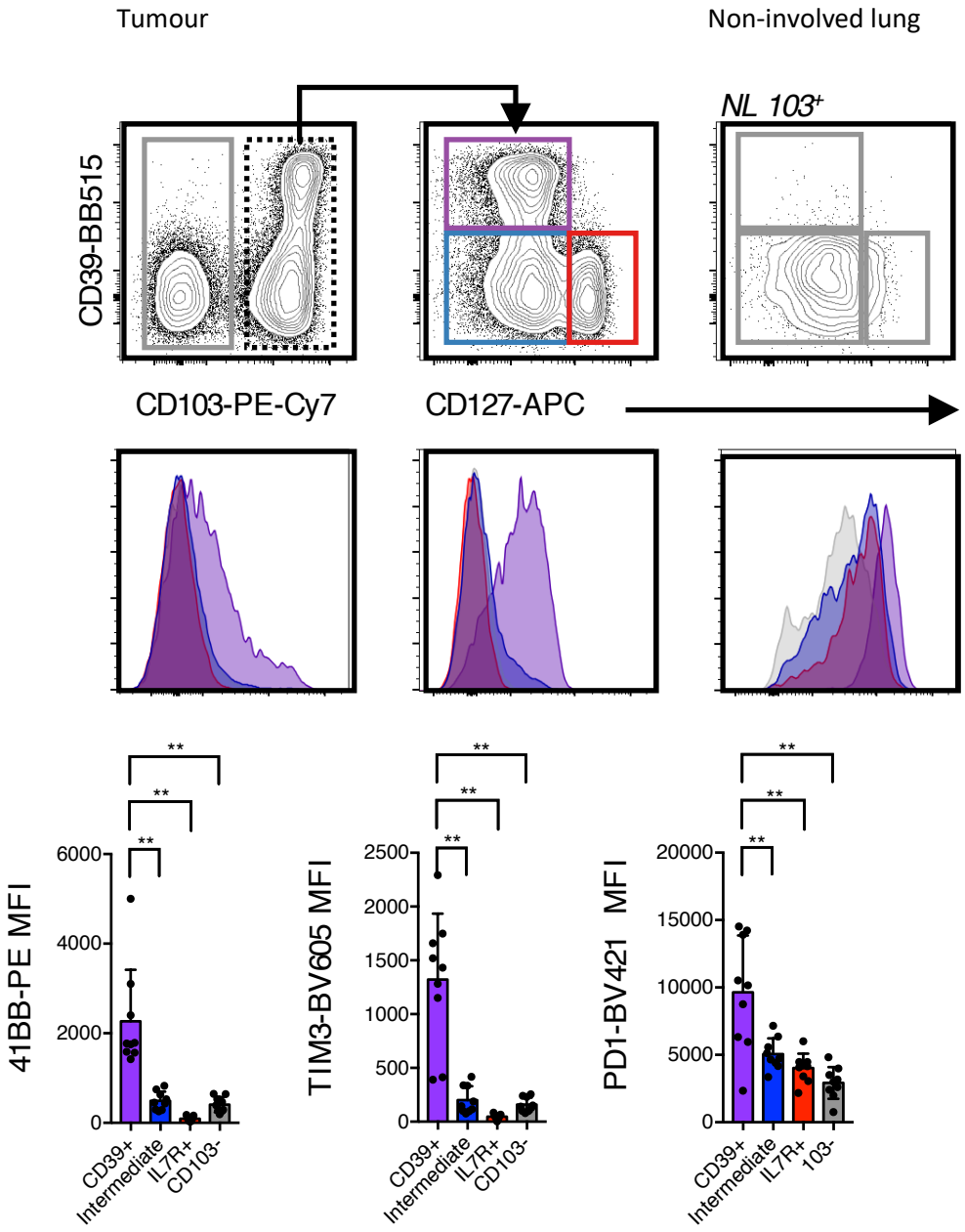


Figure 54. CD39⁺ IL7R⁻ T-T_{RM} subtype is enriched for markers of antigen-specific engagement.

Geometric mean fluorescent intensity (GMFI) of 4-1BB, TIM3 and PD1 for each tumour T_{RM} subset; bars represent the mean with standard deviation, and each symbol represents data from individual samples (**P ≤ 0.01; n=8); representative histograms shown. The CD39 and IL7R clusters were identified as strong markers to segregate the Tumour-T_{RM} cells from the transcriptional profiling. This indeed confirmed that a cluster predicted from the purified population RNA-seq and the single cell transcriptomics could be identified at the protein level. Non-involved lung (NL)

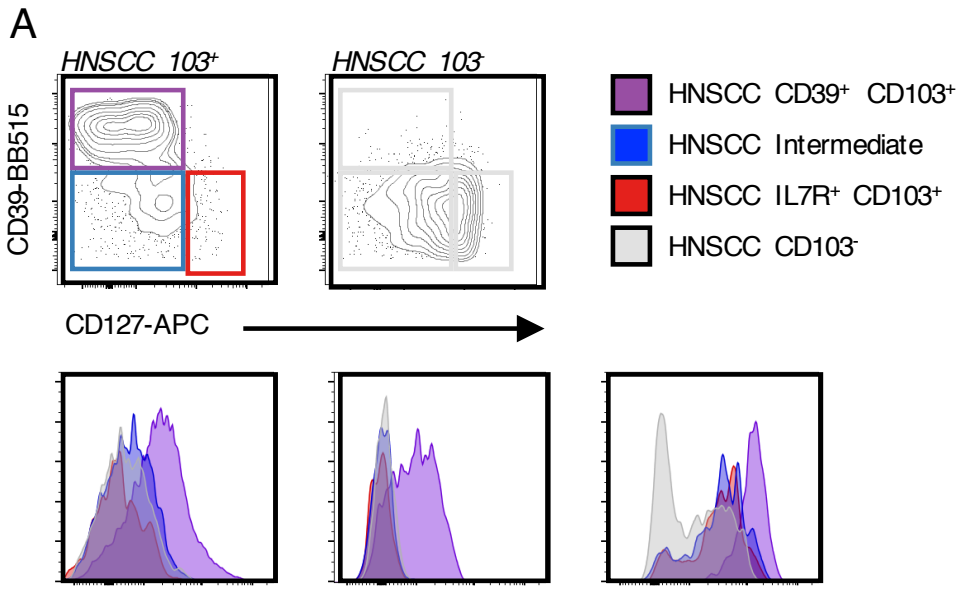


Figure 55. CD39+ IL7R- T-T_{RM} are also found in HNSCC, with enrichment for markers of antigen specific engagement.

Geometric mean fluorescent intensity (GMFI) of 4-1BB, TIM3 and PD1 for each tumour TRM subset; representative histograms shown (n = 3).

To confirm the co-expression of these markers of interest, we completed two further sets of analysis. Firstly, we generated a tSNE visualisation of FACS data, taking equivalent numbers of cells from each patient (down sampled to smallest sample size) and merged these samples into a singular dataset. We confirmed that the presence of $TIM3^+$, $4\text{-}1BB^+$, $CD39^+$, $IL7R^-$, $PD1^{high}$ was restricted to a subset of cells localised into the same location in the tSNE plot, which was a cluster dominated by cells isolated from the tumour (second box from the left). Additionally, this confirmed that the phenotype detected at the single cell transcriptomic level was indeed reflected at the protein level (Figure 56). We did not complete co-expression analysis in the other subsets at we did not readily detect CD39 and/or TIM3 expression in them. Additionally, we completed a set of single cell RNA-seq using Smart-seq2, through which we completed co-expression analysis, showing that markers of activation were co-expressed with PD1 in T_{RM} cells but not in the non-T_{RM} compartment. This suggested that although Tumour T_{RM} appear to express the highest levels of

PD1 at both the transcriptomic and protein levels, they do not have conventional features of exhaustion (Figure 57) and instead PD1 seems to be reflective of superior functionality.

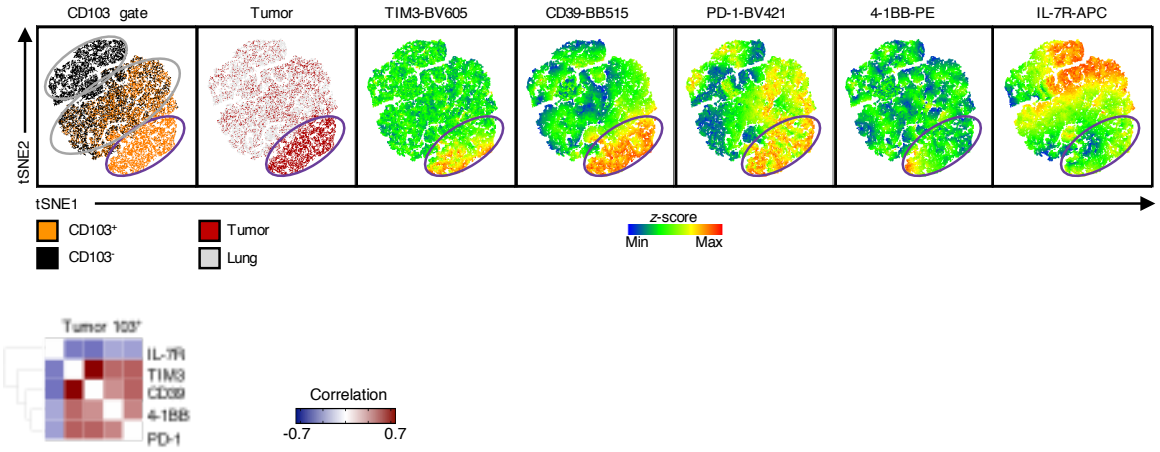


Figure 56. Features predicted from single cell RNA-seq are co-expressed at the protein level. (A) TSNE visualisation of 3,000 randomly selected cells from total $CD8^+$ cells isolated from the tumour and the lung. Each dot represents a cell and the colour is reflective of the Z-score normalised expression of the protein (above). (B) Spearman co-expression analysis of Tumour $CD103^+$ cells showing co-expression of noted proteins, samples are clustered with complete linkage analysis.

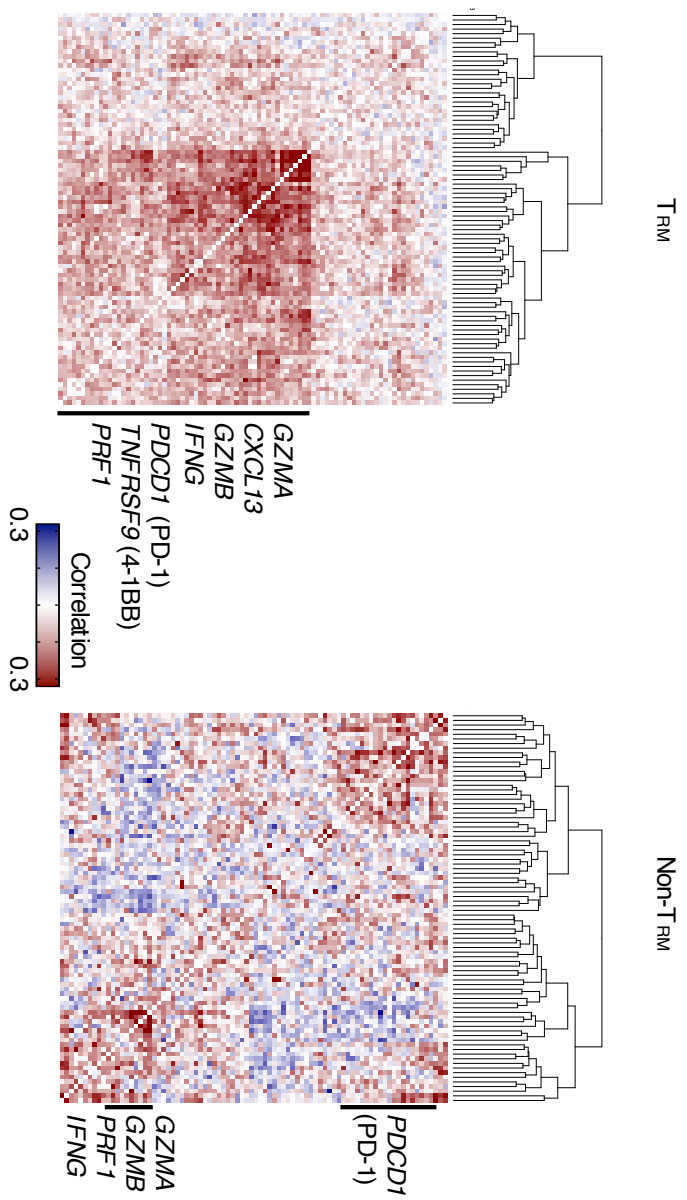


Figure 57. Co-expression of features specific to effector T_{RM} are co-expressed in T_{RM} but not in non-T_{RM} cells.

