Features of the intratumoural T cell receptor repertoire associated with antigen exposure in cancer patients

Marc Louis Robert de Massy

University College London

PhD Supervisor: Professor Sergio A. Quezada

A thesis submitted for the degree of

Doctor of Philosophy

University College London

May 2021

Declaration

I, Marc Robert de Massy, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Abstract

The clinical success of immunotherapies demonstrates the importance of the immune system in tumour control, but the response rates remain low and many biological mechanisms underlying how these therapies work are still uncharacterised. In particular, the specificity of the anti-tumour immune response pre-existing in treatment-naive patients or induced by treatment remains poorly described.

In this thesis, I explore how T cell receptor (TCR) sequencing data in multi-omics contexts can be utilised to identify features associated with antigen exposure in cancer patients.

In treatment-naive non-small cell lung cancer (NSCLC) patients, multi-region TCR sequencing revealed a pattern of heterogeneity in the TCR repertoire resembling the heterogeneity observed in the mutational profile of these tumours and a range of clonotype frequency values associated with tumour specificity. A novel method was built in order to identify distinct TCR populations that spatially follow the pattern of the well-established clonal/subclonal mutational dichotomy.

The impact of immune checkpoint blockade therapy on the TCR repertoire distribution was assessed in advanced renal cell carcinoma in the context of anti-PD1 treatment. TCRs with frequency distribution characteristics similar to what was observed in NSCLC were maintained upon treatment and associated with clinical response. In addition, RNA-sequencing analysis identified a gene expression profile consistent with specific activation of T cells through TCR signalling.

Finally, the same methodology was applied to bone marrow samples harvested from B cell acute lymphoblastic leukaemia (B-ALL) patients. A statistical framework was developed in order to efficiently distinguish leukaemic re-arrangements from the non-leukaemic TCR repertoire of B-ALL patients. Subsequently, longitudinal analysis revealed TCR distributions that suggested the presence of cytotoxic T cells which was further characterised in matched single-cell RNA sequencing data.

Impact Statement

The findings presented in this thesis highlight that immune repertoire data is a valuable resource for the identification of features associated with antigen exposure within the intratumoural TCR repertoire. This thesis further demonstrates that these features can be leveraged for the characterisation of the potentially tumour-reactive TCR repertoire and can be used as biomarkers of response to immunotherapy. The analyses conducted throughout the result chapters are straightforward to implement and capture insightful signals in three separate types of cancer: non-small cell lung cancer (NSCLC), clear cell renal cell carcinoma (ccRCC) and childhood B cell acute lymphoblastic leukaemia (B-ALL). Many of the methodologies developed were transferrable across these three cancers, even though their immunological microenvironments differ in many aspects, demonstrating that the bioinformatics framework developed throughout this thesis is of general application in the analysis of numerous types of immune repertoire data.

In NSCLC, non-synonymous mutations in the cancer cells' genome have been shown to give rise to immunogenic antigens, capable of triggering an anti-tumour immune response. The heterogeneity of the cancer genomic landscape is critical to the outcome of the immunotherapies currently being developed for lung cancer. Analyses performed in the TCR repertoire study of patients with NSCLC demonstrate that this genomic heterogeneity is mirrored in the intratumoural TCR repertoire. The findings suggest that TCR repertoire heterogeneity may be a critical feature in determining the outcome of immunotherapy, and may help design future more elaborate therapeutic strategies to overcome this heterogeneity.

In ccRCC, the sources of antigenic stimuli remain elusive and the mutation-derived antigens described above for NSCLC do not correlate with longer survival or clinical response to immunotherapy. Whilst identifying such antigens remains a key challenge, the TCR repertoire analyses performed in the study of metastatic ccRCC patients treated with anti-PD-1 therapy suggest that tumour-specific T cells can be identified without knowing the antigenic source. TCRs identified by such analyses could form the basis of adoptive cellular therapy despite the limited description of the antigenic landscape in renal cancer. In addition, the findings demonstrate that

quantitative characterisation of TCR repertoire may provide a biomarker of response to checkpoint blockade in ccRCC.

In childhood B-ALL, most of the immunotherapies that are currently being developed focus on targeting well-established cell-surface antigens, such as CD19, with engineered CAR T cells. Little is known about the endogenous immune response in this disease and how it could be leveraged by different therapeutic approaches. Analyses performed in the study of childhood B-ALL patients identify features of the TCR repertoire associated with reactivity similar to the ones captured in the NSCLC cohort and the ccRCC cohort. The findings suggest that cancer-reactive T cells are present and can be identified in the bone marrow of B-ALL patients, which potentially opens new immune-therapeutic paths leveraging checkpoint inhibitors, vaccines or TCR-based cell therapies for instance.

Acknowledgements

I would like to thank my supervisors Professor Sergio A. Quezada, Professor Javier Herrero and Professor Benny Chain. I am truly grateful to have been able to benefit from their guidance and expertise and have been nothing but inspired by their mentorship.

Thank you to the past and present members of the Quezada group and the Chain group for countless hours of discussing ideas, fascinating collaborations and precious friendships.

I am very thankful to have been involved in large collaborative efforts and to have had the opportunity to work with the TRACERx consortium, the Turajlic group at the Francis Crick Institute and the Mansour group at UCL. I am truly grateful to Cancer Research UK for the funding of my work and to all patients involved in the projects presented in this thesis, and their families.

Finally, thank you to my friends and my family for believing in me and giving me perspective. A very special thank you to my partner Alice for her support throughout the journey.

Table of Contents

Table of figures

Abbreviations

ADAPTeR	A Study of Anti-PD-1 (Nivolumab) Therapy as Pre- and Post-operative therapy in Metastatic Renal Cell Cancer
B-ALL	B cell acute lymphoblastic leukaemia
BCR	B cell receptor
BMMC	bone marrow mononuclear cell
CAR T cell	chimeric antigen receptor-modified T cell
cDNA	complementary DNA
ccRCC	clear cell renal cell carcinoma
CDR	complementary-determining region
CTLA-4	cytotoxic T-lymphocyte protein 4
DAPI	4’,6-diamidino-2-phenylindole
DGEA	differential gene expression analysis
DNA	deoxyribonucleic acid
DC	dendritic cell
EBV	Epstein-Barr virus
FACS	fluorescence-activated cell sorting
GOBP	gene ontology biological process
GLIPH	Grouping of Lymphocytes Interactions by Paratope Hotspots
GSEA	gene set enrichment analysis
HSC	hematopoietic stem cell
HIV-1	human immunodeficiency virus
HLA	human leukocyte antigen
ICB	immune checkpoint blockade
IDO	indoleamine-pyrrole 2,3-dioxygenase
IF	immunofluorescence
IFN	interferon
IHC	immunohistochemistry
IQR	interquartile range
IRAE	immune related adverse event
MAIT cells	mucosal associated invariant T cells
MCA	methylcholanthrene
MHC	major histocompatibility complex

MRD	minimal residual disease
NGS	next-generation sequencing
NK cells	natural killer cells
NK T cells	natural killer T cells
NSCLC	non-small cell lung cancer
NSM	non-synonymous mutation
PBMC	peripheral blood mononuclear cell
PC	principal component
PCR	polymerase chain reaction
PD-1	programmed cell death 1
PD-L1	programmed cell death ligand 1
PEACE	Posthumous Evaluation of Advanced Cancer Environment
QC	quality control
RAG	recombination activating gene
RCC	renal cell carcinoma
RECIST	response evaluation criteria in solid tumours
RNA	ribonucleic acid
RNA-Seq	RNA sequencing
scRNA-Seq	single-cell RNA sequencing
scTCR-Seq	single-cell TCR sequencing
STAT3	signal transducer and activator of transcription 3
T-ALL	T cell acute lymphoblastic leukaemia
TCR	T cell receptor
TCR-Seq	T cell receptor sequencing
TIL	tumour-infiltrating lymphocyte
TMB	tumour mutational burden
TME	tumour microenvironment
TRACERx	Tracking Cancer Evolution through therapy (Rx)
UMAP	uniform manifold approximation and projection
UMI	unique molecular identifier
VAF	variant allele frequency
WES	whole-exome sequencing
3D	three-dimensional

Chapter 1. Introduction

1.1 Cancer immunosurveillance

The idea that the immune system could provide protection against cancer was first proposed at the beginning of the 20th century (Ehrlich, 1908). Since then, the scientific community has been adding more layers to the complex relationship that exists between cancer and the immune system. One framework commonly used to describe this relationship consists of three main phases: Elimination, Equilibrium and Escape (Dunn, Old, & Schreiber, 2004a) (Figure 1.1).

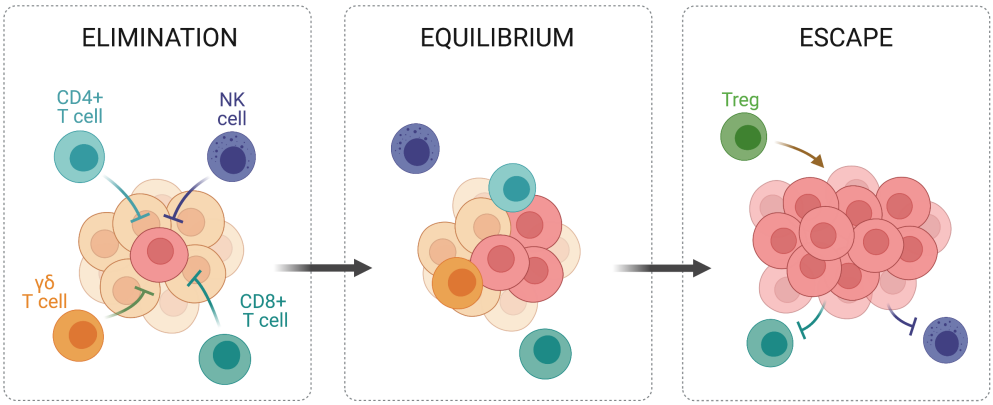

Figure 1.1 Cancer cells can be recognised and eliminated by immune cells

Adapted from "Cancer Immunoediting" by BioRender.com (2021). Retrieved from https://biorender.com/biorender-templates. Adapted from Weaver & Murphy, 2016. Tumour cells can be recognised by a variety of immune cells including CD4 T cells, CD8 T cells, NK cells, γδ T cells and regulatory T cells. Some tumour cells can acquire the ability to escape the immune system and spread.

1.1.1 Immune cells have the capacity to eliminate nascent tumours

In the late 1950s, Frank MacFarlane Burnet and Lewis Thomas hypothesised that immune cells have the capacity to detect and eliminate tumour cells (Burnet, 1957; L. Thomas, 1959). A series of studies conducted in the late 90's and early 2000's (Dunn, Old, & Schreiber, 2004b) leveraged the exponential development, at the time, of monoclonal antibodies and genetically manipulated mouse models to validate this hypothesis and demonstrate that both the innate and the adaptive immune systems are required to provide protection against tumours. Most evidence supporting the existence of this process, called immunosurveillance and also referred to as the

elimination phase of immunoediting, and the three-phase model as a whole, is primarily based on mouse models. How this model translates to the human setting is still debatable. In particular, many of these experiments rely on mouse models in which tumours are induced by chemicals such as methylcholanthrene (MCA). These models involve induction of multiple mutations simultaneously in a way quite different from how we think human tumours develop (Wall & Shani, 2008). Caveats aside, the Elimination, Equilibrium and Escape model is extremely valuable to introduce several key concepts of immuno-oncology.

The recombination activating gene (RAG) pathway is responsible for the proper maturation of lymphocyte populations. Mice lacking either the RAG-1 or the RAG-2 gene are unable to generate functional αβ T cells, γδ T cells, natural killer (NK) T cells or B cells (Alt, Rathbun, Oltz, Taccioli, & Shinkai, 1992). RAG knock-out experiments constitute a pillar of immuno-oncology animal modelling and provided early evidence for the immunosurveillance process. In 2001, Shankaran et al. showed that mice lacking RAG-2 developed significantly more sarcomas than control animals following injection of MCA (Shankaran et al., 2001).

In parallel, the same year, Smyth et al. published a study demonstrating that innate NK cells were also necessary to the immunosurveillance process, leveraging another fundamental technique: antibody-mediated depletion of a specific subset of cells. They first utilised a depleting anti-NK1.1 antibody to show that treated mice, hence lacking both NK cells and NK T cells, were developing more MCA-induced sarcomas than their untreated counterparts (Smyth, Crowe, & Godfrey, 2001).