Spearman co-expression analysis of 500 and 167 single cell transcriptomes analysed by Smart-seq2. Each column and row is clustered with complete linkage analysis. Note, in the T_{RM} *PDCD1* expression correlates with effector functions which is not found in the $CD103^-$ cells.

6.3 Discussion.

Single cell analysis of almost 9000 cells demonstrated that even at the single cell level, markers of tissue residency and antigen-specific activation were co-expressed and dominated the majority of the variation found across the $CD8^+$ cells. We identified 8 unique clusters of which 4 were highly represented by $CD103^+$ cells. This emphasized the importance of sorting on $CD103^+$ to identify features of tissue residency. From this we concluded that these two populations must represent Tumour-specific subtypes. As we had identified a module that contained *IL7R, CXCR4, FOS, KLF6* that was the most strikingly correlated to the level of immune infiltrate, we concluded that our analysis at population and single cell levels had identified the same population through complementary approaches. We believe this to be one of the first demonstrations of the use of purified population network analysis to predict single cell phenotypic changes. Given the increasing availability of transcriptomic approaches, we foresee this process being useful across disease and cell type applications.

Further characterization demonstrated that these effector T_{RM} are found in other solid tumours and were also enriched for markers of antigen-specific engagement in HNSCC; the HNSCC cohort will need to be expanded to reach statistical significance. Future work may identify shared clonotypes between each cluster which could be demonstrative of the $IL7R^+$ cells being precursors as we have shown previously in CD4-cytotoxic lymphocytes (Patil et al., 2018). Further studies will help delineate the dynamics that cause the development of effector or intermediate phenotype T_{RM} and their role in long-term survival. This will become clearer particularly as survival datasets mature with further immunotherapy trials ongoing. Alternatively, *in-vivo* models may help facilitate the temporal dynamics of each of these populations.

The transcriptomic similarities between the cycling cluster and the effector cluster led us to suggest that this population appears to be a key population responding to antigen-specific engagement with clonal expansion. Interestingly the genes that were found to be differentially expressed in the effector cluster includes genes encoding proteins reflective of superior functionality such as GZMA, IFNG, CCL3 and CXCL13. It was particularly noteworthy that PD1 expression was highest in this cluster alongside the expression of cytoxicity and effector molecules traditionally thought to be downregulated in $PD1^+$, 'exhausted' cells. This suggests that PD1 expression in T_{RM} cells may not represent an exhaustion molecular signature, like in T_{FH} cells, and may instead represent a molecule preventing apoptosis following antigen-specific engagement. This population which has not been described in previous studies of murine TILS or murine T_{RM} cells, appears to be proliferating actively and has the highest PD1 expression, suggesting this is the population that responds following PD1 therapy. Discussion.

6.4 Final discussion.

This thesis started with an initial question about the stability and variability of the microscopic and molecular pathological features in cancer tissue samples. Given that we wanted to work in human tissue, our starting point was to investigate the consistency of immune infiltrates in replicate tumour biopsies and resection samples within and between patients. For practical reasons (feasibility of repeated sampling) this question was addressed in patients with HNSCC. Using variance filtering of the top 4000 most variable genes (Engel et al., 2016), we maximised the chance to detect differences, which we then used test correlation, to undertake Euclidean distance analysis and evaluation by principal component analysis (Figure 10). The power of these molecular technologies to investigate gene expression is remarkable, as it is possible to compare the expression of every gene in contrast to more traditionally methodologies which is protein by protein as with Immunohistochemistry (Figure 11). We found a high degree of immunological similarity within a given patient's samples, compared to the samples from other patients in the cohort. In this study two methodologies were used, the correlation and the Euclidean distance to demonstrate this (dis)similarity (Figure 14).

This finding was unexpected in light of the heterogeneity found at the level of the genome within tumours (Ledgerwood et al., 2016), but clinically is important and reassuring for treatment choices in the individual patient. It will be interesting to follow up with exome sequencing analysis on the sections consecutive to those used for the transcriptome analysis to identify if these patients have an expected level of mutational burden. These molecular analyses are ongoing in our laboratory. It was reassuring that the gene and protein expression were tightly correlated. It was reassuring that the gene and protein expression were tightly correlated (Figure 15) as the following chapter was protein validation of gene expression profiles. This work contributed to a manuscript in which, when investigating known genes of relevance for cancer immunotherapy (such as CTLA4), a high degree of similarity was identified in multiple samples from the same patient, even when taken at different timepoints. We concluded that bulk molecular profiles allow a reliable assessment of a patient's unique immune signature. However, these bulk transcriptomic profiles do not investigate expression profiles of individual subpopulations. The analysis reads out the total transcriptome of all cellular populations present in the tissue; therefore, if a gene of interest is expressed in multiple population such as in both $CD4^+$ and $CD8^+$ cells, an assessment of the contribution of individual immune cell types is not possible.

This could only be resolve by single cell or purified population transcriptomic analysis. In a pilot study, we found that it was possible to recover gene expression profiles and the TCR from single cells (Figure 16). Another member of our research group had found transcriptomic evidence

of T_{RM} cells in the $CD8^+$ TILS. The next stage was to validate these markers at the protein level and the data (lack of expression of CCR7, CD62L, KLRG1 (Figure 20 and Figure 21) and expression of CD49A and CD69, were consistent with the conclusion that the $CD103^+CD8^+$ T cells were indeed T_{RM} cells. Furthermore, we found interesting potential immune-oncological targets, for example of CD38 and CD39, which were consistently found at the protein level and by RNA analysis. CD39 has recently been shown to be a marker of human tumour-specific TILS, compared to 'bystander' TILS (Simoni et al., 2018). However, this analysis of particular markers by flow cytometry was limited to a set of proteins that were selected a priori, reflecting the limited number of fresh tissue samples that were available. Therefore, to understand the expression profiles at the genome wide level we moved to population analysis of the $CD103^+$ and $CD103^-$ cells in the tumour and human lung.

Given that sorting on CD103 enriched for a tissue resident phenotype, we next aimed to detect the signature of tissue resident cells at a greater resolution using genome-wide molecular approaches and to test, which features were found only in tumour T_{RM} compared to tumour-free lung from the same patient. To this end, we undertook RNA-seq on populations of $CD103^+$ and $CD103^-$ cells. The enrichment for markers of tissue residency by sorting on CD103 was important, as it resolved previous controversies in the literature: CD69 had been reported as a T_{RM} marker but we were able to detect expression in both T_{RM} and non-T_{RM} cells (Figure 31-Figure 35).

We found clonally expanded T_{RM}, and, consistent with cell division, observed overexpression of cell cycle signatures. The data suggest that this population is undergoing antigen-specific TCR engagement (likely with tumour-related (neo-)antigens). Other signatures included those reflective of immune cell activation (Figure 39-Figure 42). Intriguingly, this latter feature was enriched in T_{RM}-high tumours, but in individual patients in both T_{RM} and nonT_{RM} $CD8^+$ TILS. These data suggest that conditioning occurs in the tumour microenvironment and we intend to examine in future work whether T_{RM} shape the non-T_{RM} or whether both cell types follow the same cues in good prognosis cancers.

We extended our study of $CD8^+$ T_{RM} using co-expression analysis, as it can provide extra insight into a particular population. This analysis identified several distinct transcriptomic modules in the data, suggestive of multiple T_{RM} cell subsets (Figure 43). This was confirmed by the single cell analysis which identified several subpopulations of tumour T_{RM} cells alongside subsets of non-T_{RM} cells (Figure 46-Figure 50).

Single cell analysis strongly highlighted that the genes identified from the purified population analysis were enriched at the single cell level. Moreover, even at the single cell level, there was

distinct co-expression of 'exhaustion' markers, T_{RM} features and genes associated with cytotoxicity (Figure 57). This further suggests that PD1 is not a feature of exhaustion in T_{RM}. The cluster of T_{RM} with effector phenotype was enriched in TIL^{high} patients; this led us to postulate that induction and expansion of these cells must be an aim in immunotherapy.

Future work will focus on several key areas: the first will be improving our understanding of populations T_{RM} ($IL7R^{+}$, Intermediate and Effector) and comparing and contrasting these with both the $CD103^{-}$ population, but also the $PD1^{+}$ $CD103^{-}$ to confirm whether these cells have a functionally exhausted phenotype or not. There are several methodologies by which to interrogate these populations: one is traditional flow cytometry of markers of activation and cytotoxicity in each population, another would be population RNA-seq following *ex-vivo* stimulation to dissect how each population responds; a further option could be ATAC-seq which describes chromatin accessibility and may allow us to identify features that underpin cell behaviour. An additional step would be to interrogate these cells in the context of an immunotherapy regime such as anti-PD1 therapy, where patients have responded or not. The presence or absence of a clonally expanded 'effector' T_{RM} phenotype may provide an insight into the role of this population. We also need to confirm if the presence of particular subtypes is linked to greater patient survival. Finally, given we have suggested that T_{RM} in the non-malignant tissue also contain $PD1^{+}$ cells, it would be interesting to study tissue biopsies from patients with immune related adverse effects, such as colitis, to understand if these organs/tissues contain activated, clonally expanded T_{RM} populations.

A striking observation from recent publications has been the identification of a stem-cell like precursor cell, that is activated by immune checkpoint blockade in the murine LCMV-13 system from Rafi Ahmed's group. It is thought that this *TCF7* expressing (TCF1) population provides a pool of T cells from which the anti-tumour immunity is formed. Although these cells are clearly the antigen specific cells that are activated in this viral system, in the human system, it is clear that the class I neo-antigen specific T cells appear to be these $PD1^{high}$ $CD39^{+}$ $CD103^{+}$ $CD8^{+}$ population (Simoni, et al., 2018). Several studies have now highlighted that the expression of PD1 in T_{RM} cells is required for their survival in tissue, and that the expression of PD1 is maintained in an antigen-independent manner (Shwetank et al., 2017). This highlights a number of unresolved questions, firstly are $PD1^{high}$ T_{RM} cells exhausted? These cells appear to have the highest expression of cytotoxicity associated transcripts, but this analysis may be biased by the enrichment for antigen specific cells in this population. The analysis presented here demonstrating co-expression of transcripts encoding PD1 and cytotoxicity molecules would argue these cells are not exhausted and instead the molecular programs of T_{RM} activation and functionality in the context of PD1 expression are distinct from those studied in the murine system. Further given these cells are the antigen specific cells and are those expressing the highest amount of PD1 it is likely these cells may be those responding to anti-PD1

blockade. Further studies finely interrogating PD1 expression in T_{RM} and non-T_{RM} cells are likely to add clarity to this question. Further where T_{RM} cells fit into this lineage of exhausted terminally differentiated cells and these stem-cell like $CXCR5^{+}$ precursors is unknown.

To conclude, the original focus point of this thesis was to understand inter- and intra-patient immune heterogeneity. At a whole tumour level, we identified that each patient has their own molecular immune profile, although patterns began to emerge, which were shared between patients. We observed significant heterogeneity within cell types both between patients and within individual patients (Chapter 3). In highly CD8 T cell infiltrated tumours we found an enrichment for T_{RM}, which we further characterised by flow cytometry (Chapter 4) and genome-wide profiling. We identified a number of specific gene modules that correlated to immune infiltrate levels (Chapter 5). Using single cell analysis, we were able to characterise a novel subpopulation of T-T_{RM} (Chapter 6) that appears to escape exhaustion. In spite of their high expression of PD1, this populations appears to have superior functional properties and suggests that the perception of PD1 as an exhaustion marker requires re-evaluation.

List of References.

Abbey, J.L., and O'Neill, H.C. (2008). Expression of T-cell receptor genes during early T-cell development. Immunol. Cell Biol. *86*, 166–174.

Alarcón, B., Mestre, D., and Martínez-Martín, N. (2011). The immunological synapse: a cause or consequence of T-cell receptor triggering? Immunology *133*, 420–425.

Alizadeh, A.A., Aranda, V., Bardelli, A., Blanpain, C., Bock, C., Borowski, C., Caldas, C., Califano, A., Doherty, M., Elsner, M., et al. (2015). Toward understanding and exploiting tumor heterogeneity. Nat. Med. *21*, 846–853.

Amir, E.D., Davis, K.L., Tadmor, M.D., Simonds, E.F., Levine, J.H., Bendall, S.C., Shenfeld, D.K., Krishnaswamy, S., Nolan, G.P., and Pe'er, D. (2013). viSNE enables visualization of high dimensional single-cell data and reveals phenotypic heterogeneity of leukemia. Nat. Biotechnol. *31*, 545–552.

Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. Genome Biol. *11*, R106.

Anderson, A.C., Anderson, D.E., Bregoli, L., Hastings, W.D., Kassam, N., Lei, C., Chandwaskar, R., Karman, J., Su, E.W., Hirashima, M., et al. (2007). Promotion of Tissue Inflammation by the Immune Receptor Tim-3 Expressed on Innate Immune Cells. Science. *318*, 1141–1143.

Anderson, A.C., Joller, N., and Kuchroo, V.K. (2016). Lag-3, Tim-3, and TIGIT: Co-inhibitory Receptors with Specialized Functions in Immune Regulation. Immunity *44*, 989–1004.

Ariotti, S., Beltman, J.B., Chodaczek, G., Hoekstra, M.E., van Beek, A.E., Gomez-Eerland, R., Ritsma, L., van Rheenen, J., Maree, A.F.M., Zal, T., et al. (2012). Tissue-resident memory CD8+ T cells continuously patrol skin epithelia to quickly recognize local antigen. Proc. Natl. Acad. Sci. *109*, 19739–19744.

Arsenio, J., Kakaradov, B., Metz, P.J., Kim, S.H., Yeo, G.W., and Chang, J.T. (2014). Early specification of CD8+ T lymphocyte fates during adaptive immunity revealed by single-cell gene-expression analyses. Nat. Immunol. *15*, 365–372.

Arsenio, J., Metz, P.J., and Chang, J.T. (2015). Asymmetric Cell Division in T Lymphocyte Fate Diversification. Trends Immunol. *36*, 670–683.

Attaf, M., Huseby, E., and Sewell, A.K. (2015a). αβ T cell receptors as predictors of health and disease. Cell. Mol. Immunol. 1–9.

Attaf, M., Legut, M., Cole, D.K., and Sewell, A.K. (2015b). The T cell antigen receptor: The Swiss Army knife of the immune system. Clin. Exp. Immunol. *181*, 1–18.

Bacher, P., Heinrich, F., Stervbo, U., Nienen, M., Vahldieck, M., Iwert, C., Vogt, K., Kollet, J., Babel, N., Sawitzki, B., et al. (2016). Regulatory T cell Specificity Directs Tolerance versus Allergy against Aeroantigens in Humans. Cell.

Banchereau, R., Hong, S., Cantarel, B., Baldwin, N., Baisch, J., Edens, M., Cepika, A.-M., Acs, P., Turner, J., Anguiano, E., et al. (2016). Personalized Immunomonitoring Uncovers Molecular Networks that Stratify Lupus Patients. Cell *165*, 551–565.

Bardou, P., Mariette, J., Escudié, F., Djemiel, C., and Klopp, C. (2014). jvenn: an interactive Venn diagram viewer. BMC Bioinformatics *15*, 293.

Bendall, S.C., Simonds, E.F., Qiu, P., Amir, E.D., Krutzik, P.O., Finck, R., Bruggner, R. V, Melamed, R., Trejo, A., Ornatsky, O.I., et al. (2011). Single-cell mass cytometry of differential immune and drug responses across a human hematopoietic continuum. Science *332*, 687–696.

Bergsbaken, T., Bevan, M.J., and Fink, P.J. (2017). Local Inflammatory Cues Regulate Differentiation and Persistence of CD8+ Tissue-Resident Memory T cells. Cell Rep. *19*, 114–124.

Best, K., Oakes, T., Heather, J.M., Shawe-Taylor, J., and Chain, B. (2015). Computational analysis of stochastic heterogeneity in PCR amplification efficiency revealed by single molecule barcoding. Sci. Rep. *5*, 14629.

Bhati, M., Cole, D.K., McCluskey, J., Sewell, A.K., and Rossjohn, J. (2014). The versatility of the αβ T-cell antigen receptor. Protein Sci. *23*, 260–272.

Bindea, G., Mlecnik, B., Hackl, H., Charoentong, P., Tosolini, M., Kirilovsky, A., Fridman, W.-H., Pagès, F., Trajanoski, Z., and Galon, J. (2009). ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. Bioinformatics *25*, 1091–1093.

Bindea, G., Mlecnik, B., Tosolini, M., Kirilovsky, A., Waldner, M., Obenauf, A.C., Angell, H., Fredriksen, T., Lafontaine, L., Berger, A., et al. (2013). Spatiotemporal dynamics of intratumoral immune cells reveal the immune landscape in human cancer. Immunity *39*, 782–795.

Birnbaum, M.E., Mendoza, J.L., Sethi, D.K., Dong, S., Glanville, J., Dobbins, J., Ozkan, E., Davis, M.M., Wucherpfennig, K.W., and Garcia, K.C. (2014). Deconstructing the peptide-MHC specificity of T cell recognition. Cell *157*, 1073–1087.

Bluestone, J.A., Mackay, C.R., O'Shea, J.J., and Stockinger, B. (2009). The functional plasticity of T cell subsets. Nat. Rev. Immunol. *9*, 811–816.

Borg, N.A., Wun, K.S., Kjer-Nielsen, L., Wilce, M.C.J., Pellicci, D.G., Koh, R., Besra, G.S., Bharadwaj, M., Godfrey, D.I., McCluskey, J., et al. (2007). CD1d–lipid-antigen recognition by the semi-invariant NKT T-cell receptor. Nature *448*, 44–49.

Bowen, S., Sun, P., Livak, F., Sharrow, S., and Hodes, R.J. (2014). A novel T cell subset with trans-rearranged Vγ-Cβ TCRs shows Vβ expression is dispensable for lineage choice and MHC restriction. J. Immunol. *192*, 169–177.

Braggio, E., Egan, J.B., Fonseca, R., and Stewart, A.K. (2013). Lessons from next-generation sequencing analysis in hematological malignancies. *3*, e127-10.

Breen, M.S., Beliakova-Bethell, N., Mujica-Parodi, L.R., Carlson, J.M., Ensign, W.Y., Woelk, C.H., and Rana, B.K. (2015a). Acute psychological stress induces short-term variable immune response. Brain. Behav. Immun.

Breen, M.S., Maihofer, A.X., Glatt, S.J., Tylee, D.S., Chandler, S.D., Tsuang, M.T., Risbrough, V.B., Baker, D.G., O'Connor, D.T., Nievergelt, C.M., et al. (2015b). Gene networks specific for innate immunity define post-traumatic stress disorder. Mol. Psychiatry.

Brennecke, P., Anders, S., Kim, J.K., Kołodziejczyk, A.A., Zhang, X., Proserpio, V., Baying, B., Benes, V., Teichmann, S.A., Marioni, J.C., et al. (2013). Accounting for technical noise in single-cell RNA-seq experiments. Nat. Methods *10*, 1093–1095.

Brownlie, R.J., and Zamoyska, R. (2013). T cell receptor signalling networks: branched, diversified and bounded. Nat. Rev. Immunol. *13*, 257–269.

de Bruin, E.C., McGranahan, N., Mitter, R., Salm, M., Wedge, D.C., Yates, L., Jamal-Hanjani, M., Shafi, S., Murugaesu, N., Rowan, A.J., et al. (2014). Spatial and temporal diversity in genomic instability processes defines lung cancer evolution. Science. *346*, 251–256.

Buchan, S.L., Fallatah, M., Thirdborough, S.M., Taraban, V.Y., Rogel, A., Thomas, L.J., Penfold, C.A., He, L.-Z., Curran, M.A., Keler, T., et al. (2018). PD-1 blockade and CD27 stimulation activate distinct transcriptional programs that synergize for CD8+ T-cell driven anti-tumor immunity. Clin. Cancer Res. clincanres.3057.2017.

Capone, M., Romagnoli, P., Beermann, F., MacDonald, H.R., and van Meerwijk, J.P. (2001). Dissociation of thymic positive and negative selection in transgenic mice expressing major histocompatibility complex class I molecules exclusively on thymic cortical epithelial cells. Blood *97*, 1336–1342.

Carding, S.R., and Egan, P.J. (2002). Gammadelta T cells: functional plasticity and heterogeneity. Nat. Rev. Immunol. *2*, 336–345.

Cha, E., Klinger, M., Hou, Y., Cummings, C., Ribas, A., Faham, M., and Fong, L. (2014). Improved survival with T cell clonotype stability after anti-CTLA-4 treatment in cancer patients. Sci. Transl. Med. *6*, 238ra70.

Chan, C.J., Martinet, L., Gilfillan, S., Souza-Fonseca-Guimaraes, F., Chow, M.T., Town, L., Ritchie, D.S., Colonna, M., Andrews, D.M., and Smyth, M.J. (2014). The receptors CD96 and CD226 oppose each other in the regulation of natural killer cell functions. Nat. Immunol. *15*, 431–438.

Chang, J.T., Palanivel, V.R., Kinjyo, I., Schambach, F., Intlekofer, A.M., Banerjee, A., Longworth, S.A., Vinup, K.E., Mrass, P., Oliaro, J., et al. (2007). Asymmetric T Lymphocyte Division in the Initiation of Adaptive Immune Responses. Science. *315*, 1687–1691.

Chang, J.T., Wherry, E.J., and Goldrath, A.W. (2014). Molecular regulation of effector and memory T cell differentiation. Nat. Immunol. *15*, 1104–1115.

Chee, S.J., Lopez, M., Mellows, T., Gankande, S., Moutasim, K.A., Harris, S., Clarke, J., Vijayanand, P., Thomas, G.J., and Ottensmeier, C.H. (2017). Evaluating the effect of immune cells on the outcome of patients with mesothelioma.

Chen, L., and Flies, D.B. (2013). Molecular mechanisms of T cell co-stimulation and co-inhibition. Nat. Rev. Immunol. *13*, 227–242.

Cheuk, S., Schlums, H., Gallais Sérézal, I., Martini, E., Chiang, S.C., Marquardt, N., Gibbs, A., Detlofsson, E., Introini, A., Forkel, M., et al. (2017). CD49a Expression Defines Tissue-Resident CD8(+) T cells Poised for Cytotoxic Function in Human Skin. Immunity *46*, 287–300.

Chevrier, S., Levine, J.H., Zanotelli, V.R.T., Silina, K., Schulz, D., Bacac, M., Ries, C.H., Ailles, L., Jewett, M.A.S., Moch, H., et al. (2017). An Immune Atlas of Clear Cell Renal Cell Carcinoma. Cell *169*, 736–749.e18.

Coghill, J.M., Sarantopoulos, S., Moran, T.P., Murphy, W.J., Blazar, B.R., and Serody, J.S. (2011). Effector CD4+ T cells, the cytokines they generate, and GVHD: something old and something new. Blood *117*, 3268–3276.

Collins, S., Waickman, A., Basson, A., Kupfer, A., Licht, J.D., Horton, M.R., and Powell, J.D. (2012). Regulation of CD4+ and CD8+ Effector Responses by Sprouty-1. PLoS One *7*, e49801.

Corces, M.R., Trevino, A.E., Hamilton, E.G., Greenside, P.G., Sinnott-Armstrong, N.A., Vesuna, S., Satpathy, A.T., Rubin, A.J., Montine, K.S., Wu, B., et al. (2017). An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. Nat. Methods *14*, 959–962.

Couzin-Frankel, J. (2013). Breakthrough of the year 2013. Cancer immunotherapy. Science *342*, 1432–1433.

Croft, M. (2009). The role of TNF superfamily members in T-cell function and diseases. Nat. Rev. Immunol. *9*, 271–285.

Croft, M., So, T., Duan, W., and Soroosh, P. (2009). The significance of OX40 and OX40L to T-cell biology and immune disease. Immunol. Rev. *229*, 173–191.

Crotty, S. (2014). T Follicular Helper Cell Differentiation, Function, and Roles in Disease. Immunity *41*, 529–542.

Cui, W., Liu, Y., Weinstein, J.S., Craft, J., and Kaech, S.M. (2011). An interleukin-21-interleukin-10-STAT3 pathway is critical for functional maturation of memory CD8+ T cells. Immunity *35*, 792–805.

D'Asaro, M., Dieli, F., Caccamo, N., Musso, M., Porretto, F., and Salerno, A. (2006). Increase of CCR7− CD45RA+ CD8 T cells (TEMRA) in chronic graft-versus-host disease. Leukemia *20*, 545–547.

D'haeseleer, P. (2005). How does gene expression clustering work? Nat. Biotechnol. *23*, 1499–1501.

Davis, M.M. (2012). Immunology taught by humans. Sci. Transl. Med. *4*, 117fs2.