Additional studies conducted around that time further dissected the type of T cells required for efficient elimination of tumours (Gao et al., 2003; Girardi et al., 2003; Girardi et al., 2001; Hayakawa, Rovero, Forni, & Smyth, 2003; Smyth et al., 2000) and demonstrated that αβ T cells, γδ T cells and NK T cells were all individually essential to that process: animals lacking any of the three developed more cancer.

Together, this experimental data was the demonstration that, in animal models, both the innate and the adaptive compartment of the immune system are essential to continuously detect and eliminate growing tumours.

1.1.2 Immunosurveillance in humans

Equivalent evidence of immunosurveillance in humans is limited as the experimental strategies mentioned above are not applicable to human beings for obvious reasons. However, several indirect studies on the general population strongly make the case for the instrumental role of the adaptive and innate immune system in controlling nascent tumours (Finn, 2008).

By definition, cytotoxic T cells are cells that recognise and kill infected cells or cancer cells (Weaver & Murphy, 2016). In 2000, Imai et al. demonstrated in a cohort of 3,625 people followed for 11 years that subjects displaying a higher degree of cytotoxicity in blood derived CD8 T cells at the beginning of the study had a lower probability of developing a cancer of any type than individuals with lower levels of cytotoxic CD8 T cells (Imai, Matsuyama, Miyake, Suga, & Nakachi, 2000). In 2007, Roithmaier et al. studied a cohort of patients who received a lung or heart transplant and were subjected to immunosuppression in order to reduce the chances of graft rejection. In these immunosuppressed individuals, the prevalence of different cancers (including leukaemia, lymphomas, head and neck cancers and lung cancers) was much higher than in the general population (Roithmaier et al., 2007). This observation was in line with what was observed in mice, suggesting that impaired immune systems were less efficient at eliminating tumours over time than “normal” immune systems.

1.2 Equilibrium and Escape

1.2.1 Equilibrium

Although immune cells have the capacity to eliminate tumour cells, Elimination is undoubtedly failing in many instances which lead to the development of cancer in patients. In that regard, it is clear that Elimination is paired with Escape and it becomes natural to hypothesise that the Elimination-Escape system also comprises of a latent state: Equilibrium.

In 2007, Koebel et al. built a murine model to study this concept further. They induced tumours in mice with low dose MCA to generate tumour masses in the animals’ skin

that remained at a stable volume up to 200 days post first MCA injection, hence effectively creating a stable state in which tumours are neither eliminated nor do they grow uncontrollably. At this point, they treated a first cohort with a monoclonal antibody depleting both CD4 and CD8 T cells and neutralising interferon γ (IFN-γ) when a second control was treated with a control antibody. Strikingly, doing so triggered Escape in 60% of treated animals, which developed progressive sarcomas, against none in the control group (Koebel et al., 2007). This model of dormancy of primary tumour lesions and triggerable Escape by immune depletion was replicated by Eyles et al. in 2010 and extended to recurrent metastatic lesions in a melanoma mouse model (Eyles et al., 2010).

In humans, although it is not trivial to measure Equilibrium prior to diagnosis, its prevalence between primary and recurrent disease is routinely monitored in the clinical setting (Aguirre-Ghiso, 2007). Cancer patients can develop distant metastatic disease many years after the primary diagnosis. In many cases, disseminated or local tumour cells can still be detected over long periods of time after treatment. A proportion of cancer patients will relapse, often within five years of the primary disease but sometime up to 25 years later. Although a period of dormancy is apparent in these cases, the mechanism which triggers the progression of dormant lesions and breaks the equilibrium is still unclear.

1.2.2 Tumour evolution and immune escape

In the Equilibrium phase of the immune response to cancer, the action of cytotoxic T cells is essential in maintaining tumour masses to a stable volume (Koebel et al., 2007), enabling a continuous infiltration of lymphocytes. From the cancer cell's perspective, this is a very challenging context for survival which requires rapid adaptation in order to avoid Elimination (see Section 1.1.1).

Tumours exploit the plasticity and heterogeneity of their genome in order to tune down their interaction with the immune system and survive through long periods of time, a process called immunoediting. There are many aspects to the interaction

between tumours cells and immune cells, which can all be targeted by the tumour (**Figure 1.2**) (Weaver & Murphy, 2016).

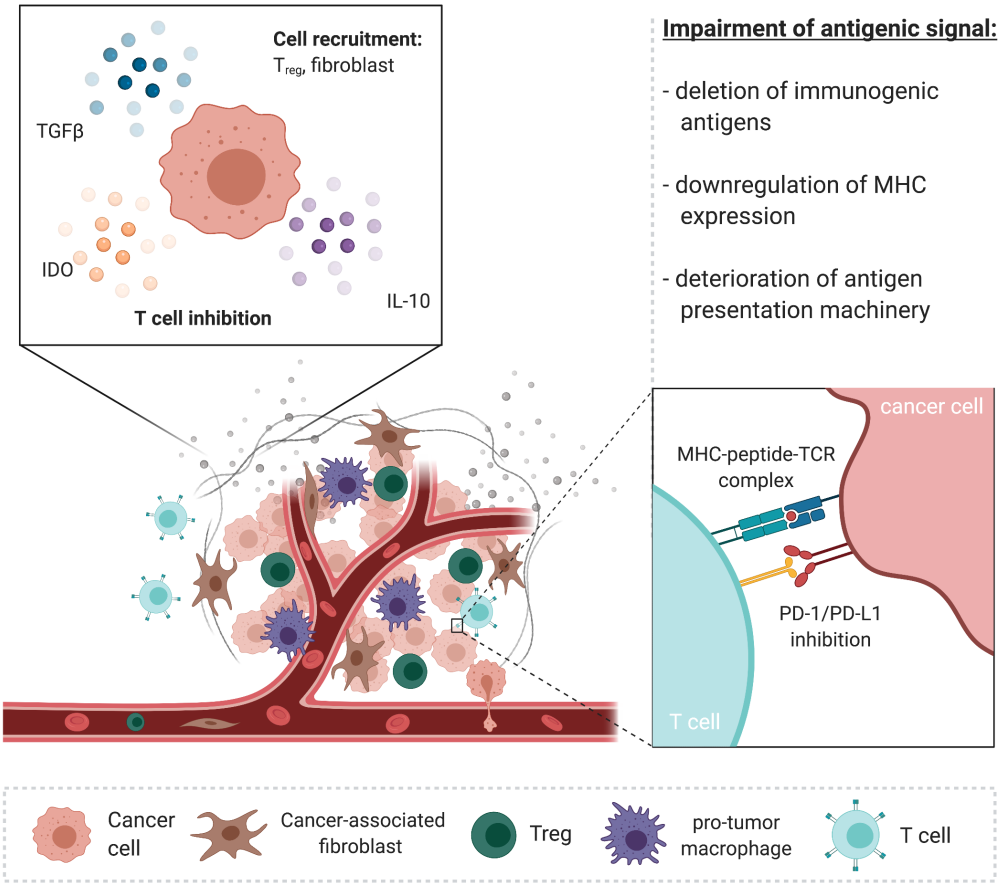

Figure 1.2 Tumours can avoid immune recognition in a variety of ways

Adapted from "The Tumour Microenvironment: Overview of Cancer-Associated Changes" and "Cold vs Hot Tumours" by BioRender.com (2021). Retrieved from https://biorender.com/biorender-templates. Adapted from Weaver & Murphy, 2016. Many aspects to the interaction between tumours cells and immune cells can be targeted by the tumour.

As mentioned above, a critical feature of the Elimination phase and the Equilibrium phase is the T cells' ability to recognise cancer cells. This process, which will be detailed and discussed further in Sections 1.3.1, 1.3.2 and 1.5.2, relies on the key interaction between the TCR on a T cell and the peptide-MHC complex on a cancer cell. Briefly, major histocompatibility complex (MHC) molecules are expressed on the surface of a cell and present peptides sampled from the cell's internal proteome. T cells scan these peptides via the TCR. A healthy proteome will not trigger any

response from the T cell. Conversely, in cancer cells, the internal proteome contains aberrant proteins translated from unstable genomic material which will result in foreign peptides, antigens by definition, being presented by MHC molecules and inducing an immune response.

The first class of mechanisms that impair the cytotoxic action of T cells is the deterioration of the antigenic signal, which can happen in at least three well documented ways (Schreiber, Old, & Smyth, 2011): modification of antigenic landscape through Darwinian selection of cancer subclones, loss of MHC proteins and impairment of intracellular antigen presentation machinery. The latter will not be discussed further in this thesis.

The plasticity tumour genomes enables the Darwinian selection of subclones with greater survival potential which modifies the antigenic landscape upon which the cancer-immune cycle (D. S. Chen & Mellman, 2013) is built. When and where the antigenic landscape is changing, the specificity of the current tumour-infiltrating lymphocytes (TILs) becomes irrelevant. If that process is dynamic enough, the tumour can continuously escape the specificity of newly infiltrated lymphocytes. In 2019, Rosenthal et al. described and bioinformatically documented two ways by which such reshaping occurs. In a cohort of NSCLC patients, they show that neoantigens (a subtype of tumour specific antigens) can be erased at the DNA level through copy number loss and at the RNA level through selective downregulation of transcripts (Rosenthal et al., 2019).

Starting 30 years ago, several studies demonstrated at the protein levels that human tumours are enriched for defects in MHC expression, either through direct loss (Algarra, Cabrera, & Garrido, 2000; Lopez-Nevot et al., 1989; Ohnmacht & Marincola, 2000) or loss of other components of the pathway (Kloor, von Knebel Doeberitz, & Gebert, 2005; Seliger, Maeurer, & Ferrone, 2000). More recently, it was bioinformatically shown that impairment of this mechanism was selected by cancer evolution and detected in the DNA of as much as 40% of NSCLC patients in the TRACERx cohort (McGranahan et al., 2017).

Another important phenomenon that falls under this first category is cancer immune tolerance (Lu & Finn, 2008). Unlike the mechanisms described above which rely solely on tumour cells, immune occurs in lymph nodes, where naïve T cells presented with cancer-specific antigens are activated (see Section 1.3.3). Immune tolerance can be induced by suboptimal cross-presentation of cancer-specific antigens by dendritic cells (DCs), which triggers apoptosis of T cells rather than activation and prevents the development of a cancer-specific T cell population. In addition, a significant proportion of the antigens produce by the tumour are similar to the endogenously expressed protein. Consequently, mutated peptides which are not distinct enough from the corresponding canonical peptides will be tolerised, similarly preventing the activation of cancer-specific T cells.

The second class of escape mechanisms does not rely on tumour cell intrinsic features but instead on secretion or expression of molecules by tumour cells that influence the balance of the tumour microenvironment (TME). These molecules include signal transducer and activator of transcription 3 (STAT3), which impairs DC maturation, indoleamine-pyrrole 2,3-dioxygenase (IDO), which inhibits T cell responses, transforming growth factor beta (TGF-β) which recruits regulatory T cells or factors that recruit cells that over time will encapsulate the tumour and keep immune cells away (Weaver & Murphy, 2016).

Throughout their development, tumours explore these strategies to survive and remain under control as long as the immune system is responsive enough. Cancer arises when a combination of these escape mechanisms eventually breaks the equilibrium.

1.2.3 The tumour microenvironment at diagnosis

The improvement of immune data acquisition and analysis has enabled to characterise the TME in many ways which can be broadly organised under three major high level categories (D. S. Chen & Mellman, 2017): immune-desert, immune-excluded and immune-inflamed. The first two categories indicate that the immune response presents massive defects which makes it incomplete or virtually inexistent

at the time of immune Escape. These phenotypes are hard to reconcile with the immunoediting continuum and are currently failing to respond to immunotherapy.

This thesis will focus on the third category which has been studied extensively for the past 10 years, following the exponential development of technologies such as fluorescence-activated cell sorting (FACS), immunohistochemistry (IHC), immunofluorescence (IF), mass spectrometry, RNA sequencing (RNA-Seq) and analytic tools dedicated to the analysis of the immune compartment of the TME. The main immune cell populations present in the TME of cancer patients are CD4 T cells, CD8 T cells, Regulatory T cells, NK cells, NKT cells, γδ T cells, DCs and macrophages (Whiteside, 2008).

The past 30 years have only confirmed that T cells are absolutely central to the immune response to cancer and are the key actors of cytotoxicity against tumours. Their behaviour is highly influenced by all compartments of the TME and understanding the role of each of them is of critical importance. This thesis will focus on T cells and describe the anti-tumour immune response from their perspective. In 2017, Chen & Mellman introduced the cancer immune set point concept which takes the view that the key players are cancer specific cytotoxic T cells which receive both stimulatory and inhibitory signals from the TME. This interaction can be represented by the following symbolic quantity (adapted from Chen & Mellman, 2017):

$$\sum \left(TCR_{affinity} \cdot TCR_{frequency}\right) \cdot \left[\int (F_{stim}) - \int (F_{inhib})\right]$$

Although this will be extensively detailed and discussed in Section 1.3, it should be mentioned here that the same TCR can be expressed by multiple T cells and that a given T cell normally expresses one TCR. In particular, the relative cellular prevalence of each intratumoural TCR is variable, with some TCRs shared by a few T cells only and some TCRs shared many T cells.

At a given point in time, the above quantity symbolically measures the strength of the TCR signalling (sum of all MHC-peptide-TCR complexes multiplied by corresponding abundance of TCR clone) weighted by the differential between

stimulatory and inhibitory forces in the TME at that time. This model is significantly less complex than the reality of cell to cell interactions in the TME but attempts to capture the core of the signal.

The following thought experiment helps to visualise the meaning of this quantity: if we could fix the TCR signalling component and change the values of F_{stim} and F_{inhib}, that would correspond to a system where the intratumoural TCR repertoire and corresponding antigenic landscape do not change. In that scenario, increasing the F_{stim}/F_{inhib} ratio would indeed intuitively strengthen the cytotoxic activity of T cells. Conversely, at a fixed influx of F_{stim} and F_{inhib}, a greater number of specific T cell clones would similarly correspond to greater killing capabilities of tumour-infiltrating lymphocytes (TILs).

This simplistic model is helpful to conceptualise how immunotherapies can act on the cytotoxic activities of intratumoural T cells, with adoptive cell therapy acting primarily on the left component and immune checkpoint blockage on the right component. In reality, the picture is much more complex, which will be discussed in Section 1.5.

Whether one chooses to describe the immune response to cancer from this modern and abstract perspective or rely on more seminal experimental work like the RAG knock-out experiments performed in the early 2000s (see Section 1.1.1), the specificity of TILs, relying on the T cell receptor, stands out as the central pillar of anti-tumour cytotoxicity.

The next section will introduce the concept of the TCR repertoire, explain how next-generation sequencing (NGS) enabled its measurement and description and what is our current understanding of its importance in describing the immune response to cancer.

1.3 The T cell receptor repertoire

1.3.1 Introduction

The mechanisms of Elimination, Equilibrium and Escape described above rely on specific subsets of T cells, corresponding to advanced stages of differentiation within the broad spectrum of T cell phenotypes. This results from a cascade of events that transforms precursor T cells into tumour specific T cells.

Briefly, precursor T cells (CD4/CD8 double-negative and lacking TCR expression) are generated from hematopoietic stem cells (HSCs) in the bone marrow before they migrate to the thymus where they commit to the T-cell lineage and where the TCR re-arrangement process is initiated (Weaver & Murphy, 2016). At this stage, the T-cell lineage can be broadly split between αβ T cells, γδ T cells and innate-like T cells. This thesis focuses on the former compartment, although the other two will be touched upon in Chapter 5.

The αβ TCR is encoded by a set of homologous genes (V, D and J genes) and is assembled by somatic recombination into two membrane bound protein chains that both include three hypervariable loops called complementary-determining regions (CDR) 1,2 and 3 (Figure 1.3). The VDJ recombination process is able to generate an incredible amount of diversity, estimated at 10^{13} unique possible TCR re-arrangement obtainable in a given individual (Miho et al., 2018). This diversity is predominantly captured by the α and β CDR3 sequences as they span across the junctional region of the TCR re-arrangement which contains insertions and deletions in addition to the germline DNA sequences (Attaf, Huseby, & Sewell, 2015; Weaver & Murphy, 2016).

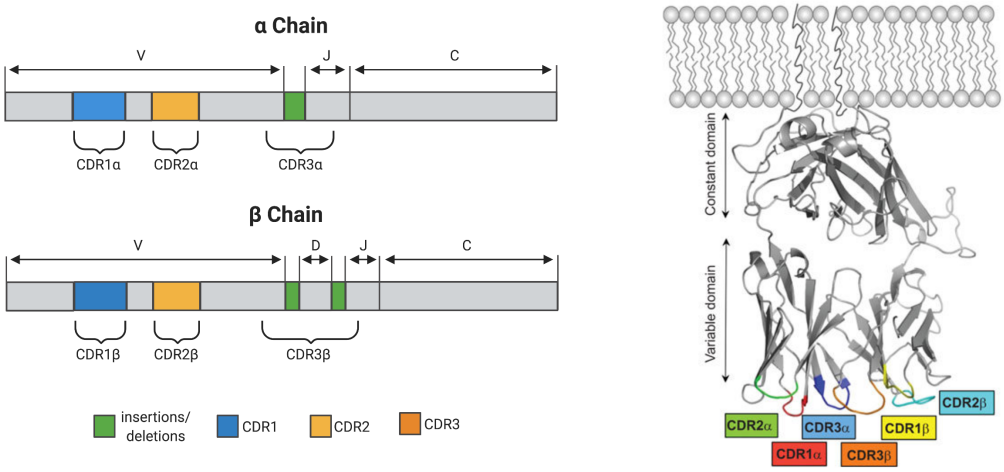

Figure 1.3 TCR gene and protein structure

Adapted from "T Cell Receptors (Alpha and Beta Chains)" by BioRender.com (2021). Retrieved from https://biorender.com/biorender-templates. Adapted from Attaf, Huseby and Sewell, 2015. CDR1 and CDR2 are entirely encoded in the germline V genes, whereas CDR3 lies at the junction between the rearranged V and J segments (TCR-α) and V, D and J segments (TCR-β).

From this point onwards, unless stated otherwise, "TCR" and "re-arrangement" will be used interchangeably to refer to the αβ TCR.

1.3.2 TCR repertoire diversity in the thymus

10^{13} unique TCRs could not possibly fit in the human body, which approximately contains 10^{11} T cells (Zarnitsyna, Evavold, Schoettle, Blattman, & Antia, 2013) but this is not the sole factor limiting the diversity of the TCR repertoire. Indeed, the number of unique TCRs in a given individual is currently estimated at 10^{8}-10^{9} (Mora & Walczak, 2019). Although randomness is an important part of VDJ recombination, the list of all TCR present in an individual and their cellular prevalence also called TCR repertoire is driven by precise mechanisms and is far from being a uniform sample of 10^{8}-10^{9} million out of 10^{13} (Attaf et al., 2015).

In the thymus, the β chain is the first one to re-arrange, leading to an intermediate stage of thymocytes harbouring a pre-TCR that signals the re-arrangement of the α chain, together with CD4 and CD8 expression and proliferation. Even at this early stage of double positive T cells, the diversity of TCRs is restricted and diverts from pure randomness. In 2012, Li et al. demonstrated in a mouse model that the β TCR

repertoire of thymocytes is systematically biased towards preferential usage of certain V and J genes, which resulted in a significant inter-animal overlap of β TCRs (H. Li et al., 2012). Although the mechanisms underlying this bias are still mostly undescribed, they can be summarised by the "accessibility hypothesis": the transcription machinery does not have equal access to all V and J segments (Ndifon et al., 2012).

Once double–positive thymocytes express fully re-arranged αβ TCRs on their surface, they undergo a process called thymic selection that filters out 99% of thymocytes together with shaping the TCR repertoire further, in a deterministic manner. This process relies on the interaction between TCRs and self-peptides presented on MHC molecules by antigen-presenting cell, as mentioned in Section 1.2.2. In the 1990s, several groups described that both the peptides themselves and the MHC molecules in which they fit impact the TCR repertoire obtained post selection in the thymus (Anderson, Partington, & Jenkinson, 1998; Ashton-Rickardt, Van Kaer, Schumacher, Ploegh, & Tonegawa, 1993; Bevan, 1997; Hogquist, Gavin, & Bevan, 1993; Messaoudi, Patiño, Dyall, LeMaoult, & Nikolich-Žugich, 2002; Nikolic-Zugic & Bevan, 1990).

In summary, TCRs that do not bind or bind too weakly to self MHC-peptide complexes will be eliminated by neglect and TCRs that bind too strongly to such complexes will be eliminated by negative selection. Conversely, double positive thymocytes harbouring TCRs with intermediate binding affinity (not too weak, not to strong), are positively selected and differentiate into single positive mature CD4 and CD8 T cells. This binding affinity is determined both by the peptide being recognised and the MHC molecule that presents it as they both interact with the TCRs. MHC molecules are encoded by highly polymorphic genes, also called human leukocyte antigen (HLA) genes, and are organised in two major classes (I & II) that determine the type of T cells they can interact with. Indeed, the non-polymorphic part of the MHC class II protein binds to CD4 whereas the non-polymorphic part of the MHC class I protein binds to CD8. In parallel, the polymorphic part the MHC molecule, which differs between individuals, restricts the pool of peptides that can fit into the MHC cleft, hence restricting the corresponding pool of TCRs that will be able to bind to the resulting complex (Weaver & Murphy, 2016).

Together, these events in the thymus apply a harsh deterministic filter onto the potential list of all possible re-arrangements one individual can generate. Nonetheless, the resulting TCR repertoire remains highly diverse and optimised to recognise a broad spectrum of foreign peptides, which are encountered by naive T cells as they exit the thymus into the periphery.

1.3.3 TCR repertoire diversity in the periphery

When naive T cells exit the thymus, they encounter specific antigens in secondary lymphoid tissues, presented by dendritic cells in most cases. This encounter initiates activation, differentiation and proliferation (Weaver & Murphy, 2016). Several studies in animal models and human cohorts have demonstrated that such encounters skew the naive TCR repertoire towards expansion of antigen experienced clonotypes (Figure 1.4). As a result, TCR repertoire diversity is at its highest in the naive compartment and diminishes in more differentiated T cell phenotypes (Oakes et al., 2017; Qi et al., 2014; Rudd, Venturi, Davenport, & Nikolich-Zugich, 2011).

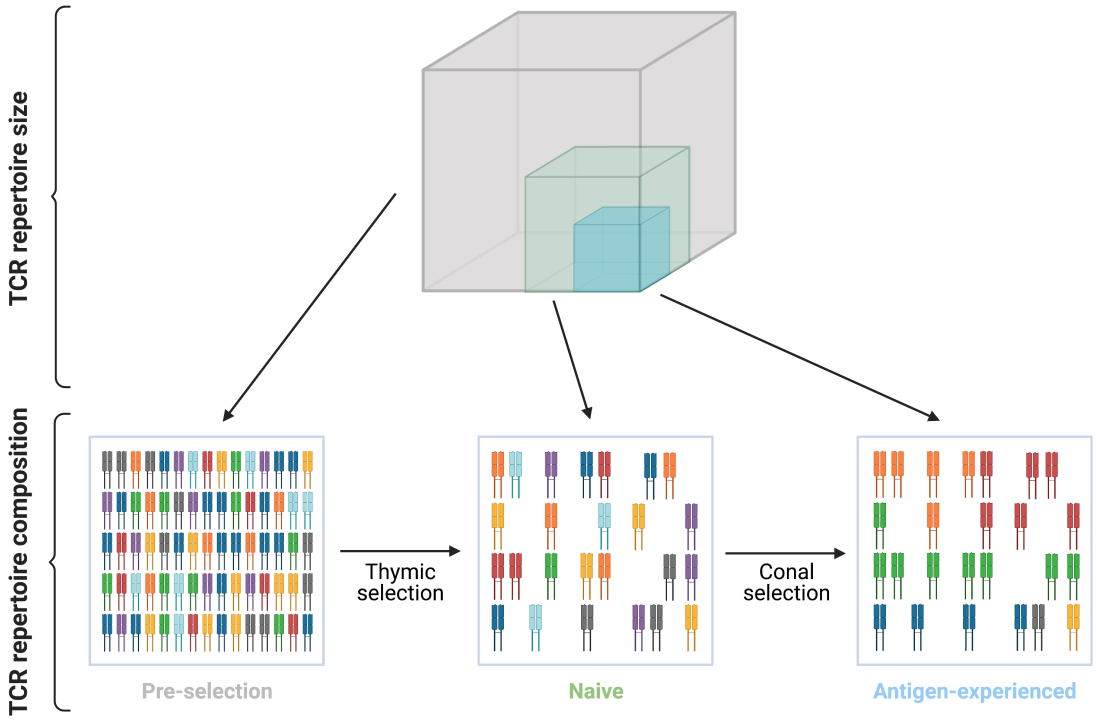

Figure 1.4 Size and composition of the pre-selection, naive and antigen experienced repertoires

Adapted from Attaf, Huseby and Sewell, 2015. Created with BioRender.com. TCR diversity is greatest in the pre-selection repertoire (grey). Positive and negative selection in the thymus purges the pre-selection repertoire of most specificities, creating a peripheral naive repertoire that is substantially less diverse (green). In the periphery, antigen exposure further narrows the repertoire over time leading to clonal expansion of antigen-specific populations (blue).