Davis, M.M., Altman, J.D., and Newell, E.W. (2011). Interrogating the repertoire: broadening the scope of peptide-MHC multimer analysis. Nat. Rev. Immunol. *11*, 551–558.

Deluca, L.S., and Gommerman, J.L. (2012). Fine-tuning of dendritic cell biology by the TNF superfamily. Nat. Rev. Immunol. *12*, 339–351.

Dillies, M.-A., Rau, A., Aubert, J., Hennequet-Antier, C., Jeanmougin, M., Servant, N., Keime, C., Marot, G., Castel, D., Estelle, J., et al. (2013). A comprehensive evaluation of normalization methods for Illumina high-throughput RNA sequencing data analysis. Brief. Bioinform. *14*, 671–683.

Djenidi, F., Adam, J., Goubar, A., Durgeau, A., Meurice, G., de Montpréville, V., Validire, P., Besse, B., and Mami-Chouaib, F. (2015). CD8+CD103+ tumor-infiltrating lymphocytes are tumor-specific tissue-resident memory T cells and a prognostic factor for survival in lung cancer patients. J. Immunol. *194*, 3475–3486.

Dolton, G., Tungatt, K., Lloyd, A., Bianchi, V., Theaker, S.M., Trimby, A., Holland, C.J., Donia, M., Godkin, A.J., Cole, D.K., et al. (2015). More Tricks with Tetramers: A Practical Guide to Staining T cells with peptide-MHC Multimers. Immunology n/a-n/a.

Draghiciu, O., Lubbers, J., Nijman, H.W., and Daemen, T. (2015). Myeloid derived suppressor cells-An overview of combat strategies to increase immunotherapy efficacy. Oncoimmunology *4*, e954829.

Drake, C.G. (2010). Prostate cancer as a model for tumour immunotherapy. Nat. Rev. Immunol. *10*, 580–593.

Dunn, G.P., Bruce, A.T., Ikeda, H., Old, L.J., and Schreiber, R.D. (2002). Cancer immunoediting: from immunosurveillance to tumor escape. Nat. Immunol. *3*, 991–998.

DuPage, M., and Bluestone, J.A. (2016). Harnessing the plasticity of CD4+ T cells to treat immune-mediated disease. Nat. Rev. Immunol.

Elliott, T., and Neefjes, J. (2006). The complex route to MHC class I-peptide complexes. Cell *127*, 249–251.

Engel, I., Seumois, G., Chavez, L., Samaniego-Castruita, D., White, B., Chawla, A., Mock, D., Vijayanand, P., and Kronenberg, M. (2016). Innate-like functions of natural killer T cell subsets result from highly divergent gene programs. Nat. Immunol. *17*, 728–739.

Faherty, S.L., Campbell, C.R., Larsen, P.A., and Yoder, A.D. (2015). Evaluating whole transcriptome amplification for gene profiling experiments using RNA-Seq. BMC Biotechnol. *15*, 65.

Fan, X., and Rudensky, A.Y. (2016). Hallmarks of Tissue-Resident Lymphocytes. Cell *164*, 1198–1211.

Fan, J., Salathia, N., Liu, R., Kaeser, G.E., Yung, Y.C., Herman, J.L., Kaper, F., Fan, J.-B., Zhang, K., Chun, J., et al. (2016). Characterizing transcriptional heterogeneity through pathway and gene set overdispersion analysis. Nat. Methods *13*, 241–244.

Farber, D.L., Yudanin, N.A., and Restifo, N.P. (2014). Human memory T cells: generation, compartmentalization and homeostasis. Nat. Rev. Immunol. *14*, 24–35.

Finkin, S., Yuan, D., Stein, I., Taniguchi, K., Weber, A., Unger, K., Browning, J.L., Goossens, N., Nakagawa, S., Gunasekaran, G., et al. (2015). Ectopic lymphoid structures function as microniches for tumor progenitor cells in hepatocellular carcinoma. Nat. Immunol.

Föhse, L., Suffner, J., Suhre, K., Wahl, B., Lindner, C., Lee, C.-W., Schmitz, S., Haas, J.D., Lamprecht, S., Koenecke, C., et al. (2011). High TCR diversity ensures optimal function and homeostasis of Foxp3+ regulatory T cells. Eur. J. Immunol. *41*, 3101–3113.

Förster, R., Davalos-Misslitz, A.C., and Rot, A. (2008). CCR7 and its ligands: balancing immunity and tolerance. Nat. Rev. Immunol. *8*, 362–371.

Fourcade, J., Sun, Z., Benallaoua, M., Guillaume, P., Luescher, I.F., Sander, C., Kirkwood, J.M., Kuchroo, V., and Zarour, H.M. (2010). Upregulation of Tim-3 and PD-1 expression is associated with tumor antigen–specific CD8 + T cell dysfunction in melanoma patients. J. Exp. Med. *207*, 2175–2186.

Fu, G., Casas, J., Rigaud, S., Rybakin, V., Lambolez, F., Brzostek, J., Hoerter, J.A.H., Paster, W., Acuto, O., Cheroutre, H., et al. (2013). Themis sets the signal threshold for positive and negative selection in T-cell development. Nature *504*, 441–445.

Gajewski, T.F., Schreiber, H., and Fu, Y.-X. (2013). Innate and adaptive immune cells in the tumor microenvironment. Nat. Immunol. *14*, 1014–1022.

Gallego Romero, I., Pai, A.A., Tung, J., and Gilad, Y. (2014). RNA-seq: impact of RNA degradation on transcript quantification. BMC Biol. *12*, 42.

Galon, J., Costes, A., Sanchez-Cabo, F., Kirilovsky, A., Mlecnik, B., Lagorce-Pagès, C., Tosolini, M., Camus, M., Berger, A., Wind, P., et al. (2006). Type, density, and location of immune cells within human colorectal tumors predict clinical outcome. Science. *313*, 1960–1964.

Ganesan, A.-P., Clarke, J., Wood, O., Garrido-Martin, E.M., Chee, S.J., Mellows, T., Samaniego-Castruita, D., Singh, D., Seumois, G., Alzetani, A., et al. (2017). Tissue-resident memory features are linked to the magnitude of cytotoxic T cell responses in human lung cancer. Nat. Immunol. *18*, 940–950.

Garten, A., Petzold, S., Körner, A., Imai, S.-I., and Kiess, W. (2009). Nampt: linking NAD biology, metabolism and cancer. Trends Endocrinol. Metab. *20*, 130–138.

Gattinoni, L., Lugli, E., Ji, Y., Pos, Z., Paulos, C.M., Quigley, M.F., Almeida, J.R., Gostick, E., Yu, Z., Carpenito, C., et al. (2011). A human memory T cell subset with stem cell–like properties. Nat. Med. *17*, 1290–1297.

Gaublomme, J.T., Yosef, N., Lee, Y., Gertner, R.S., Yang, L.V., Wu, C., Pandolfi, P.P., Mak, T., Satija, R., Shalek, A.K., et al. (2015). Single-Cell Genomics Unveils Critical Regulators of Th17 Cell Pathogenicity. Cell.

Gaudillière, B., Fragiadakis, G.K., Bruggner, R. V, Nicolau, M., Finck, R., Tingle, M., Silva, J., Ganio, E.A., Yeh, C.G., Maloney, W.J., et al. (2014). Clinical recovery from surgery correlates with single-cell immune signatures. Sci. Transl. Med. *6*, 255ra131.

Gerlach, C., van Heijst, J.W.J., Swart, E., Sie, D., Armstrong, N., Kerkhoven, R.M., Zehn, D., Bevan, M.J., Schepers, K., and Schumacher, T.N.M. (2010). One naive T cell, multiple fates in CD8+ T cell differentiation. J. Exp. Med. *207*, 1235–1246.

Gierliński, M., Cole, C., Schofield, P., Schurch, N.J., Sherstnev, A., Singh, V., Wrobel, N., Gharbi, K., Simpson, G., Owen-Hughes, T., et al. (2015). Statistical models for RNA-seq data derived from a two-condition 48-replicate experiment. Bioinformatics *31*, 3625–3630.

Godfrey, D.I., Uldrich, A.P., McCluskey, J., Rossjohn, J., and Moody, D.B. (2015). The burgeoning family of unconventional T cells. Nat. Immunol. *16*, 1114–1123.

Goodwin, S., McPherson, J.D., and McCombie, W.R. (2016). Coming of age: ten years of next-generation sequencing technologies. Nat. Rev. Genet. *17*, 333–351.

Gras, S., Chen, Z., Miles, J.J., Liu, Y.C., Bell, M.J., Sullivan, L.C., Kjer-Nielsen, L., Brennan, R.M., Burrows, J.M., Neller, M.A., et al. (2010). Allelic polymorphism in the T cell receptor and its impact on immune responses. J. Exp. Med. *207*, 1555–1567.

Greiff, V., Miho, E., Menzel, U., and Reddy, S.T. (2015). Bioinformatic and Statistical Analysis of Adaptive Immune Repertoires. Trends Immunol. *36*, 738–749.

Gridelli, C., Rossi, A., Carbone, D.P., Guarize, J., Karachaliou, N., Mok, T., Petrella, F., Spaggiari, L., and Rosell, R. (2015). Non-small-cell lung cancer. Nat. Rev. Dis. Prim. 15009.

Gros, A., Parkhurst, M.R., Tran, E., Pasetto, A., Robbins, P.F., Ilyas, S., Prickett, T.D., Gartner, J.J., Crystal, J.S., Roberts, I.M., et al. (2016). Prospective identification of neoantigen-specific lymphocytes in the peripheral blood of melanoma patients. Nat. Med.

Gupta, P.K., Godec, J., Wolski, D., Adland, E., Yates, K., Pauken, K.E., Cosgrove, C., Ledderose, C., Junger, W.G., Robson, S.C., et al. (2015). CD39 Expression

Identifies Terminally Exhausted CD8+ T cells. PLOS Pathog. *11*, e1005177.

Hammerman, P.S., Lawrence, M.S., Voet, D., Jing, R., Cibulskis, K., Sivachenko, A., Stojanov, P., McKenna, A., Lander, E.S., Gabriel, S., et al. (2012). Comprehensive genomic characterization of squamous cell lung cancers. Nature *489*, 519–525.

Han, A., Glanville, J., Hansmann, L., and Davis, M.M. (2014). Linking T-cell receptor sequence to functional phenotype at the single-cell level. Nat. Biotechnol. *32*, 684–692.

Hanahan, D., and Coussens, L.M. (2012). Accessories to the crime: functions of cells recruited to the tumor microenvironment. Cancer Cell *21*, 309–322.

Hanahan, D., and Weinberg, R.A. (2000). The hallmarks of cancer. Cell *100*, 57–70.

Hanahan, D., and Weinberg, R.A. (2011). Hallmarks of cancer: The next generation. Cell *144*, 646–674.

Hanley, C.J., Noble, F., Ward, M., Bullock, M., Drifka, C., Mellone, M., Manousopoulou, A., Johnston, H.E., Hayden, A., Thirdborough, S., et al. (2015). A subset of myofibroblastic cancer-associated fibroblasts regulate collagen fiber elongation, which is prognostic in multiple cancers. Oncotarget.

Hawse, W.F., De, S., Greenwood, A.I., Nicholson, L.K., Zajicek, J., Kovrigin, E.L., Kranz, D.M., Garcia, K.C., and Baker, B.M. (2014). TCR scanning of peptide/MHC through complementary matching of receptor and ligand molecular flexibility. J. Immunol. *192*, 2885–2891.

Held, K., Beltrán, E., Moser, M., Hohlfeld, R., and Dornmair, K. (2015). T-cell receptor repertoire of human peripheral CD161(hi)TRAV1-2(+) MAIT cells revealed by next generation sequencing and single cell analysis. Hum. Immunol.

Hensley, C.T., Faubert, B., Yuan, Q., Lev-Cohain, N., Jin, E., Kim, J., Jiang, L., Ko, B., Skelton, R., Loudat, L., et al. (2016). Metabolic Heterogeneity in Human Lung Tumors. Cell *164*, 681–694.

Herndler-Brandstetter, D., Ishigame, H., Shinnakasu, R., Plajer, V., Stecher, C., Zhao, J., Lietzenmayer, M., Kroehling, L., Takumi, A., Kometani, K., et al. (2018). KLRG1 + Effector CD8 + T cells Lose KLRG1, Differentiate into All Memory T cell Lineages, and Convey Enhanced Protective Immunity. Immunity 1–14.