Repertoire skewness is largely driven by the environment that challenges one's immune system and defines the antigenic landscape that is presented to the TCR repertoire (Brodin et al., 2015). In 1994 Davey et al. demonstrated that monozygotic twins have almost identical repertoires at birth but drastically separate ones later in time (Davey, Meyer, & Bakke, 1994). TCR repertoires are also highly heterogeneous between tissues or conditions in a matched individual (Newell & Becht, 2018), which further illustrates that the cellular environment of T cells impacts the composition of the TCR repertoire.

As different environments lead to different repertoires, one could hypothesise that similar environment would lead to similar repertoires. This hypothesis was investigated in the context of HLA-matched individuals, which limits the impact of the

genetic background on the TCR repertoire. In such context, shared V and J gene usage and even shared CDR3 sequences was observed across people infected with Epstein-Barr virus (EBV) (Miles et al., 2010), human immunodeficiency virus (HIV-1) (H. Chen et al., 2012) or T large granulocyte leukaemia patients (Clemente et al., 2013). In addition, certain clonotypes have been observed as enriched and shared in the general population, agnostic of HLA haplotype, associated with T cell populations with reduced repertoire diversity such as mucosal associated invariant T (MAIT) cells or NK T cells (Attaf et al., 2015).

Even if similar environments have been linked to similarities in the TCR repertoire, the picture appears to be much more complex. As an example, in 2007, Yu et al. presented a longitudinal case study of monozygotic twins infected with the same HIV-1 variant yielding to several concordant epitopes. Despite the exceptional matched setting, the corresponding TCR repertoires were distinct between individuals, indicating that preferential TCR expansion was not fully determined by antigenic stimulation (Yu, Lichterfeld, Williams, Martinez-Picado, & Walker, 2007).

Together, these studies demonstrate that the TCR repertoire is rich, dynamic and complex and that although patterns can be predicted and observed in specific contexts, the rules that determine the shape of the TCR repertoire in health and disease remain incompletely characterised. In particular, the inherent stochasticity of the TCR re-arrangement process implies that not all features of a given repertoire can be attributed to specific events. However, the ability to distinguish between the random part and the deterministic part of the TCR repertoire is achievable, provided that the right tools are available and the right analyses are performed.

1.4 Quantifying the TCR repertoire

1.4.1 Introduction

In the 1980s, the discovery of the first TCR specific DNA sequences (Yanagi et al., 1984) was accompanied by the development of techniques to robustly and systematically derive metrics from the TCR repertoire which were primarily applied to study haematological malignancies like acute lymphoblastic leukaemia (Bertness, Kirsch, Hollis, Johnson, & Bunn, 1985; Korsmeyer et al., 1983). These techniques

involved southern blotting of known TCR genes, targeted with dedicated probes, and formed the basis of a technology that has since then enabled a much deeper and more complete description of the TCR repertoire (Mahe, Pugh, & Kamel-Reid, 2018).

Later, different techniques such as flow cytometry and CDR3 spectratyping were developed and applied to the analysis of the TCR repertoire (Six et al., 2013), which started to depict important metrics such as TCR counts or CDR3 lengths distribution. Undoubtedly however, the advent of TCR repertoire analysis was enabled by the development of NGS sequencing, also known as high-throughput sequencing, and only within the last two decades did we start to scratch the surface of the immense amount of information that is hidden inside TCR repertoire data (Heather, Ismail, Oakes, & Chain, 2018).

Today, most TCR repertoire data, including the data that will be discussed in this thesis, is obtained by high-throughput sequencing. Although all pipelines are not identical, they all include three major stages: library preparation and sequencing, low-level processing and high-level processing **(**Figure 1.5).

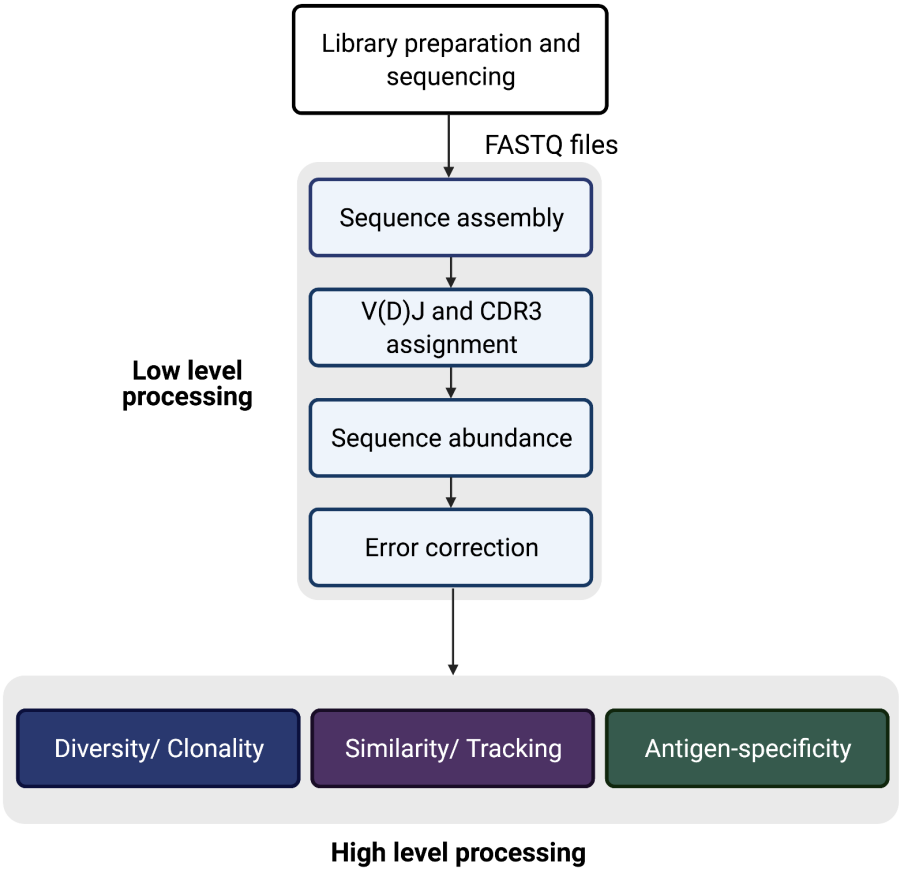

Figure 1.5 The main stages involved in the study of immune repertoires.

Adapted from "Data Processing Flow Chart" by BioRender.com (2021). Retrieved from https://biorender.com/biorender-templates. Adapted from Heather et al., 2018. The main stages involved in the study of immune repertoires: library preparation and sequencing, low-level processing and high-level processing.

The work presented in this thesis obviously relies on library preparation and sequencing steps together with low-level processing, but it primarily comprises of high-level processing. Indeed, all the library preparation and sequencing as well as most of the low-level processing of the data presented in this thesis were performed by colleagues, whilst the implementation of existing tools or development of novel high-level processing analyses discussed in Chapter 3, Chapter 4 and Chapter 5 are my own work. Although they will not be discussed further in the result chapters of the thesis, a brief description of the first two stages will be given in this section and specific technical details relating to each project can be found in Chapter 2.

1.4.2 Library preparation, sequencing and low-level processing

Briefly, all TCR sequencing protocols start from either DNA or RNA. Following nucleic acid extraction, recombined TCR genes are amplified using multiplex polymerase chain reaction (PCR) with a set of primers specific to V and J genes. After amplification, sequencing is performed, most of the time through the Illumina platform (Bentley et al., 2008), which results in the generation of a *.fasta* (FASTA) or *.fastq* (FASTQ) file (Heather et al., 2018).

After sequencing, several solutions exist to retrieve the list of TCR re-arrangements present in a sample and their corresponding abundance (number of reads) from the FASTQ files. This task often comprises of two steps: V, D and J gene assignment and CDR3 identification and translation.

The gene assignment step requires a reference database listing V, D and J sequences to be looked for in the FASTQ file. The most widely used one is the IMGT/GENE-DB database, which contains annotations for V, D and J genes across several species (Giudicelli, Chaume, & Lefranc, 2005). Mapping reads to a specific annotated V, D or J transcript is a well described task that generated several solutions over the past decade, which vary in efficacy but all perform well (Heather et al., 2018). Once V, D and J genes are identified, reconstructing the CDR3 region is a fairly trivial task consisting in identifying where it starts and whether the translated amino acid sequence is productive (in frame with respect to the start of the V gene leader sequence and not containing premature stop codon) or not.

1.4.3 High-level processing

Low-level processing usually outputs one file per sample, which can vary in format but is systematically organised as a list of TCR re-arrangements and matched count values, which can be aggregated at different levels (gene level, nucleotide level, amino acid level). Such files contain thousands to millions of different re-arrangements (Gerlinger et al., 2013), which are mostly distinct between individuals and often distinct between different samples of the same individual. As mentioned in Section 1.4.1, high-level processing or the mining of this rich and complex data has seen significant improvement in the recent years but, although some groups have

attempted to organise such analysis into publically available tools (Dmitriy V. Bagaev et al., 2016; ImmunoMind, 2019; Kaplinsky & Arnaout, 2016; Nazarov et al., 2015; Shugay et al., 2015), no gold standard has yet emerged (Rosati et al., 2017).

A number of summary metrics are however recurrent in TCR-Seq data analysis and commonly accepted as informative and useful. The simplest metrics that can be derived resemble what could be obtained with the early techniques generating TCR repertoire data mentioned in Section 1.4.1. They include V and J gene usage, total and unique TCR counts and distribution of CDR3 lengths.

Diversity/ Clonality

A more elaborate concept that is very commonly looked at in TCR repertoire data is diversity. TCR-Seq data often approximates to a power-law distribution (Oakes et al., 2017), meaning that most re-arrangements are very infrequent, a large proportion are singlets (only seen once in the repertoire) but the bulk of the total re-arrangements is occupied by a few sequences which are individually very abundant. The manner in which the data is distributed between infrequent or frequent clones is captured by diversity or clonality scores. A repertoire in which the frequency of TCR re-arrangements is evenly distributed will generate a high diversity score or a low clonality score, whereas a repertoire in which a small number of unique re-arrangements occupies a large proportion of total counts will generate a low diversity score or high clonality score. The assumption is that antigen experienced TCR repertoire will be more clonal (high clonality score) than non-antigen experienced ones. Consequently, these scores are used to infer antigenic exposure in a sample: high clonality scores should indicate the presence of antigen exposed T cells. Many different scores capturing diversity exist and are routinely applied to TCR-Seq data including Rényi indices, the Gini index, the Simpsons index, the Shannon entropy (Mora & Walczak, 2016) or the clonality index (Kirsch, Vignali, & Robins, 2015). These metrics capture the same idea and, in practice, strongly positively or negatively correlate to one another.

Similarity

Another important category of metrics aims to capture how two or more TCR-Seq samples compare, which enables to infer how the repertoire differs between

individuals or, more interestingly, how it changes in space and time within the same individual. This is particularly useful when additional information is known about the samples, such as comparing a tumour sample to an adjacent non-tumour sample or comparing different conditions of the same tissue, for instance blood samples of cancer patients before or after a specific treatment. These metrics fall under the concept of similarity and include the Jaccard index, Pearson and Spearman correlation coefficients and the dot product.

Tracking

A sub-category of similarity consists in tracking a particular TCR or set of TCRs of interest across several samples. In such instances, the data is looked at on a per TCR basis and simply asks whether a particular re-arrangement is present or absent in a given sample and/or at which frequency. This is of particular interest in the management of haematological malignancies which was actually one the first applications of repertoire data analysis, as mentioned in Section 1.4.1. This will be further discussed in Section 1.4.5 and in Chapter 5.

Antigen-specificity

Finally, the last category of analysis that can be performed on TCR-Seq data constitutes the holy grail of high-level processing: determining the specificity of a given TCR. A first approach is to generate large databases of known TCR specificities, which can be obtained through multimer sorting experiments for instance. Two such databases exist, VDJdb (D. V. Bagaev et al., 2020) and McPAS-TCR (Tickotsky, Sagiv, Prilusky, Shifrut, & Friedman, 2017), and are already gathering a large amount of TCRs, which is however still very limited compared to the titanic number of possible unique TCRs.

A second category of approaches aimed at determining the specificity of a given TCR consists in modelling the binding affinity of TCR-peptide-MHC complexes (Borrman, Pierce, Vreven, Baker, & Weng, 2020; Moris et al., 2020; Springer, Besser, Tickotsky-Moskovitz, Dvorkin, & Louzoun, 2020; Tong et al., 2020). Building such models often requires training an algorithm on a dataset of known specificities such as the ones described above. Although the performance of such methods is

improving on the recurrent publically available benchmarking datasets, it remains unclear how they perform outside the training setting.

Lastly, a less direct approach yet significantly less constrained than the methods described above is the grouping of similar CDR3 sequences based on amino-acid motif usage (N. Thomas et al., 2014). The idea behind this approach, which has been demonstrated in several instances (Glanville et al., 2017; H. Huang, Wang, Rubelt, Scriba, & Davis, 2020), is that different TCRs harbouring very similar CDR3 regions are likely to be recognising the same peptide-MHC complex. Several variations exist, and implement more or less stringent grouping criteria (Pogorelyy et al., 2019). With this approach, the corresponding peptide is unknown, which makes it less powerful than the ones mentioned above are attempting in theory but is in practice convenient to derive metrics of interest (see Chapter 3, Chapter 4 and Chapter 5).

1.4.4 Challenges

A number of challenges are associated with every step of TCR repertoire data generation and analysis (library preparation, sequencing, low-level processing and high-level processing). Very early in the process, sequencing errors can be introduced because of the error rate of the polymerases used in the PCR amplification or the DNA sequencing reaction itself. These errors can artificially inflate the diversity of a sample by introducing novel TCRs in silico that are not actually present in the sample. The source of these technical errors will vary depending on the technical specificities of a given protocol and several solutions are currently implemented. These include using the Phred quality score of sequencing reads, filtering out low frequency TCRs and the use of UMIs (Heather et al., 2018). This matter will not be discussed further in this thesis but is worth keeping in mind.

Another level of limitations to TCR-Seq data analysis, which is common to all NGS data analyses, is the representativeness of a sample. As mentioned above, the immense diversity of the TCR repertoire results in a single sample containing thousands to millions of unique TCR sequences with virtually no overlap between individuals and very little overlap between samples of the same individual. In addition, in several instances such as cancer TCR-Seq analysis, sampling the tissue of interest itself presents representativeness issues. Firstly, the tissue of interest

might be difficult to access hence it might be preferable to sample more accessible tissue, such as blood, and try to extrapolate findings from one compartment to the other (Lucca et al., 2021). This is very attractive, in particular for the development of clinical biomarkers or any analysis aiming to be applicable to routinely acquirable samples, which can be quite different from what an academic project can offer. However, determining how well does the TCR repertoire of a certain tissue depict the immune response occurring in another tissue is certainly not straightforward and will be discussed in Chapter 3, and Chapter 4.

Secondly, even when the tissue of interest can be accessed, the sampling process is often unable to sample the entire tissue. For instance, in the context of the same tumour mass, biopsies of different parts of the tumour can be extremely different. How one sample is representative of another sample of the same tissue is not trivial to determine, which will be discussed in the specific context of NSCLC in Chapter 3.

1.4.5 Applications

For the past 20 years, the interest for TCR-sequencing data analysis has only grown, together with the number of fields in which its clinical relevance has been demonstrated. One of the first applications of TCR-Seq, still very active today, is the measurement of minimal residual disease (MRD) in haematological malignancies. Indeed, diseases such as acute (D. Wu et al., 2012) or chronic (Logan et al., 2011) lymphoblastic leukaemia are associated with the emergence of aberrant clonotypes, sometimes productive but in many instances non-productive (Allam & Kabelitz, 2006). Consequently, measuring the prevalence of these conventional or trans TCRs via TCR-Seq is a powerful biomarker in the clinical management of such diseases (O'Connor et al., 2018). Chapter 5 will showcase repertoire sequencing generated in this context for a cohort of childhood B-ALL patients and show that such data contains a rich and complex signal that goes beyond MRD assessment.

Although they will not be discussed in detail in this thesis, it should be mentioned that the main applications of TCR-Seq data analysis include the characterisation of a number of non-infectious and infectious diseases such as inflammatory and autoimmune disorders (Chapman et al., 2016; Goronzy, Zettl, & Weyand, 1998),

infection (Heather et al., 2015; Wisskirchen et al., 2017) and immune responses following vaccination (Ademokun et al., 2011; Qi et al., 2016).

1.5 The TCR Repertoire in cancer

1.5.1 TCR clonality as a biomarker

In the last decade, numerous studies have utilised TCR sequencing data analysis as a measure of the immune response to cancer, especially in the context of checkpoint blockade therapy, as eluded to in Section 1.2.3. Most of these studies used repertoire data derived from blood samples and attempted to define TCR metrics that could be robustly used as biomarkers of response to immunotherapies.

In 2014, several studies investigated how TCR-Seq could inform response to anti-cytotoxic T-lymphocyte protein 4 (CTLA-4) treatment, particularly in the context of melanoma. Robert et al. described a global effect of tremelimumab on the blood TCR repertoire of 21 metastatic melanoma patients. When compared to healthy control, patients exhibited increased diversification (measured by Shannon entropy) of the blood TCR repertoire post treatment (Robert et al., 2014). In a cohort of metastatic castration-resistant prostate cancer and metastatic melanoma patients, Che et al. repeated this observation and also found that it was the maintenance of TCRs expanded pre-treatment in the post-treatment blood that associated with response (Cha et al., 2014). Kvistborg et al. added to the complexity of this picture by longitudinal monitoring of the reactivity against 145 melanoma associated epitopes in a cohort of 40 patients treated with ipilimumab. In this context, the authors confirmed that maintenance of pre-existing specificities was occurring but argued that the striking increase in the number of specificities detected post-treatment was probably a more important mechanism of response to anti-CTLA-4 therapy (Kvistborg et al., 2014). In 2017, Oh et al., published a study on immune related adverse events (IRAEs) in cohort of metastatic castration-resistant prostate cancer patients treated with Ipilimumab. The increase of the diversification of the repertoire could be preferentially linked to the patients with IRAEs. Together, these findings indicate that the interpretation of how TCR-Seq can inform the response to anti-CTLA-4 treatment is not trivial and probably comprises of several layers that are yet to be de-convoluted (Oh et al., 2017).

In parallel, similar efforts were conducted in an attempt to understand the clinical response to therapies targeting the programmed cell death 1 (PD-1)/programmed death ligand 1 (PD-L1) axis. In the intratumoural setting, Tumeh et al. showed in 2014 a pre-treatment association between response and more clonal TCR repertoire in a cohort of 46 metastatic melanoma patients treated with pembrolizumab (Tumeh et al., 2014). In 2018, Amaria et al. repeated this observation in a cohort of 23 high-risk resectable melanoma patients treated with nivolumab alone or in combination with Ipilimumab. In addition, they demonstrated an association between response and the maintenance of pre-treatment expanded TCRs in the tumour (Amaria et al., 2018).

Two studies highlight even further these discordant observations between TCR measurements and response to anti-CTLA4 treatment or anti-PD-1/PD-L1 treatment for melanoma patients. In 2017, Roh et al. demonstrated an association with more clonal intratumoural repertoires and response to PD-1 blockade but not CTLA-4 blockade (Roh et al., 2017). In 2018, Levesque et al. showed that the same diversity score had the opposite predictive power depending on the treatment type: melanoma patients responding to anti-PD1 treatment had less diverse repertoires in the blood whereas melanoma patients responding to anti-CTLA-4 treatment had more diverse repertoires in the blood (Hogan et al., 2019). Together, these studies in melanoma cohorts highlight the complexity of the TCR repertoire both in blood and in the tumour and demonstrate that even supposedly simple measures could have drastically different meanings depending on the context.

Since then, numerous studies have performed TCR sequencing in the context of immune checkpoint blockade (ICB) therapy for many different cancer types (Fairfax et al., 2020; Han et al., 2020; Naidus et al., 2021; Schalper et al., 2019; Zhao et al., 2019). The range of metrics that have been derived from such data and associated to clinical response is still very heterogeneous and lacks clarity. In addition to leveraging TCR-Seq data in order to develop biomarkers of response to treatment, there is a clear need to take a step back and to build a deep understanding of the biology that underlies these metrics. This is required to guide the design of

meaningful metrics and better interpretation of TCR-Seq data derived from cancer patients.

Chapter 4, which forms sections of a submitted manuscript under review, will show an example of integration of TCR-Seq data within a complex multi-omics cohort of nivolumab treated metastatic renal cell carcinoma patients.

1.5.2 Antigen specificity

As explained in Sections 1.1 and 1.2, the central feature of the cytotoxic response to cancer is the specificity of infiltrated T cells. The amount of reactivity to cancer within the intratumoural TCR repertoire is highly variable across cancer types and across patients within the same indication as demonstrated in ovarian and colorectal cancer by Scheper et al. in 2019 (Scheper et al., 2019). Although it should be mentioned that there is evidence of a potential anti-cancer function of the non-cancer-specific TIL compartment (Danahy, Berton, & Badovinac, 2020; Erkes et al., 2017), the key challenge that this thesis addresses is the identification of cancer-specific TCRs and the ability to distinguish them from the rest of the repertoire, enriched for bystander T cells (Simoni et al., 2018).

Several studies have now demonstrated that such specificities are indeed detectable in TCR-Seq data. The Rosenberg group first demonstrated that neoantigen specific CD8 T cells could be isolated from the tumours of melanoma patients (Gros et al., 2014) via direct tetramer sorting. They subsequently showed that tumour specific TCRs could be also be obtained in silico by focusing on the most expanded intratumoural clones (Pasetto et al., 2016).

Another indirect way to link TCR-Seq data to specificity is to correlate certain TCR metrics to cancer intrinsic metrics, such as tumour mutational burden (TMB), which serve as a proxy for antigenicity of the tumour. In certain cancer types such as melanoma or lung cancers where non-synonymous mutations (NSMs) have been identified as an important source of tumour-specific antigens and TMB has been shown to correlate with several clinical features, it can be used to shed light onto specificity within TCR-Seq data. In 2017, for instance, Huang et al. showed an association between expanded TCR clones and tumour burden in patients with stage

IV melanoma (A. C. Huang et al., 2017). This concept will be extensively discussed in Chapter 3, in the context of NSCLC.

In addition, as presented in Section 1.4, several methods have been developed to indirectly characterise antigenic specificity within TCR-Seq data intrinsically, relying on amino acid motifs usage, and are starting to be applied to immuno-oncology. Deriving meaningful metrics from such analysis is challenging but can prove useful, which will be discussed in sections of Chapter 3, Chapter 4 and Chapter 5.

1.5.3 Spatial and temporal heterogeneity

As seen in Sections 1.1, 1.2 and 1.3, within a single individual, both the cancer and the TCR repertoire are highly dynamic in space and time, and the sequencing of a particular specimen at a particular point in time will provide an incomplete view, as mentioned in Section 1.4.4. In addition, the cancer and TCR repertoire of an individual are evolving interdependently and changes in one will most certainly induce changes in the other.