Hodi, F.S., O'Day, S.J., McDermott, D.F., Weber, R.W., Sosman, J.A., Haanen, J.B., Gonzalez, R., Robert, C., Schadendorf, D., Hassel, J.C., et al. (2010). Improved Survival with Ipilimumab in Patients with Metastatic Melanoma. N. Engl. J. Med. *363*, 711–723.

Holgate, S.T. (2012). Innate and adaptive immune responses in asthma. Nat. Med. *18*, 673–683.

Holland, S.J., Bartok, I., Attaf, M., Genolet, R., Luescher, I.F., Kotsiou, E., Richard, A., Wang, E., White, M., Coe, D.J., et al. (2012). The T-cell receptor is not hardwired to engage MHC ligands. Proc. Natl. Acad. Sci. U. S. A. *109*, E3111-8.

Hombrink, P., Helbig, C., Backer, R.A., Piet, B., Oja, A.E., Stark, R., Brasser, G., Jongejan, A., Jonkers, R.E., Nota, B., et al. (2016). Programs for the persistence, vigilance and control of human CD8 + lung-resident memory T cells. Nat. Immunol. *17*, 1467–1478.

Honey, K. (2006). CCL3 and CCL4 actively recruit CD8+ T cells. Nat. Rev. Immunol. *6*, 427.

Huang, Y.-H., Zhu, C., Kondo, Y., Anderson, A.C., Gandhi, A., Russell, A., Dougan, S.K., Petersen, B.-S., Melum, E., Pertel, T., et al. (2015). CEACAM1 regulates TIM-3-mediated tolerance and exhaustion. Nature *517*, 386–390.

Iadecola, C., and Anrather, J. (2011). The immunology of stroke: from mechanisms to translation. Nat. Med. *17*, 796–808.

Iijima, N., and Iwasaki, A. (2015). Tissue instruction for migration and retention of TRM cells. Trends Immunol. *36*, 556–564.

Imielinski, M., Berger, A.H., Hammerman, P.S., Hernandez, B., Pugh, T.J., Hodis, E., Cho, J., Suh, J., Capelletti, M., Sivachenko, A., et al. (2012). Mapping the hallmarks of lung adenocarcinoma with massively parallel sequencing. Cell *150*, 1107–1120.

Jabri, B., and Abadie, V. (2015). IL-15 functions as a danger signal to regulate tissue-resident T cells and tissue destruction. Nat. Rev. Immunol. *15*, 771–783.

Lieberman, J. (2010). Granzyme A activates another way to die. Immunol. Rev. 235, 93–104.

Jamal-Hanjani, M., Quezada, S.A., Larkin, J., and Swanton, C. (2015). Translational Implications of Tumor Heterogeneity. Clin. Cancer Res. *21*, 1258–1266.

Joffre, O.P., Segura, E., Savina, A., and Amigorena, S. (2012). Cross-presentation by dendritic cells. Nat. Rev. Immunol. *12*, 557–569.

Johnston, H.E., Carter, M.J., Larrayoz, M., Clarke, J., Garbis, S.D., Oscier, D., Strefford, J.C., Steele, A.J., Walewska, R., and Cragg, M.S. (2018). Proteomics Profiling of CLL Versus Healthy B-cells Identifies Putative Therapeutic Targets and a Subtype-independent Signature of Spliceosome Dysregulation. Mol. Cell. Proteomics *17*, 776–791.

June, C.H., Warshauer, J.T., and Bluestone, J.A. (2017). Is autoimmunity the Achilles' heel of cancer immunotherapy? Nat. Med. *23*, 540–547.

Junttila, M.R., and de Sauvage, F.J. (2013). Influence of tumour micro-environment heterogeneity on therapeutic response. Nature *501*, 346–354.

Kaiserman, D., Bird, C.H., Sun, J., Matthews, A., Ung, K., Whisstock, J.C., Thompson, P.E., Trapani, J.A., and Bird, P.I. (2006). The major human and mouse granzymes are structurally and functionally divergent. J. Cell Biol. 175, 619–630.

Kalos, M., and June, C.H. (2013). Adoptive T cell transfer for cancer immunotherapy in the era of synthetic biology. Immunity *39*, 49–60.

Kaplan, M.H., Hufford, M.M., and Olson, M.R. (2015). The development and in vivo function of T helper 9 cells. Nat. Rev. Immunol. *15*, 295–307.

Kappes, D.J. (2007). CD4 and CD8: Hogging All the Lck. Immunity *27*, 691–693.

Khan, T.N., Mooster, J.L., Kilgore, A.M., Osborn, J.F., and Nolz, J.C. (2016). Local antigen in nonlymphoid tissue promotes resident memory CD8 $^{+}$ T cell formation during viral infection. J. Exp. Med. *213*, 951–966.

Kharchenko, P. V, Silberstein, L., and Scadden, D.T. (2014). Bayesian approach to single-cell differential expression analysis. Nat. Methods *11*, 740–742.

Kim, C.J., Nazli, A., Rojas, O.L., Chege, D., Alidina, Z., Huibner, S., Mujib, S., Benko, E., Kovacs, C., Shin, L.Y.Y., et al. (2012). A role for mucosal IL-22 production and Th22 cells in HIV-associated mucosal immunopathogenesis. Mucosal Immunol. *5*, 670–680.

Kim, H., Kwon, B., and Sin, J.-I. (2013a). Combined stimulation of IL-2 and 4-1BB receptors augments the antitumor activity of E7 DNA vaccines by increasing Ag-specific CTL responses. PLoS One *8*, e83765.

Kim, H.K., Lee, H.Y., Choi, Y.-L., Choi, S.-J., Choi, H., Lee, J., Han, J., Ahn, M.-J., Lee, K.S., and Kim, J. (2014). Assessment of intratumoral heterogeneity of oncogenic driver mutations in surgically-resected lung adenocarcinoma: implications of percutaneous biopsy-based molecular assay for target-directed therapy. Anticancer Res. *34*, 707–714.

Kim, I.-K., Kim, B.-S., Koh, C.-H., Seok, J.-W., Park, J.-S., Shin, K.-S., Bae, E.-A., Lee, G.-E., Jeon, H., Cho, J., et al. (2015). Glucocorticoid-induced tumor necrosis factor receptor-related protein co-stimulation facilitates tumor regression by inducing IL-9-producing helper T cells. Nat. Med.

Kim, S. V., Xiang, W. V., Kwak, C., Yang, Y., Lin, X.W., Ota, M., Sarpel, U., Rifkin, D.B., Xu, R., and Littman, D.R. (2013b). GPR15-Mediated Homing Controls Immune Homeostasis in the Large Intestine Mucosa. Science. *340*, 1456–1459.

King, E. V, Ottensmeier, C.H., and Thomas, G.J. (2014). The immune response in HPV(+) oropharyngeal cancer. Oncoimmunology *3*, e27254.

Kirsch, I.R., Watanabe, R., O'Malley, J.T., Williamson, D.W., Scott, L.-L., Elco, C.P., Teague, J.E., Gehad, A., Lowry, E.L., LeBoeuf, N.R., et al. (2015). TCR sequencing facilitates diagnosis and identifies mature T cells as the cell of origin in CTCL. Sci. Transl. Med. *7*, 308ra158-308ra158.

Klein, A.M., Mazutis, L., Akartuna, I., Tallapragada, N., Veres, A., Li, V., Peshkin, L., Weitz, D.A., and Kirschner, M.W. (2015). Droplet Barcoding for Single-Cell Transcriptomics Applied to Embryonic Stem Cells. Cell *161*, 1187–1201.

Klein, L., Kyewski, B., Allen, P.M., and Hogquist, K.A. (2014). Positive and negative selection of the T cell repertoire: what thymocytes see (and don't see). Nat. Rev. Immunol. *14*, 377–391.

Kniemeyer, O., Brakhage, A.A., Ferreira, F., Wallner, M., and Sawitzki, B. (2016). Regulatory T cell Specificity Directs Tolerance versus Allergy against Aeroantigens in Humans. Cell *167*, 1067–1078.e16.

Kobayashi, K.S., and van den Elsen, P.J. (2012). NLRC5: a key regulator of MHC class I-dependent immune responses. Nat. Rev. Immunol. *12*, 813–820.

Kolodziejczyk, A.A., Kim, J.K., Tsang, J.C.H., Ilicic, T., Henriksson, J., Natarajan, K.N., Tuck, A.C., Gao, X., Bühler, M., Liu, P., et al. (2015). Single Cell RNA-Sequencing of Pluripotent States Unlocks Modular Transcriptional Variation. Cell Stem Cell *17*, 471–485.

Koyama, S., Akbay, E.A., Li, Y.Y., Herter-Sprie, G.S., Buczkowski, K.A., Richards, W.G., Gandhi, L., Redig, A.J., Rodig, S.J., Asahina, H., et al. (2016). Adaptive resistance to therapeutic PD-1 blockade is associated with upregulation of alternative immune checkpoints. Nat Commun *7*, 1–9.

Kreiter, S., Vormehr, M., van de Roemer, N., Diken, M., Löwer, M., Diekmann, J., Boegel, S., Schrörs, B., Vascotto, F., Castle, J.C., et al. (2015). Mutant MHC

class II epitopes drive therapeutic immune responses to cancer. Nature.

Kumar, B. V, Ma, W., Miron, M., Friedman, A.L., Shen, Y., Farber, D.L., Granot, T., Guyer, R.S., Carpenter, D.J., Senda, T., et al. (2017). Human Tissue-Resident Memory T cells Are Defined by Core Transcriptional and Functional Signatures in Lymphoid and Mucosal Sites. Cell Rep. *20*, 2921–2934.

Van Laethem, F., Tikhonova, A.N., Pobezinsky, L.A., Tai, X., Kimura, M.Y., Le Saout, C., Guinter, T.I., Adams, A., Sharrow, S.O., Bernhardt, G., et al. (2013). Lck availability during thymic selection determines the recognition specificity of the T cell repertoire. Cell *154*, 1326–1341.

Langfelder, P., and Horvath, S. (2007). Eigengene networks for studying the relationships between co-expression modules. BMC Syst. Biol. *1*, 54.

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

Langfelder, P., Luo, R., Oldham, M.C., Horvath, S., and Horvath, S. (2011). Is My Network Module Preserved and Reproducible? PLoS Comput. Biol. *7*, e1001057.

Laufer, T.M., DeKoning, J., Markowitz, J.S., Lo, D., and Glimcher, L.H. (1996). Unopposed positive selection and autoreactivity in mice expressing class II MHC only on thymic cortex. Nature *383*, 81–85.

Lawrence, M.S., Sougnez, C., Lichtenstein, L., Cibulskis, K., Lander, E., Gabriel, S.B., Getz, G., Ally, A., Balasundaram, M., Birol, I., et al. (2015). Comprehensive genomic characterization of head and neck squamous cell carcinomas. Nature *517*, 576–582.

Ledgerwood, L.G., Kumar, D., Eterovic, A.K., Wick, J., Zhao, H., Tazi, L., Manna, P., Kerley, S., Joshi, R., Wang, L., et al. (2016). The degree of intratumor mutational heterogeneity varies by primary tumor sub-site. Oncotarget *7*, 27185–27198.

Lee, J.A., Spidlen, J., Boyce, K., Cai, J., Crosbie, N., Dalphin, M., Furlong, J., Gasparetto, M., Goldberg, M., Goralczyk, E.M., et al. (2008). MIFlowCyt: the minimum information about a Flow Cytometry Experiment. Cytometry. A *73*, 926–930.

Li, X., and Zheng, Y. (2015). Regulatory T cell identity: formation and maintenance. Trends Immunol. *36*, 344–353.

Li, X.C., and Turka, L.A. (2010). An update on regulatory T cells in transplant tolerance and rejection. Nat. Rev. Nephrol. *6*, 577–583.

Li, B., Li, T., Pignon, J.-C., Wang, B., Wang, J., Shukla, S. a, Dou, R., Chen, Q., Hodi, F.S., Choueiri, T.K., et al. (2016). Landscape of tumor-infiltrating T cell repertoire of human cancers. Nat. Genet.

Liao, W., Lin, J.-X., and Leonard, W.J. (2013). Interleukin-2 at the crossroads of effector responses, tolerance, and immunotherapy. Immunity *38*, 13–25.

Liu, Q., Kriksunov, I. a, Graeff, R., Munshi, C., Lee, H.C., and Hao, Q. (2005). Crystal structure of human CD38 extracellular domain. Structure *13*, 1331–1339.

Lloyd, C.M., and Hessel, E.M. (2010). Functions of T cells in asthma: more than just T(H)2 cells. Nat. Rev. Immunol. *10*, 838–848.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 550.