The metastatic setting is immediately relevant to the question of TCR repertoire spatial heterogeneity. In 2013, Emerson et al. showed in a small cohort of 5 metastatic ovarian cancer patients that there was a striking coherence of the intratumoural TCR repertoire across different tumour regions of the same patient compared to matched blood samples (Emerson et al., 2013). In contrast, in 2017 Reuben et al., highlighted a very heterogeneous TCR repertoire across metastatic sites of melanoma patients, with an observed proportion of overlapping clones below 8% on average (Reuben et al., 2017). Descriptions of the TCR repertoire in the metastatic setting were also conducted in glioblastoma (Feng et al., 2017), breast (Wang et al., 2017) and gastric cancer (Kuang et al., 2017), depicting a complex picture of both overlapping and private TCRs across metastatic sites.

Similar to the cancer genomic landscape, the intratumoural TCR repertoire has also been described to present heterogeneity within a single site, and not only across distant sites. In 2013, Gerlinger et al., demonstrated that distinct tumour regions harvested from the primary sites of a cohort of clear cell renal cell carcinoma patients systematically contained private TCRs and that the prevalence of TCRs not found

across all regions could reach up to 93%. Similar observations were made in oesophageal squamous cell carcinoma (M. F. Chen et al., 2016) and localised lung adenocarcinomas (Reuben et al., 2017). Chapter 3 will discuss a model showing that such heterogeneity is not random in a non-small cell lung cancer patients cohort.

Finally, as mentioned above, the temporal heterogeneity of the TCR repertoire also seems to contain important information about the immune response to cancer. Both maintenance and replacement of TCRs have been observed intratumourally and in the blood of cancer patients, between separate time-points. Depending on the setting, it is still unclear what signal is a direct consequence of treatment or a simple correlation. For instance, in 2019, Yost et al. described a clear TCR replacement phenomenon occurring in basal or squamous cell carcinoma patients treated with PD-1 blockade, which however did not associate with response or non-response (Yost et al., 2019). Chapter 4 will specifically discuss maintenance and replacement of TCRs upon anti-PD-1 treatment in a cohort of metastatic renal cell carcinoma patients.

1.5.4 Linking TCRs and T cell phenotypes

As briefly mentioned in Section 1.2.3, the TME is itself a very heterogeneous system and the T cell compartment alone contains a complex gradient of phenotypes which is still being actively characterised. This is a very important component of TCR repertoire analysis as the same TCR found on a regulatory CD4 T cell, a naive CD8 T cells or a cytotoxic CD8 T cell will have a drastically different role in the immune response to cancer. Cancer specific TCRs have been robustly associated to late differentiated CD8 phenotypes, often expressing PD-1 but this vision has been refined with the recent development of paired RNA and TCR single-cell sequencing.

In 2019, Li et al. showed in a cohort of 25 melanoma patients that the most expanded TCR clones, hypothesised as enriched for tumour specificity, were predominantly of the CD8 dysfunctional phenotype (displaying high levels of PD-1 expression) but also comprised of cytotoxic, memory, transitional and even naive like cells, indicating a gradient of phenotypes sharing the same expanded TCRs (H. Li et al., 2019).

In the context of nivolumab treatment, Yost et al. showed transition of CD8 T cells from one phenotype to another in basal and squamous cell carcinoma patients. In addition to exact TCR matches, the authors leveraged CDR3 amino acid clustering via GLIPH (Glanville et al., 2017) to show that frequent transition were observed between memory and activated states. As mentioned above, no association with response was observed and the phenotype(s) of T cells that respond to anti-PD1 therapy remain(s) to be fully determined (Yost et al., 2019).

Integrating transcriptomic information to TCR data at the single cell level is still a novel challenge which is seeing a growing number of attempts to normalise approaches and interpretation. This year, Zhang et al. proposed an interesting approach combining both CDR3 sequence based distances and transcriptomic distances when comparing single cells. They show that the two types of distances are highly correlated in healthy controls but not in tumour samples, confirming active transition between T cell phenotypes in the TME (Zhang, Xiong, Wang, Liu, & Wang, 2021).

Such analysis is only beginning to reveal that combining TCR-Seq and RNA-Seq at the single-cell level contains more information than the summation of both and that elaborate integration approaches has the potential to unravel novel biology rather than simply confirming what has already been described. Chapter 5 will present a project that aims to integrate scTCR-Seq and scRNA-Seq in order to reveal novel specificities against cancer in a cohort of childhood B-ALL patients.

1.6 Aims of the thesis

In this chapter, I gave a brief view of key concepts of immuno-oncology, namely the evidence for the existence of an endogenous immune response to cancer in many settings and the evidence that tumour cells often escape this response. After highlighting that the recognition of cancer cells by T cells through the MHC-peptide-TCR interaction is at the heart of this immune response and several immune-escape mechanisms, I introduced the notion of the TCR repertoire. Finally, I detailed the different aspects of the TCR repertoire that can be measured today, the techniques

that exist and the main applications of TCR repertoire data analysis, including its importance in cancer research.

As mentioned in the previous section, in cancer research, TCR-Seq data has already proven useful as a biomarker of response to treatment in certain contexts and has already been used to unravel numerous insights into the immune response to cancer. However, a proportion of the studies mentioned above fail to converge towards a shared understanding of how TCR-Seq data behaves and how to interpret it. Whether it is a cause or a consequence of the previous statement, a joint observation is that analytical gold standards dedicated to TCR-Seq have not yet emerged. There is currently no tool dedicated to the analysis of TCR-Seq data that is equivalent to what DESeq2 (Love, Huber, & Anders, 2014) or edgeR (Robinson, McCarthy, & Smyth, 2010) are to RNA-Seq or to what Seurat (Satija, Farrell, Gennert, Schier, & Regev, 2015) is to scRNA-Seq.

The lack of interest for developing such tools is partially explained by the fact that TCR-Seq is, today, mainly considered as a technique that complements other techniques, such as RNA-Seq, and helps confirm findings rather than discover novel biology. In this thesis, I aim to demonstrate that TCR-Seq can be used as a powerful discovery tool and can be the central driving element in complex multi-omics collaborative studies. In particular, just like differential expression analysis can capture an insightful signal in RNA-Seq data or the expression profile of a particular cluster of cells in scRNA-Seq data can identify cytotoxic T cells, there are questions in immuno-oncology that preferentially benefit from leveraging immune repertoire data.

In this thesis, I analyse TCR-Seq data in three different clinical oncology scenarios to ask the following fundamental questions:

1. Can cancer-specific TCRs be identified and distinguished from non-cancer-specific TCRs?
2. Can the TCR be used to link T cell specificity to T cell phenotype?
3. What features of the intratumoural TCR repertoire relate to clinical features of cancer?

The three separate result chapters of this thesis present the metrics that can be extracted with which to address these questions in TCR-Seq data in a range of situations, from clean, dedicated TCR-Seq data to noisy, repurposed bulk repertoire data. Some metrics are common to all projects and some are tailored to data-specific requirements, showcasing the versatility of TCR-Seq data analysis. In each project, TCR-Seq data analysis revealed key insights that were further integrated with matched –omics data in order to identify and characterise features consistent with antigen exposure in NSCLC, metastatic ccRCC and childhood B-ALL.

Chapter 2. Materials and Methods

In this chapter, I outline the bioinformatics and experimental procedures adopted throughout the thesis. As mentioned in Section 1.4.2, unless stated otherwise, all the sample collection, experimental procedures, library preparation, sequencing and low-level processing were performed by colleagues. Details of contribution are given at the beginning of Chapter 3, Chapter 4 and Chapter 5**.** All high-level bioinformatics analyses and visualisations are my own work and were carried out in the R statistical environment, unless otherwise specified.

2.1 Datasets

2.1.1 Lung TRACERx cohort

All patients within the study presented in Chapter 3 were recruited to the Tracking Cancer Evolution through therapy (Rx) (TRACERx) study (Research Ethics Committee no. 13/LO/1546). Patients with sufficient RNA from at least two tumour regions were selected for the TCR sequencing study. Samples from adjacent non-tumour lung and peripheral blood mononuclear cells (PBMCs) taken at the time of resection as well as a number of follow-up PBMC samples were sequenced whenever available. All tissue specimens were reviewed by a lung pathologist before being selected, as previously described (Jamal-Hanjani et al., 2017). For all analyses, data was locked on April 2018.

TCR-sequencing and pre-processing

TCR α-chain and β-chain sequencing was performed by utilising whole RNA extracted from NSCLC tumour samples and non-tumour lung tissue or from cryopreserved PBMC samples using the Chain protocol extensively described in Oakes et al., 2017. TCR identification, error correction and CDR3 extraction was performed using the Decombinator suite (Peacock, Heather, Ronel, & Chain, 2020; N. Thomas, Heather, Ndifon, Shawe-Taylor, & Chain, 2013).

DNA&RNA-sequencing and pre-processing

Whole-exome sequencing and subsequent non-synonymous mutation calling of multi-region tumour specimens and matched germline samples derived from whole

blood were performed as detailed in Jamal-Hanjani et al., 2017. All mutations classified as non-synonymous variants plus all mutations that introduced or removed a stop codon or introduced a frame shift were selected for further analysis. As detailed in Joshi et al. 2019, all mutations were classified as present or absent based on a 10% variant allele frequency (VAF) threshold and subsequently classified as ubiquitous if present in all regions of a tumour and regional if absent from at least one region.

RNA was extracted by using a modification of the AllPrep kit (Qiagen) as described in Jamal-Hanjani et al., 2017. The prepared libraries were size-selected and multiplexed, and underwent quality control before paired-end sequencing. FASTQ data underwent quality control and were aligned to the hg19 genome with STAR (Dobin et al., 2013). Transcript quantification was performed using RSEM (B. Li & Dewey, 2011) with default parameters.

RNA sequencing of sorted TIL subsets and pre-processing

The BD FACSAria II flow cytometer was used to sort $CD8^+$ TILs from NSCLC samples obtained from CRUK0017, CRUK0024 and CRUK0069. $CD8^+$ TILs were sorted into two populations: (i) $CD45RA^-CCR7^-CD57^-PD\text{-}1^+$ cells and (ii) all other $CD8^+$ T cells ('not gate'). All cells were sorted into TRIzol followed by phenol-chloroform RNA extraction. Subsequent RNA extraction and sequencing was performed as detailed in Joshi et al., 2019.

To search for tumour-expanded TCRs within the RNA-Seq data, a tailor-made script in R was used. Briefly, the algorithm gives an estimate of the proportion of the TCRs present in a particular sample that can be attributed to the set of expanded TCRs in the tumour as a whole. More details can be found in Joshi et al., 2019.

scRNA-sequencing of neoantigen-reactive T cells and pre-processing

$CD8^+$ T cells targeted against a clonal neoantigen (arising from the mutated *MTFR2* gene) in NSCLC tumour regions derived from patient L011 were previously identified by the Quezada group and the Swanton group (McGranahan et al., 2016). The staining of neoantigen-reactive T cells was based on dual-fluorescent multimer

labelling, using a freshly thawed vial of cryopreserved TILs from the same patient. Multimer-positive and multimer- negative single $CD8^+$ T cells from NSCLC specimens were sorted and subjected to scRNA-Seq as detailed in Joshi et al., 2019.

The bioinformatic reconstruction of TCRs from single-cell RNA-Seq data was performed with a modified version of Decombinator (https://github.com/innate2adaptive/Single-Tag-Decombinator).

2.1.2 ADAPTeR cohort

All patients within the study presented in Chapter 4 were recruited to the study of Anti-PD1 (Nivolumab) Therapy as Pre- and Post-operative Therapy in Metastatic Renal Cell Cancer (ADAPTeR) (NCT02446860). ADAPTeR is a single-arm, open-label, phase 2 study of nivolumab therapy as pre-operative therapy in metastatic ccRCC. Response status was derived from the Response Evaluation Criteria in Solid Tumours (RECIST) score as detailed in Au et al., 2021. ADAPTeR was initially approved by NRES Committee London Fulham on 01/12/2014. ADAPTeR is performed in accordance with the ethical principles in the Declaration of Helsinki, Good Clinical Practice and applicable regulatory requirements. For all analyses, data was locked on December 2018.

For translational study sample collection, baseline tumour biopsy via appropriate guidance (ultrasound or computer tomography) at least three days and up to 14 days prior to starting nivolumab was obtained. Where possible, multiple regions of nephrectomy specimen were sampled, as well as image guided biopsy of regressing lesions or at disease progression either at site of progression or, if not possible, percutaneous primary renal tumour biopsy, prior to commencement of any subsequent treatment. Blood samples were collected at each tumour sampling time-point.

Autopsy samples from ADR001, ADR005 and ADR015 were obtained through the Posthumous Evaluation of Advanced Cancer Environment (PEACE) Study (NIHR 18422; NCT03004755), where samples are harvested within 48 hours from death.

TCR-sequencing and pre-processing

TCR α-chain and β-chain sequencing and processing through Decombinator were performed by utilising whole RNA extracted from tumour samples or from cryopreserved PBMC samples as described above for the lung TRACERx bulk TCR cohort.

DNA&RNA-sequencing and pre-processing

RNA was co-extracted from fresh-frozen tumour tissue using AllPrep DNA/RNA mini kit (Qiagen) as detailed in Au et al., 2021. As described above for lung, the prepared libraries were size selected and multiplexed, and underwent quality control before paired-end sequencing. FASTQ data underwent quality control and were aligned to the hg19 genome with STAR. Transcript quantification was performed by using RSEM with default parameters. Samples with less than 15,000 genes detected were excluded.

scRNA-sequencing and pre-processing

Tumour infiltrating lymphocytes from ADR001 and ADR013 were stained with CD3 and IgG4 antibodies for flow cytometry. Stained cells were FACS sorted as $CD3^{+}IgG4^{-}$ (40,000 cells) and $CD3^{+}IgG4^{+}$ (20,000 cells) for ADR001 and $CD3^{+}IgG4^{-}$ (50,000 cells) and $CD3^{+}IgG4^{+}$ (90,000 cells) for ADR013. FACS sorted cells were single cell sorted using the 10X Genomic machine. The sorted cells were processed using the 10X Genomic Chromium Next GEM Single Cell 5' Reagents Kit V2 (dual index) for 5'gene expression library construction and V(D)J library construction. The samples were sequenced on the NextSeq using the High Output Kit v2.5 (150 Cycles).

FASTQ files containing gene expression (GEX) and VDJ were demultiplexed using *cellranger mkfastq* (https://github.com/10XGenomics/cellranger). GEX reads were aligned to GRCh38 and counted using *cellranger count*, VDJ reads were aligned to cellranger's GRCh38 VDJ reference dataset using *cellranger vdj*. Expression matrices were analysed using the Seurat package version 3 (Stuart et al., 2019) and QC was performed as detailed in Au et al., 2021. 8382 $CD3^{+}IgG4^{-}$ and 10083

$CD3^+IgG4^+$ cells in ADR013; and 4648 $CD3^+IgG4^-$ and 3343 $CD3^+IgG4^+$ cells in ADR001 were retained after QC.

2.1.3 B-ALL cohort

Ethical approval was given for use of appropriately consented material from patients with B cell acute lymphoblastic leukaemia at the Great Ormond Street Hospital for Children (London, UK) and at the Royal Hospital for Children (Bristol, UK). All samples within the study presented in Chapter 5 were processed in the MRD laboratory at Great Ormond Street Hospital for Children and the Bristol Genetics Laboratory at Southmead Hospital, Bristol.

Repertoire sequencing and pre-processing

For the BIOMED-2 cohort, bone marrow mononuclear cell (BMMC) samples were processed using the protocol described by Bartram et al. in 2016, in which the IGH/TCR amplification step consists of a two stage PCR based on the BIOMED-2 family primers (Bartram et al., 2016; van Dongen et al., 2003).

In order to reconstruct the CDR3 region of IG and VDJ re-arrangements, FASTQ files were processed with MiXCR's *analyse amplicon* pipeline with default parameters (Bolotin et al., 2015).

All *.clonotypes.*.txt* files (*TRA, TRB, TRG, IGH, IGK, IGL*) but *.clonotypes.TRD.txt* (https://github.com/milaboratory/mixcr/issues/541#issuecomment-545963300) were pooled per sample and used for downstream analysis.

Single-cell RNA and VDJ-sequencing and pre-processing

For patients 913304, 1017078 and 1020374, freshly thawed vials of BMMCs were stained for cell surface markers at 4C in the dark and DAPI added to sort buffer at a final concentration of 1μg/ml to exclude dead or dying cells. Up to 100,000 live $CD3^+$ T cells (LIVE, $CD3^+$, $CD19^-$, $CD56^-$) and 100,000 leukaemia cells (LIVE, $CD3^-$, $CD19^+$, $CD10^+$) were sorted separately into cold PBS + 0.04% BSA, with cell counts determined by the sorter. Cells were counted manually, using trypan blue to assess viability and confirm cell counts, spun (350g, 5 min 4°C) and re-suspended in PBS +

0.04% BSA at a final concentration of 1000 cells/μl. T cells and leukaemia cells from each sample were combined at a 9:1 ratio for capture on the 10X Chromium Controller.

As described above for the ADAPTeR scRNA-Seq and scTCR-Seq data sets, the sorted cells were processed using the 10X Genomic Chromium Next GEM Single Cell 5' Reagents Kit V2 (dual index) for 5'gene expression library construction and V(D)J library construction. The samples were sequenced on the NextSeq using the High Output Kit v2.5 (150 Cycles). FASTQ files containing gene expression (GEX) and VDJ were de-multiplexed using *cellranger mkfastq* (10x Genomics). GEX reads were aligned to GRCh38 and counted using *cellranger count,* VDJ reads were aligned to cellranger's GRCh38 VDJ reference dataset using *cellranger vdj*.

Expression matrices were analysed using the Seurat package version 4 (Hao et al., 2020). For the VDJ data, the *filtered_contig_annotations.csv* files (generated with the *cellranger vdj* method) were loaded into R as data frames using the *read.csv* command. For simplicity and continuity with the B-ALL BIOMED-2 repertoire cohort, only beta chains were considered. For each sample, clonotypes were defined by unique beta CDR3 re-arrangement. Only productive CDR3 re-arrangements were considered (*cdr3 != "None"*) and cells expressing more than one productive beta re-arrangement were excluded from further analysis.

2.2 TCR Metrics

2.2.1 Clonality

For the lung TRACERx bulk TCR cohort, the ADAPTeR bulk TCR cohort and the B-ALL BIOMED-2 cohort, the clonality index defined as 1 – the normalised Shannon entropy was estimated for each sample by using the command *entropy* from the *entropy* R package, on the basis of the observed frequency of the TCRs in that sample.

$$Clonality = 1 + (\sum(p_i \cdot \ln(p_i))/\ln(N)$$

where p_i is the frequency of the ith TCR in the repertoire and N is the number of TCRs in that repertoire.

For the B-ALL BIOMED-2 cohort and the lung TRACERx bulk TCR cohort, to smooth out library size bias, the clonality score was computed as the arithmetic mean of 100 repetitions of the above formula computed for a random subset of 1000 total TCRs, per sample.

2.2.2 Classification of expanded/outlying TCR

For the ADAPTeR bulk TCR cohort and the lung TRACERx bulk TCR cohort, TCRs present in a sample above a threshold frequency of 2/1,000 were labelled as expanded. This threshold was optimised for the lung TRACERx bulk TCR cohort (Joshi et al., 2019).

Tissue specific expansion

For the lung TRACERx bulk TCR cohort, the relative abundance of the TCRs in the tumour (averaged over all regions) versus in paired non-tumour lung from the same patient was computed. The *P* value for the difference in abundance between tumour and non-tumour lung was calculated with the *poisson.test* function from the *stats* R package.

Longitudinal expansion/contraction

For the ADAPTeR bulk TCR cohort, the relative abundance of the TCRs in the pre-treatment sample versus the abundance in paired post-treatment sample from the same patient was computed for both the tumour compartment and the blood compartment, separately. As described above, the *P* value for the difference in abundance between pre- and post- was calculated with the *poisson.test* function in R.

Outlier detection

For the B-ALL BIOMED-2 cohort, the labelling of outlying re-arrangements was performed via a two-step process. Firstly, a derivative of the IQR outlier detection method was applied on the group (Ω) of re-arrangements with frequencies only observed once in the sample, corresponding to a y axis value equals to 1 on the log-log plots displayed in Figure 5.2. These frequencies were compared to the following threshold value:

$$threshold = Q_3 + 2.5\times(Q_3 - Q_1)$$

where Q_1 and Q_3 are the first and third quartiles of frequency values within Ω. If a given frequency was above this threshold, the corresponding re-arrangement was labelled as outlying.

Secondly, all log-log plots resulting from step one were manually screened to assess whether the threshold computed as above was splitting the data at a reasonable value. If the threshold value was not reasonable (for example classifying re-arrangements as aberrant when they did not appear as outliers by visual inspection when plotted), a more appropriate threshold was manually picked and re-applied to the sample.

2.2.3 Pairwise similarity

For the lung TRACERx bulk TCR cohort, the ADAPTeR bulk TCR cohort and the B-ALL BIOMED-2 cohort, the similarity between two TCR repertoires was assessed with the normalised dot product (also known as the cosine similarity) between the vectors of TCR abundance. The similarity between the two vectors is given as:

$$similarity = \frac{TCR_1 \cdot TCR_2}{\|TCR_1\|\times\|TCR_2\|}$$

where TCR_1 and TCR_2 are the abundance vectors, · represents the vector product and paired vertical bars represent the Euclidean norm of the vector.

2.2.4 Spatial diversity

For the lung TRACERx bulk TCR cohort, an additional spatial diversity score was computed per patient. The normalised Shannon diversity was estimated for each expanded TCR by using the command *entropy.empirical* from the *entropy* R package (Hausser & Strimmer, 2009) on the basis of the observed frequency of the TCR across all regions as:

$$H = -(\sum(p_i \cdot \ln(p_i))/\ln(N))$$

where H is the diversity, p_i is the observed frequency in the ith region and N is the number of regions. p_i is obtained by dividing the observed frequency in region *i* by the sum of the frequencies across regions such that $\sum(p_i) = 1$. H lies between 0 (TCR found in one region only) and 1 (TCR evenly found across all regions). To

derive a metric for each patient, the average of the diversity scores obtained for all expanded TCRs was computed.

A similar index was used to calculate the genomic mutational diversity. In this case, Shannon diversity was obtained for each nonsynonymous mutation with the same formula, where p_i was obtained by dividing the observed corrected prevalence of the mutation in the ith region by the sum of the corrected prevalence across regions such that $\sum(p_i) = 1$.

2.2.5 Ubiquitous and regional definition

In the lung TRACERx bulk TCR cohort, expanded intratumoural TCRs were subsequently classified as ubiquitous or regional. First, the probability that a TCR was absent from a region owing to sampling was determined. For each TCR, the likelihood of the data given two alternative models was compared. In model 1 (the null model), the TCR counts are drawn from a single Poisson distribution with the mean equal to the mean of all regions. In model 2, the TCR counts are drawn from a mixed distribution where one or more regions has no TCR with a probability of 1, and the remaining regions are drawn from a Poisson distribution. Then, the log-likelihood ratio between the two models was calculated. Finally, for each TCR, both models were run 1,000 times, drawing independent deviates from a Poisson distribution with the mean equal to the mean of all regions. The proportion (*p*) of simulations in which the log-likelihood ratio was greater or equal to the one observed with the real data was calculated. This procedure gives a non-parametric estimation of the *P* value correcting for the increased complexity of model 2. The algorithm was implemented in R and was run on all TCRs in each tumour. A TCR was deemed absent if the *P* value (corrected for multiple testing) was less than 0.05 (TCRs where the null model was significantly less likely to explain the data than the alternative model 2).

Expanded TCRs were then classified as regional if they were absent from at least one region of the tumour. Otherwise, they were classified as ubiquitous. Ubiquitous TCRs can therefore be absent from the data for specific regions but this is attributed to sampling rather than true spatial heterogeneity.

The same classification strategy was applied to the somatic mutation data obtained from matched WES sample. All mutations classified as non-synonymous variants, plus all mutations that introduced or removed a stop codon or introduced a frame shift were similarly classified as regional or ubiquitous. All mutations with frequencies of less than 10% were first labelled as absent and all mutations with frequencies greater than 10% as present. Then, each mutation was classified as ubiquitous if it was present in all the regions of a tumour and regional if it was absent from at least one region.