Luo, Y., Coskun, V., Liang, A., Yu, J., Cheng, L., Ge, W., Shi, Z., Zhang, K., Li, C., Cui, Y., et al. (2015). Single-cell transcriptome analyses reveal signals to activate dormant neural stem cells. Cell *161*, 1175–1186.

Mackay, L.K., Rahimpour, A., Ma, J.Z., Collins, N., Stock, A.T., Hafon, M.-L., Vega-Ramos, J., Lauzurica, P., Mueller, S.N., Stefanovic, T., et al. (2013). The developmental pathway for CD103+CD8+ tissue-resident memory T cells of skin. Nat. Immunol. *14*, 1294–1301.

Mackay, L.K., Minnich, M., Kragten, N.A.M., Liao, Y., Nota, B., Seillet, C., Zaid, A., Man, K., Preston, S., Freestone, D., et al. (2016). Hobit and Blimp1 instruct a universal transcriptional program of tissue residency in lymphocytes. Science. *352*, 459–463.

Macosko, E.Z., Basu, A., Satija, R., Nemesh, J., Shekhar, K., Goldman, M., Tirosh, I., Bialas, A.R., Kamitaki, N., Martersteck, E.M., et al. (2015). Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. Cell *161*, 1202–1214.

Malik, B.T., Byrne, K.T., Vella, J.L., Zhang, P., Shabaneh, T.B., Steinberg, S.M., Molodtsov, A.K., Bowers, J.S., Angeles, C. V., Paulos, C.M., et al. (2017). Resident memory T cells in the skin mediate durable immunity to melanoma. Sci. Immunol. *2*, 6346.

Man, K., and Kallies, A. (2015). Synchronizing transcriptional control of T cell metabolism and function. Nat. Rev. Immunol. *15*, 574–584.

Martin, N.T., and Martin, M.U. (2016). Interleukin 33 is a guardian of barriers and a local alarmin. Nat. Immunol. *17*, 122–131.

Matsushita, H., Vesely, M.D., Koboldt, D.C., Rickert, C.G., Uppaluri, R., Magrini, V.J., Arthur, C.D., White, J.M., Chen, Y.-S., Shea, L.K., et al. (2012). Cancer exome analysis reveals a T-cell-dependent mechanism of cancer immunoediting. Nature *482*, 400–404.

Mayya, V., and Dustin, M.L. (2016). What Scales the T cell Response? Trends Immunol. *xx*, 1–10.

McCann, K.J., Mander, A., Cazaly, A., Chudley, L., Stasakova, J., Thirdborough, S.M., King, A., Lloyd-Evans, P., Buxton, E., Edwards, C., et al. (2016). Targeting Carcinoembryonic Antigen with DNA Vaccination: On-Target Adverse Events Link with Immunological and Clinical Outcomes. Clin. Cancer Res.

McGranahan, N., Furness, A.J.S., Rosenthal, R., Ramskov, S., Lyngaa, R., Saini, S.K., Jamal-Hanjani, M., Wilson, G.A., Birkbak, N.J., Hiley, C.T., et al. (2016). Clonal neoantigens elicit T cell immunoreactivity and sensitivity to immune checkpoint blockade. Science. aaf1490.

McKinney, E.F., Lee, J.C., Jayne, D.R.W., Lyons, P.A., and Smith, K.G.C. (2015). T-cell exhaustion, co-stimulation and clinical outcome in autoimmunity and infection. Nature *523*, 612–616.

Medzhitov, R., Preston-Hurlburt, P., and Janeway, C.A. (1997). A human homologue of the Drosophila Toll protein signals activation of adaptive immunity. Nature *388*, 394–397.

Mellone, M., Hanley, C.J., Thirdborough, S., Mellows, T., Garcia, E., Woo, J., Tod, J., Frampton, S., Jenei, V., Moutasim, K.A., et al. (2016). Induction of fibroblast senescence generates a non-fibrogenic myofibroblast phenotype that differentially impacts on cancer prognosis. Aging (Albany. NY). *9*, 114–132.

van der Merwe, P.A., and Dushek, O. (2011). Mechanisms for T cell receptor triggering. Nat. Rev. Immunol. *11*, 47–55.

Miller, J.A., Horvath, S., and Geschwind, D.H. (2010). Divergence of human and mouse brain transcriptome highlights Alzheimer disease pathways. Proc. Natl. Acad. Sci. U. S. A. *107*, 12698–12703.

Mills, K.H.G. (2011). TLR-dependent T cell activation in autoimmunity. Nat. Rev. Immunol. *11*, 807–822.

Milner, J.J., Toma, C., Yu, B., Zhang, K., Omilusik, K., Phan, A.T., Wang, D., Getzler, A.J., Nguyen, T., Crotty, S., et al. (2017). Runx3 programs CD8+ T cell residency in non-lymphoid tissues and tumours. Nature 2017.

Mitchell, C.J., Getnet, D., Kim, M.-S., Manda, S.S., Kumar, P., Huang, T.-C., Pinto, S.M., Nirujogi, R.S., Iwasaki, M., Shaw, P.G., et al. (2015). A multi-omic analysis of human naïve CD4+ T cells. BMC Syst. Biol. *9*, 75.

Morgan, R.A., Chinnasamy, N., Abate-Daga, D., Gros, A., Robbins, P.F., Zheng, Z., Dudley, M.E., Feldman, S.A., Yang, J.C., Sherry, R.M., et al. (2013). Cancer Regression and Neurological Toxicity Following Anti-MAGE-A3 TCR Gene Therapy. J. Immunother. *36*, 133–151.

Morris, G.P., and Allen, P.M. (2012). How the TCR balances sensitivity and specificity for the recognition of self and pathogens. Nat. Immunol. *13*, 121–128.

Mucida, D., Husain, M.M., Muroi, S., Wijk, F. Van, Shinnakasu, R., Naoe, Y., Reis, B.S., Huang, Y., Lambolez, F., Docherty, M., et al. (2013). Transcriptional reprogramming of mature CD4 + helper T cells generates distinct MHC class II – restricted cytotoxic T lymphocytes. Nat. Immunol. *14*, 281–289.

Mueller, S.N., and Mackay, L.K. (2015). Tissue-resident memory T cells: local specialists in immune defence. Nat. Rev. Immunol. *16*, 79–89.

Nagarajan, N.A., and Shastri, N. (2013). Immune surveillance for ERAAP dysfunction. Mol. Immunol. *55*, 120–122.

Neefjes, J., Jongsma, M.L.M., Paul, P., and Bakke, O. (2011). Towards a systems understanding of MHC class I and MHC class II antigen presentation. Nat. Rev. Immunol. *11*, 823–836.

Nemazee, D. (2006). Receptor editing in lymphocyte development and central tolerance. Nat. Rev. Immunol. *6*, 728–740.

Newell, E.W., and Davis, M.M. (2014). Beyond model antigens: high-dimensional methods for the analysis of antigen-specific T cells. Nat. Biotechnol. *32*, 149–157.

Nguyen, L.T., and Ohashi, P.S. (2014). Clinical blockade of PD1 and LAG3 — potential mechanisms of action. Nat. Rev. Immunol. *15*, 45–56.

Nguyen, L.T., Yen, P.H., Nie, J., Liadis, N., Ghazarian, D., Al-Habeeb, A., Easson, A., Leong, W., Lipa, J., McCready, D., et al. (2010). Expansion and characterization of human melanoma tumor-infiltrating lymphocytes (TILs). PLoS One *5*, e13940.

Nizard, M., Roussel, H., Diniz, M.O., Karaki, S., Tran, T., Voron, T., Dransart, E., Sandoval, F., Riquet, M., Rance, B., et al. (2017). Induction of resident

memory T cells enhances the efficacy of cancer vaccine. Nat. Commun. *8*, 15221.

Oberg, H.-H., Peipp, M., Kellner, C., Sebens, S., Krause, S., Petrick, D., Adam-Klages, S., Röcken, C., Becker, T., Vogel, I., et al. (2014). Novel bispecific antibodies increase γδ T-cell cytotoxicity against pancreatic cancer cells. Cancer Res. *74*, 1349–1360.

Ohta, T., Sugiyama, M., Hemmi, H., Yamazaki, C., Okura, S., Sasaki, I., Fukuda, Y., Orimo, T., Ishii, K.J., Hoshino, K., et al. (2016). Crucial roles of XCR1-expressing dendritic cells and the XCR1-XCL1 chemokine axis in intestinal immune homeostasis. Sci. Rep. *6*, 23505.

Oldham, M.C., Horvath, S., and Geschwind, D.H. (2006). Conservation and evolution of gene coexpression networks in human and chimpanzee brains. Proc. Natl. Acad. Sci. U. S. A. *103*, 17973–17978.

van Olffen, R.W., Koning, N., van Gisbergen, K.P.J.M., Wensveen, F.M., Hoek, R.M., Boon, L., Hamann, J., van Lier, R.A.W., and Nolte, M.A. (2009). GITR triggering induces expansion of both effector and regulatory CD4+ T cells in vivo. J. Immunol. *182*, 7490–7500.

Otero, K., Vecchi, A., Hirsch, E., Kearley, J., Vermi, W., Del Prete, A., Gonzalvo-Feo, S., Garlanda, C., Azzolino, O., Salogni, L., et al. (2010). Nonredundant role of CCRL2 in lung dendritic cell trafficking. Blood *116*, 2942–2949.

Ottensmeier, C.H., Perry, K.L., Harden, E.L., Stasakova, J., Jenei, V., Fleming, J., Wood, O., Woo, J.J., Woelk, C.H., Thomas, G.J., et al. (2016). Upregulated glucose metabolism correlates inversely with CD8+ T cell infiltration and survival in squamous cell carcinoma. Cancer Res. *76*, 4136–4148.

Palacios, E.H., and Weiss, A. (2004). Function of the Src-family kinases, Lck and Fyn, in T-cell development and activation. Oncogene *23*, 7990–8000.

Pallmann, P., Schaarschmidt, F., Hothorn, L.A., Fischer, C., Nacke, H., Priesnitz, K.U., and Schork, N.J. (2012). Assessing group differences in biodiversity by simultaneously testing a user-defined selection of diversity indices. Mol. Ecol. Resour. *12*, 1068–1078.

Pan, Y., Tian, T., Park, C.O., Lofftus, S.Y., Mei, S., Liu, X., Luo, C., O'Malley, J.T., Gehad, A., Teague, J.E., et al. (2017). Survival of tissue-resident memory T cells requires exogenous lipid uptake and metabolism. Nature *543*, 252–256.

Pardoll, D.M. (2012). The blockade of immune checkpoints in cancer immunotherapy. Nat. Rev. Cancer *12*, 252–264.

Park, C.O., and Kupper, T.S. (2015). The emerging role of resident memory T cells in protective immunity and inflammatory disease. Nat. Med. *21*, 688–697.

Park, S.L., and Mackay, L.K. (2017). PD-1: always on my mind. Immunol. Cell Biol.

Patil, V.S., Madrigal, A., Schmiedel, B.J., Clarke, J., O'Rourke, P., de Silva, A.D., Harris, E., Peters, B., Seumois, G., Weiskopf, D., et al. (2018). Precursors of human CD4 + cytotoxic T lymphocytes identified by single-cell transcriptome analysis. Sci. Immunol. *3*, 8664.

Pfirschke, C., Engblom, C., Rickelt, S., Cortez-Retamozo, V., Garris, C., Pucci, F., Yamazaki, T., Poirier-Colame, V., Newton, A., Redouane, Y., et al. (2016). Immunogenic Chemotherapy Sensitizes Tumors to Checkpoint Blockade Therapy. Immunity *44*, 343–354.

Philip, M., Fairchild, L., Sun, L., Horste, E.L., Camara, S., Shakiba, M., Scott, A.C., Viale, A., Lauer, P., Merghoub, T., et al. (2017). Chromatin states define tumour-specific T cell dysfunction and reprogramming. Nature *545*, 452–456.

Picelli, S., Faridani, O.R., Björklund, Å.K., Winberg, G., Sagasser, S., and Sandberg, R. (2014). Full-length RNA-seq from single cells using Smart-seq2. Nat. Protoc. *9*, 171–181.

Piet, B., de Bree, G.J., Smids-Dierdorp, B.S., van der Loos, C.M., Remmerswaal, E.B.M., von der Thüsen, J.H., van Haarst, J.M.W., Eerenberg, J.P., ten Brinke, A., van der Bij, W., et al. (2011). CD8+ T cells with an intraepithelial phenotype upregulate cytotoxic function upon influenza infection in human lung. J. Clin. Invest. *121*, 2254–2263.

Qureshi, O.S., Zheng, Y., Nakamura, K., Attridge, K., Manzotti, C., Schmidt, E.M., Baker, J., Jeffery, L.E., Kaur, S., Briggs, Z., et al. (2011). Trans-endocytosis of CD80 and CD86: a molecular basis for the cell-extrinsic function of CTLA-4. Science *332*, 600–603.

Raaschou-Nielsen, O., Andersen, Z.J., Beelen, R., Samoli, E., Stafoggia, M., Weinmayr, G., Hoffmann, B., Fischer, P., Nieuwenhuijsen, M.J., Brunekreef, B., et al. (2013). Air pollution and lung cancer incidence in 17 European cohorts: prospective analyses from the European Study of Cohorts for Air Pollution Effects (ESCAPE). Lancet. Oncol. *14*, 813–822.

Raj, A., and van Oudenaarden, A. (2008). Nature, Nurture, or Chance: Stochastic Gene Expression and Its Consequences. Cell *135*, 216–226.

Ramesh, M., Simchoni, N., Hamm, D., and Cunningham-Rundles, C. (2015). High-throughput sequencing reveals an altered T cell repertoire in X-linked

agammaglobulinemia. Clin. Immunol.

Reck, M., Rodríguez-Abreu, D., Robinson, A.G., Hui, R., Csőszi, T., Fülöp, A., Gottfried, M., Peled, N., Tafreshi, A., Cuffe, S., et al. (2016). Pembrolizumab versus Chemotherapy for PD-L1–Positive Non–Small-Cell Lung cancer. N. Engl. J. Med. *375*, 1823–1833.

Reeves, R.K., Li, H., Jost, S., Blass, E., Li, H., Schafer, J.L., Varner, V., Manickam, C., Eslamizar, L., Altfeld, M., et al. (2015). Antigen-specific NK cell memory in rhesus macaques. Nat. Immunol. *16*, 927–932.

Remark, R., Becker, C., Gomez, J.E., Damotte, D., Dieu-Nosjean, M.C., Saut??s-Fridman, C., Fridman, W.H., Powell, C.A., Altorki, N.K., Merad, M., et al. (2015). The non-small cell lung cancer immune contexture: A major determinant of tumor characteristics and patient outcome. Am. J. Respir. Crit. Care Med. *191*, 377–390.

Restifo, N.P. (2013). A "big data" view of the tumor "immunome". Immunity *39*, 631–632.

Restifo, N.P., Dudley, M.E., and Rosenberg, S. a (2012). Adoptive immunotherapy for cancer: harnessing the T cell response. Nat. Rev. Immunol. *12*, 269–281.

Rickert, R.C. (2013). New insights into pre-BCR and BCR signalling with relevance to B cell malignancies. Nat. Rev. Immunol. *13*, 578–591.

Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. *43*, e47–e47.

Robert, C., Schachter, J., Long, G. V., Arance, A., Grob, J.J., Mortier, L., Daud, A., Carlino, M.S., McNeil, C., Lotem, M., et al. (2015a). Pembrolizumab versus Ipilimumab in Advanced Melanoma. N. Engl. J. Med. *372*, 2521–2532.

Robert, L., Tsoi, J., Wang, X., Emerson, R., Homet, B., Thinle, C., Mok, S., Huang, R.R., Cochran, A.J., Comin-Andiux, B., et al. (2015b). CTLA4 Blockade Broadens the Peripheral T cell Receptor Repertoire. Clin. Cancer Res. *20*, 2424–2432.

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics *26*, 139–140.

Roche, P.A., and Furuta, K. (2015). The ins and outs of MHC class II-mediated antigen processing and presentation. Nat. Rev. Immunol. *15*, 203–216.

Roh, W., Chen, P.-L., Reuben, A., Spencer, C.N., Prieto, P.A., Miller, J.P., Gopalakrishnan, V., Wang, F., Cooper, Z.A., Reddy, S.M., et al. (2017). Integrated molecular analysis of tumor biopsies on sequential CTLA-4 and PD-1 blockade reveals markers of response and resistance. Sci. Transl. Med. *9*.

Rojas, J., Sampath, P., Hou, W., and Thorne, S.H. (2015). Defining Effective Combinations of Immune Checkpoint Blockade and Oncolytic Virotherapy. Clin. Cancer Res.

Rooney, M.S., Shukla, S.A., Wu, C.J., Getz, G., and Hacohen, N. (2015). Molecular and Genetic Properties of Tumors Associated with Local Immune Cytolytic Activity. Cell *160*, 48–61.

Rudolph, M.G., Stanfield, R.L., and Wilson, I.A. (2006). How TCRs bind MHCs, peptides, and coreceptors. Annu. Rev. Immunol. *24*, 419–466.

Sandberg, R. (2013). Entering the era of single-cell transcriptomics in biology and medicine. Nat. Methods *11*, 22–24.

Satija, R., and Shalek, A.K. (2014). Heterogeneity in immune responses: from populations to single cells. Trends Immunol. *35*, 219–229.

Satija, R., Farrell, J. a, Gennert, D., Schier, A.F., and Regev, A. (2015). Spatial reconstruction of single-cell gene expression data. Nat. Biotechnol. *33*.

Sato, T., Vries, R.G., Snippert, H.J., van de Wetering, M., Barker, N., Stange, D.E., van Es, J.H., Abo, A., Kujala, P., Peters, P.J., et al. (2009). Single Lgr5 stem cells build crypt-villus structures in vitro without a mesenchymal niche. Nature *459*, 262–265.

Satpathy, A.T., Saligrama, N., Buenrostro, J.D., Wei, Y., Wu, B., Rubin, A.J., Granja, J.M., Lareau, C.A., Li, R., Qi, Y., et al. (2018). Transcript-indexed ATAC-seq for precision immune profiling. Nat. Med. 1.

Schenkel, J.M., and Masopust, D. (2014). Tissue-resident memory T cells. Immunity *41*, 886–897.

Scott-Browne, J.P., López-Moyado, I.F., Trifari, S., Wong, V., Chavez, L., Rao, A., and Pereira, R.M. (2016). Dynamic Changes in Chromatin Accessibility Occur in CD8+ T cells Responding to Viral Infection. Immunity.

Seder, R.A., Darrah, P.A., and Roederer, M. (2008). T-cell quality in memory and protection: implications for vaccine design. Nat. Rev. Immunol. *8*, 247–

258.

Seumois, G., Zapardiel-Gonzalo, J., White, B., Singh, D., Schulten, V., Dillon, M., Hinz, D., Broide, D.H., Sette, A., Peters, B., et al. (2016). Transcriptional Profiling of Th2 Cells Identifies Pathogenic Features Associated with Asthma. J. Immunol.

Sewell, A.K. (2012). Why must T cells be cross-reactive? Nat. Rev. Immunol. *12*, 669–677.

Shalek, A.K., Satija, R., Shuga, J., Trombetta, J.J., Gennert, D., Lu, D., Chen, P., Gertner, R.S., Gaublomme, J.T., Yosef, N., et al. (2014). Single-cell RNA-seq reveals dynamic paracrine control of cellular variation. Nature *510*, 363–369.

Sharma, P., and Allison, J.P. (2015). The future of immune checkpoint therapy. Science. *348*, 56–61.

Shetty, K., and Schatz, D.G. (2015). Recruitment of RAG1 and RAG2 to chromatinized DNA during V(D)J recombination. Mol. Cell. Biol.

Shugay, M., Britanova, O. V, Merzlyak, E.M., Turchaninova, M.A., Mamedov, I.Z., Tuganbaev, T.R., Bolotin, D.A., Staroverov, D.B., Putintseva, E. V, Plevova, K., et al. (2014). Towards error-free profiling of immune repertoires. Nat. Methods *11*, 653–655.

Shwetank, Abdelsamed, H.A., Frost, E.L., Schmitz, H.M., Mockus, T.E., Youngblood, B.A., and Lukacher, A.E. (2017). Maintenance of PD-1 on brain-resident memory CD8 T cells is antigen independent. Immunol. Cell Biol. *95*, 953–959.

Sim, G., Martin-Orozco, N., and Jin, L. (2014). IL-2 therapy promotes suppressive ICOS+ Treg expansion in melanoma patients. J. … *124*, 99–110.

Simon, P., Omokoko, T.A., Breitkreuz, A., Hebich, L., Kreiter, S., Attig, S., Konur, A., Britten, C.M., Paret, C., Dhaene, K., et al. (2014). Functional TCR retrieval from single antigen-specific human T cells reveals multiple novel epitopes. Cancer Immunol. Res. *2*, 1230–1244.

Simoni, Y., Becht, E., Fehlings, M., Loh, C.Y., Koo, S.-L., Teng, K.W.W., Yeong, J.P.S., Nahar, R., Zhang, T., Kared, H., et al. (2018). Bystander CD8+ T cells are abundant and phenotypically distinct in human tumour infiltrates. Nature *557*, 575–579.

Skon, C.N., Lee, J.-Y., Anderson, K.G., Masopust, D., Hogquist, K.A., and Jameson, S.C. (2013). Transcriptional downregulation of S1pr1 is required for the establishment of resident memory CD8+ T cells. Nat. Immunol. *14*, 1285–1293.

Skowera, A., Ladell, K., McLaren, J.E., Dolton, G., Matthews, K.K., Gostick, E., Kronenberg-Versteeg, D., Eichmann, M., Knight, R.R., Heck, S., et al. (2015). β-cell-specific CD8 T cell phenotype in type 1 diabetes reflects chronic autoantigen exposure. Diabetes *64*, 916–925.

Smith, S.N., Wang, Y., Baylon, J.L., Singh, N.K., Baker, B.M., Tajkhorshid, E., and Kranz, D.M. (2014). Changing the peptide specificity of a human T-cell receptor by directed evolution. Nat. Commun. *5*, 5223.

Srivastava, S., and Riddell, S.R. (2015). Engineering CAR-T Cells: Design Concepts. Trends Immunol. *36*, 494–502.

Stanford, S.M., Rapini, N., and Bottini, N. (2012). Regulation of TCR signalling by tyrosine phosphatases: from immune homeostasis to autoimmunity. Immunology *137*, 1–19.

Stemberger, C., Huster, K.M., Koffler, M., Anderl, F., Schiemann, M., Wagner, H., and Busch, D.H. (2007). A Single Naive CD8+ T cell Precursor Can Develop into Diverse Effector and Memory Subsets. Immunity *27*, 985–997.

Stemberger, C., Graef, P., Odendahl, M., Albrecht, J., Dössinger, G., Anderl, F., Buchholz, V.R., Gasteiger, G., Schiemann, M., Grigoleit, G.U., et al. (2014). Lowest numbers of primary CD8(+) T cells can reconstitute protective immunity upon adoptive immunotherapy. Blood *124*, 628–637.

Stepanek, O., Prabhakar, A.S., Osswald, C., King, C.G., Bulek, A., Naeher, D., Beaufils-Hugot, M., Abanto, M.L., Galati, V., Hausmann, B., et al. (2014). Coreceptor Scanning by the T cell Receptor Provides a Mechanism for T cell Tolerance. Cell *159*, 1–13.

Stevanović, S., Pasetto, A., Helman, S.R., Gartner, J.J., Prickett, T.D., Howie, B., Robins, H.S., Robbins, P.F., Klebanoff, C.A., Rosenberg, S.A., et al. (2017). Landscape of immunogenic tumor antigens in successful immunotherapy of virally induced epithelial cancer. Science. *356*, 200–205.

Stevenson, F.K., Ottensmeier, C.H., Johnson, P., Zhu, D., Buchan, S.L., McCann, K.J., Roddick, J.S., King, A.T., McNicholl, F., Savelyeva, N., et al. (2004). DNA vaccines to attack cancer. Proc. Natl. Acad. Sci. U. S. A. *101 Suppl*, 14646–14652.

Stransky, N., Egloff, A.M., Tward, A.D., Kostic, A.D., Cibulskis, K., Sivachenko, A., Kryukov, G. V, Lawrence, M.S., Sougnez, C., McKenna, A., et al. (2011). The mutational landscape of head and neck squamous cell carcinoma. Science *333*, 1157–1160.

Stubbington, M.J., Mahata, B., Svensson, V., Deonarine, A., Nissen, J.K., Betz, A.G., and Teichmann, S.A. (2015). An atlas of mouse CD4(+) T cell

transcriptomes. Biol. Direct *10*, 14.

Stubbington, M.J.T., Lönnberg, T., Proserpio, V., Clare, S., Speak, A.O., Dougan, G., and Teichmann, S.A. (2016). T cell fate and clonality inference from single-cell transcriptomes. Nat. Methods *13*, 329–332.

Su, L.F., Kidd, B.A., Han, A., Kotzin, J.J., and Davis, M.M. (2013). Virus-Specific CD4+ Memory-Phenotype T cells Are Abundant in Unexposed Adults. Immunity *38*, 373–383.

Takada, K., Van Laethem, F., Xing, Y., Akane, K., Suzuki, H., Murata, S., Tanaka, K., Jameson, S.C., Singer, A., and Takahama, Y. (2015). TCR affinity for thymoproteasome-dependent positively selecting peptides conditions antigen responsiveness in CD8+ T cells. Nat. Immunol. *16*, 1069–1076.

Tangye, S.G., Ma, C.S., Brink, R., and Deenick, E.K. (2013). The good, the bad and the ugly - TFH cells in human health and disease. Nat. Rev. Immunol. *13*, 412–426.

Tasic, B., Menon, V., Nguyen, T.N., Kim, T.K., Jarsky, T., Yao, Z., Levi, B., Gray, L.T., Sorensen, S.A., Dolbeare, T., et al. (2016). Adult mouse cortical cell taxonomy revealed by single cell transcriptomics. Nat. Neurosci. *19*, 335–346.

Thomas, N., Heather, J., Ndifon, W., Shawe-Taylor, J., and Chain, B. (2013). Decombinator: a tool for fast, efficient gene assignment in T-cell receptor sequences using a finite state machine. Bioinformatics *29*, 542–550.

Tian, Y., Babor, M., Lane, J., Schulten, V., Patil, V.S., Seumois, G., Rosales, S.L., Fu, Z., Picarda, G., Burel, J., et al. Unique phenotypes and clonal expansions of human CD4 effector memory T cells re-expressing CD45RA.

Tirosh, I., Izar, B., Prakadan, S.M., Wadsworth, M.H., Treacy, D., Trombetta, J.J., Rotem, A., Rodman, C., Lian, C., Murphy, G., et al. (2016). Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. Science. *352*, 189–196.

Tran, E., Ahmadzadeh, M., Lu, Y.-C., Gros, A., Turcotte, S., Robbins, P.F., Gartner, J.J., Zheng, Z., Li, Y.F., Ray, S., et al. (2015). Immunogenicity of somatic mutations in human gastrointestinal cancers. Science. science.aad1253-.

Trapnell, C., Pachter, L., and Salzberg, S.L. (2009). TopHat: discovering splice junctions with RNA-Seq. Bioinformatics *25*, 1105–1111.

Trapnell, C., Cacchiarelli, D., Grimsby, J., Pokharel, P., Li, S., Morse, M., Lennon, N.J., Livak, K.J., Mikkelsen, T.S., and Rinn, J.L. (2014). The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. Nat. Biotechnol. *32*, 381–386.

Treutlein, B., Brownfield, D.G., Wu, A.R., Neff, N.F., Mantalas, G.L., Espinoza, F.H., Desai, T.J., Krasnow, M. a, and Quake, S.R. (2014). Reconstructing lineage hierarchies of the distal lung epithelium using single-cell RNA-seq. Nature *509*, 371–375.

Trombetta, J.J., Gennert, D., Lu, D., Satija, R., Shalek, A.K., and Regev, A. (2014). Preparation of Single-Cell RNA-Seq Libraries for Next Generation Sequencing. Curr. Protoc. Mol. Biol. *107*, 4.22.1-4.22.17.

Tsang, J.C.H., Yu, Y., Burke, S., Buettner, F., Wang, C., Kolodziejczyk, A.A., Teichmann, S.A., Lu, L., and Liu, P. (2015). Single-cell transcriptomic reconstruction reveals cell cycle and multi-lineage differentiation defects in Bcl11a-deficient hematopoietic stem cells. Genome Biol. *16*, 178.

Tumeh, P.C., Harview, C.L., Yearley, J.H., Shintaku, I.P., Taylor, E.J., Robert, L., Chmielowski, B., Spasic, M., Henry, G., Ciobanu, V., et al. (2014). PD-1 blockade induces responses by inhibiting adaptive immune resistance. Nature *515*, 568–571.

Turchaninova, M.A., Britanova, O. V, Bolotin, D.A., Shugay, M., Putintseva, E. V, Staroverov, D.B., Sharonov, G., Shcherbo, D., Zvyagin, I. V, Mamedov, I.Z., et al. (2013). Pairing of T-cell receptor chains via emulsion PCR. Eur. J. Immunol. *43*, 2507–2515.

Vanneman, M., and Dranoff, G. (2012). Combining immunotherapy and targeted therapies in cancer treatment. Nat. Rev. Cancer *12*, 237–251.

Van Verk, M.C., Hickman, R., Pieterse, C.M.J., and Van Wees, S.C.M. (2013). RNA-Seq: revelation of the messengers. Trends Plant Sci. *18*, 175–179.

Voskoboinik, I., Whisstock, J.C., and Trapani, J.A. (2015). Perforin and granzymes: function, dysfunction and human pathology. Nat. Rev. Immunol. *15*, 388–400.

Vyas, J., Veen, A. Van der, and Ploegh, H.L. (2008). The known unknowns of antigen processing and presentation. Nat. Rev. Immunol. *8*, 607–618.

Walker, L.S.K., and Sansom, D.M. (2011). The emerging role of CTLA4 as a cell-extrinsic regulator of T cell responses. Nat. Rev. Immunol. *11*, 852–863.

Wang, R., Natarajan, K., and Margulies, D.H. (2009). Structural basis of the CD8 alpha beta/MHC class I interaction: focused recognition orients CD8 beta

to a T cell proximal position. J. Immunol. *183*, 2554–2564.

Wang, S., Xia, P., Chen, Y., Huang, G., Xiong, Z., Liu, J., Li, C., Ye, B., Du, Y., and Fan, Z. (2016). Natural Killer-like B Cells Prime Innate Lymphocytes against Microbial Infection. Immunity *45*, 131–144.

Ward, M.J., Mellows, T., Harris, S., Webb, A., Patel, N.N., Cox, H.J., Piper, K., Ottensmeier, C.H., Thomas, G.J., and King, E. V (2014). Staging and treatment of oropharyngeal cancer in the human papillomavirus era. Head Neck.

Weng, W.-K., Armstrong, R., Arai, S., Desmarais, C., Hoppe, R., and Kim, Y.H. (2013). Minimal residual disease monitoring with high-throughput sequencing of T cell receptors in cutaneous T cell lymphoma. Sci. Transl. Med. *5*, 214ra171.

Wherry, E.J. (2011). T cell exhaustion. Nat. Immunol. *12*, 492–499.

Wolchok, J.D., Kluger, H., Callahan, M.K., Postow, M.A., Rizvi, N.A., Lesokhin, A.M., Segal, N.H., Ariyan, C.E., Gordon, R.-A., Reed, K., et al. (2013). Nivolumab plus Ipilimumab in Advanced Melanoma. N. Engl. J. Med. *369*, 122–133.

Wölfl, M., and Greenberg, P.D. (2014). Antigen-specific activation and cytokine-facilitated expansion of naive, human CD8+ T cells. Nat. Protoc. *9*, 950–966.

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. (2016). A High-Dimensional Atlas of Human T cell Diversity Reveals Tissue-Specific Trafficking and Cytokine Signatures. Immunity *45*, 442–456.

Wood, O., Clarke, J., Woo, J., Mirza, A.H., Woelk, C.H., Thomas, G.J., Vijayanand, P., King, E., and Ottensmeier, C.H. (2017). Head and neck squamous cell carcinomas are characterized by a stable immune signature within the primary tumor over time and space. Clin. Cancer Res. *23*, 7641–7649.

Wood, O., Woo, J., Seumois, G., Savelyeva, N., McCann, K.J., Singh, D., Jones, T., Peel, L., Breen, M.S., Ward, M., et al. (2016). Gene expression analysis of TIL rich HPV-driven head and neck tumors reveals a distinct B-cell signature when compared to HPV independent tumors. Oncotarget *7*, 56781–56797.

Wu, Y.-L., Ding, Y.-P., Tanaka, Y., Shen, L.-W., Wei, C.-H., Minato, N., and Zhang, W. (2014). γδ T cells and their potential for immunotherapy. Int. J. Biol. Sci. *10*, 119–135.

Wynn, T.A. (2005). TH-17: a giant step from TH1 and TH2. Nat. Immunol. *6*, 1069–1070.

Xiao, Z., Casey, K.A., Jameson, S.C., Curtsinger, J.M., and Mescher, M.F. (2009). Programming for CD8 T cell Memory Development Requires IL-12 or Type I IFN. J. Immunol. *182*, 2786–2794.

Yosef, N., Shalek, A.K., Gaublomme, J.T., Jin, H., Lee, Y., Awasthi, A., Wu, C., Karwacz, K., Xiao, S., Jorgolli, M., et al. (2013). Dynamic regulatory network controlling TH17 cell differentiation. Nature *496*, 461–468.

Yu, P., and Lin, W. (2016). Single-cell Transcriptome Study as Big Data. Genomics. Proteomics Bioinformatics.

Yu, W., Jiang, N., Ebert, P.J.R., Kidd, B.A., Müller, S., Lund, P.J., Juang, J., Adachi, K., Tse, T., Birnbaum, M.E., et al. (2015). Clonal Deletion Prunes but Does Not Eliminate Self-Specific αβ CD8(+) T Lymphocytes. Immunity *42*, 929–941.

Yu, X.G., Lichterfeld, M., Williams, K.L., Martinez-Picado, J., and Walker, B.D. (2007). Random T-cell receptor recruitment in human immunodeficiency virus type 1 (HIV-1)-specific CD8+ T cells from genetically identical twins infected with the same HIV-1 strain. J. Virol. *81*, 12666–12669.

Zhang, N., and Bevan, M.J. (2011). CD8(+) T cells: foot soldiers of the immune system. Immunity *35*, 161–168.

Zhang, B., Wang, J., Wang, X., Zhu, J., Liu, Q., Shi, Z., Chambers, M.C., Zimmerman, L.J., Shaddox, K.F., Kim, S., et al. (2014). Proteogenomic characterization of human colon and rectal cancer. Nature.

Zhang, T., Berrocal, J.G., Frizzell, K.M., Gamble, M.J., DuMond, M.E., Krishnakumar, R., Yang, T., Sauve, A. a, and Kraus, W.L. (2009). Enzymes in the NAD+ salvage pathway regulate SIRT1 activity at target gene promoters. J. Biol. Chem. *284*, 20408–20417.

Zhao, W., He, X., Hoadley, K.A., Parker, J.S., Hayes, D.N., and Perou, C.M. (2014). Comparison of RNA-Seq by poly (A) capture, ribosomal RNA depletion, and DNA microarray for expression profiling. BMC Genomics *15*, 419.

Zheng, C., Zheng, L., Yoo, J.-K., Guo, H., Zhang, Y., Guo, X., Kang, B., Hu, R., Huang, J.Y., Zhang, Q., et al. (2017). Landscape of Infiltrating T cells in Liver Cancer Revealed by Single-Cell Sequencing. Cell *169*, 1342–1356.e16.

Zheng, G.X.Y., 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. (2016). Massively parallel digital transcriptional profiling of single cells. bioRxiv *8*, 65912.

Zhou, L., Chong, M.M.W., and Littman, D.R. (2009). Plasticity of CD4+ T cell Lineage Differentiation. Immunity *30*, 646–655.

Zhou, Y., Zhang, L., Romaguera, J., Delasalle, K., Han, X., Du, X., Kwak, L., Yi, Q., and Wang, M. (2008). Immunotherapy in mantle cell lymphoma: anti-CD20-based therapy and beyond. Am. J. Hematol. *83*, 144–149.

Ziegenhain, C., Vieth, B., Parekh, S., Reinius, B., Guillaumet-Adkins, A., Smets, M., Leonhardt, H., Heyn, H., Hellmann, I., and Enard, W. (2017). Comparative Analysis of Single-Cell RNA Sequencing Methods. Mol. Cell *65*, 631–643.e4.

Zou, W., and Restifo, N.P. (2010). TH17 cells in tumour immunity and immunotherapy. Nat. Rev. Immunol. *10*, 248–256.

Zvyagin, I. V., Pogorelyy, M. V., Ivanova, M.E., Komech, E.A., Shugay, M., Bolotin, D.A., Shelenkov, A.A., Kurnosov, A.A., Staroverov, D.B., Chudakov, D.M., et al. (2014). Distinctive properties of identical twins' TCR repertoires revealed by high-throughput sequencing. Proc. Natl. Acad. Sci. U. S. A. *111*, 5980–5985.