2.2.6 CDR3 amino acid clustering

For the lung TRACERx bulk TCR cohort, the ADAPTeR bulk TCR cohort and the B-ALL BIOMED-2 cohort, the pairwise similarity between pairs of TCRs was measured on the basis of amino acid triplet sharing. Sharing was quantified using the normalised string kernel function *stringdot* (with parameters *type = 'spectrum', length = 3, normalised = TRUE*) from the *Kernlab* package (Karatzoglou, Smola, Hornik, & Zeileis, 2004). The kernel is calculated as the number of amino acid triplets (sets of three consecutive amino acids) shared by two CDR3s, normalised by the number of triplets in each CDR3 being compared. The TCR similarity matrix was converted into a network diagram by using the *iGraph* package in R (Csardi & Nepusz, 2005). Two TCRs were considered connected if the similarity index was greater than 0.82.

For the ADAPTeR bulk TCR cohort, to normalise the counts of clusters obtained (N_{real}) for the input size for each sample, a random sample (of size equal to the number of expanded CDR3s N_{in}) of CDR3s was drawn and subjected to the same clustering strategy. This control step was repeated 10 times for each sample and the average number of clusters obtained for those controls (N_{ctrl}) was used to build the normalised cluster count value: N_{real}/N_{ctrl}.

For the lung TRACERx bulk TCR cohort, as no inter-patient comparison was computed, the normalised cluster count value was simply computed as N_{real}/N_{in}.

2.2.7 Cluster diversity

In the lung TRACERx bulk TCR cohort, for each CDR3 cluster constructed as described above, in order to capture the contribution of each possible combination of regions to the cluster, the Shannon diversity was computed. If *n* is the number of regions for a given patient, there are $N = 2^n - 1$ possible combinations of regions of any given size. The Shannon diversity is given by:

$$H = -(\sum(p_i \cdot \log(p_i))/\ln(N))$$

where H is the diversity, p_i represents the relative contribution of the ith combination to the cluster and N is the number of combinations. p_i is obtained by dividing the count of CDR3s belonging to the ith combination by the total number of CDR3s in the cluster such that $\sum(p_i) = 1$. H lies between 0 (a cluster composed of CDR3s belonging to one combination only) and 1 (a cluster evenly composed of all possible combinations). To derive a metric for each patient, the average of diversity scores obtained for all clusters was computed.

2.2.8 Frequency ratio

For the ADAPTeR bulk TCR cohort, for each expanded TCR at baseline that could also be detected after treatment, the ratio of the observed frequency at pre-treatment divided by the observed frequency post-treatment was computed. To derive a metric for each patient, the average of ratio scores obtained for all pre-treatment expanded TCRs was computed. Those that could not be detected after treatment were excluded from this analysis.

2.3 RNA-Seq metrics

2.3.1 Differential gene expression and gene set enrichment analyses

In the ADAPTeR bulk TCR cohort, *DESeq2* (Love et al., 2014) was used for differential expression analysis, using the binomial Wald test after estimation of size factors and estimation of dispersion. To identify genes differentially expressed between responders and non-responders, only transcripts with normalised count number >5 in at least 5 patients were considered. Pathway analysis was performed using the R package *XGR* (Fang, Knezevic, Burnham, & Knight, 2016) using the

gene ontology biological process (GOBP) databases (Ashburner et al., 2000; Gene Ontology, 2021). Induced and suppressed transcripts were analysed separately against the background of all tested transcripts. The “lea” ontology algorithm was used.

2.3.2 Danaher immune score

For the lung TRACERx bulk TCR cohort and the ADAPTeR bulk TCR cohort, the Danaher immune score was evaluated using RSEM abundance, z score scaled across all samples for which RNA-Seq was available. The Danaher immune score is a 60-marker gene signature derived from pan-cancer RNA-Seq analysis for 14 immune cell populations, where marker genes have been benchmarked against histological tumour-infiltrating lymphocyte (TIL) estimates and flow cytometry data (Danaher et al., 2017):

(1) Danaher Tcells: CD3D, CD3E, CD3G, CD6, SH2D1A, TRAT1
(2) Danaher CD8: CD8A, CD8B
(3) Danaher Cytotoxic: CTSW, GNLY, GZMA, GZMB, GZMH, KLRB1, KLRD1, KLRK1, PRF1, NKG7
(4) Danaher Bcells: BLK, CD19, MS4A1, TNFRSF17, FCRL2, KIAA0125, PNOC, SPIB, TCL1A
(5) Danaher NKcells: NCR1, XCL2, XCL1
(6) Danaher CD45: PTPRC
(7) Danaher DC: CCL13, CD209, HSD11B1
(8) Danaher CD8Ex: CD244, EOMES, LAG3, PTGER4
(9) Danaher Mac: CD163, CD68, CD84, MS4A4A
(10) Danaher Mast: MS4A2, TPSAB1, CPA3, HDC, TPSB2
(11) Danaher Neut: CSF3R, S100A12, CEACAM3, FCAR, FCGR3B, FPR1, SIGLEC5
(12) Danaher NKCD56: IL21R, KIR2DL3, KIR3DL1, KIR3DL2
(13) Danaher Th1: TBX21
(14) Danaher Treg: FOXP3

For each signature, the signature score was calculated as the mean of z-score scaled expression of all genes in that signature, and this for each sample.

2.4 scRNA metrics

2.4.1 QC

For the B-ALL 10X cohort, one Seurat object was created for each sample with the *Read10X* function performed on the *filtered_feature_bc_matrix* folder, obtained with the *cellranger count* method. A first round of basic Seurat QC was applied to each Seurat object with defaults parameters (*min.cells = 5, nFeature_RNA > 200, nFeature_RNA < 2500 and percent.mt < 5*). Then, the four samples were merged into one using the *merge* function and the Seurat functions *NormaliseData*, *FindVariableFeatures*, *ScaleData*, *RunPCA* were sequentially run on the merged object with default parameters. Clustering and UMAP was then performed with the *FindNeighbors*, *FindClusters* and *RunUMAP* functions with a clustering resolution of 0.5 and using the first 30 principal components (PCs). A UMAP projection of cells overlaid with expression of CD3E enabled to immediately separate T cells from leukaemic cells and to identify one cluster with mixed expression that was excluded from further analysis. Barcodes identifying T cells were stored as a separate Seurat object.

Following this first step, a second round of basic Seurat QC was applied to the initial Seurat objects created via the Read10X function, this time with less stringent filtering (*min.cells = 2, nFeature_RNA > 200 and percent.mt < 10*) as leukaemic cells were much less numerous and expected to be transcriptionally damaged. The procedure described above was then applied with the resolution parameter set to 0.8. Again, a UMAP projection of cells overlaid with expression of CD3E enabled to immediately identify the three clusters of leukaemic cells, which were extracted and stored as a separate Seurat object.

2.4.2 Integration

Following the construction of the T cell object and the leukaemic cell object, an integration procedure aimed at smoothing out any potential batch effect was applied to the former but not to the latter, as T cells were expected to be similar across

samples whereas leukaemic cells were expected to be extremely different between B-ALL patients (Caron et al., 2020).

The T cell Seurat object was split back into four separate objects (one per individual) and the Seurat functions *NormaliseData*, *FindVariableFeatures* were reapplied to each object with default parameters. The *FindIntegrationAnchors* and *IntegrateData* functions were then ran on the first 20 dimensions. Then, the functions *ScaleData*, *RunPCA*, *RunUMAP*, *FindNeighbors* and *FindClusters* were run on the integrated assay with the same parameters as above.

2.4.3 GSE132509 mapping

The *filtered_feature_bc_matrix* folders from Caron et al., 2020 were downloaded from https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE132509). Each sample was imported as a Seurat object into R with the *Read10X* function and merged as described above. The strategy described above to separate T cells from leukaemic cells was applied in the same way. The resulting object was split back into one T cell object per sample and only samples with at least 200 T cells were kept for further analysis and integrated back together with the method described in the previous section.

The resulting integrated object was then mapped onto the primary T cell object built as described in the previous section by sequentially applying the Seurat functions *SCTransform*, *FindTransferAnchors* and *MapQuery* on the first 20 PCs.

2.4.4 CellphoneDB

To analyse cell-cell interactions between leukaemic populations and T cell populations, the counts and cluster labels were extracted from the RNA assay of the pooled Seurat object and the *statistical analysis* method of the cellphoneDB algorithm (Efremova, Vento-Tormo, Teichmann, & Vento-Tormo, 2020) was ran on all populations together with default parameters. The resulting interactions were then extracted from the *significant_means.txt* output file and split per patient, only retaining significant interactions between a T cell population on one side and a leukaemic population on the other.

Chapter 3. Identification of distinct TCR populations that spatially mirror the genomic profile of NSCLC tumours

3.1 Introduction

The lung TRACERx study is a large prospective multi-institutional study (Jamal-Hanjani et al., 2014) that aims to describe the genomic evolution of NSCLC and the impact of tumour genomic and immune heterogeneity on disease progression through the analysis of multi-region tumour samples. Led by Professor Swanton's group, recruitment began in April 2014 and has recruited over 700 patients so far. The study provides a unique opportunity to explore the heterogeneity of the intratumoural immune landscape in NSCLC and how it relates to the genomic heterogeneity that has been described in that disease.

The analysis of whole-exome sequencing (WES) data of tumour specimens from the first 100 Lung TRACERx cases revealed high levels of intratumoural heterogeneity in the mutational and copy-number profile of these tumours, driven by genome doubling and chromosomal instability (Jamal-Hanjani et al., 2017). In a separate study, Rosenthal et al. demonstrated that the TME exerts a selective pressure on the tumour that triggers various immune escape mechanisms directed towards impairment of neoantigen presentation (Rosenthal et al., 2019). This work shows the importance of antigen presentation for an effective anti-tumour immune response which suggests that the corresponding recognition by T-cells through TCR signalling is equally important.

Taken together, these data suggested that the genomic heterogeneity that is described in NSCLC should be mirrored in the TCR repertoire of TILs and that the description of the immune landscape of NSCLC through the lens of TCR sequencing would provide novel insights on the anti-tumour immune response.

In this chapter, I carry out an in-depth analysis of the TCR repertoires of multi-region tumour samples, matched adjacent non-tumour tissue and PBMC samples form 72

TRACERx patients with early stage NSCLC. Firstly, I reveal a pattern of heterogeneity in the TCR repertoire resembling the heterogeneity observed in the matched mutational profile of these patients and a range of clonotype frequency values associated with tumour specificity. I also describe how a novel method was built in order to identify distinct TCR populations that spatially follow the pattern of the well-established clonal/subclonal mutational dichotomy. Secondly, I provide indirect characterisation of these TCR populations through integration of matched transcriptomic data and in-depth analysis of networks of similar CDR3 amino acid chains associated with antigenic specificity. Lastly, I demonstrate that these TCRs can be detected in the blood, which is critical for potential clinical applications of this work.

The TRACERx project is a collaborative effort involving multiple research groups. Details of individual contributions to the work presented in this chapter are as follows. Sample acquisition, preparation and sequencing were jointly done by Kroopa Joshi (oncologist at the Royal Marsden, formerly PhD student in the Quezada group), Imran Uddin (single cell genomics technician, single cell genomics facility, UCL, formerly lab technician in the Chain group) and Annemarie Woolston (post-doctoral fellow in the Chain group). Processing of raw FASTQ sequence data from TCR repertoire sequencing and single cell RNA-Seq was performed by Mazlina Ismail (post-doctoral fellow in the Attard group, formerly PhD student in the Chain group). Processing of raw FASTQ sequence data from RNA sequencing and prediction of neoantigen load was carried out by Rachel Rosenthal (Senior Bioinformatician at Achilles Therapeutics UK Limited, formerly PhD student in the Swanton group). The algorithm for CDR3 clustering and the custom script developed to detect pre-defined TCRs in RNA-Seq data were conceived by Benny Chain. Neoantigen reactive T cell experiments and FACS experiments were carried out by James Reading (group leader at UCL Cancer Institute, former post-doctoral fellow the Quezada group). I performed all downstream bioinformatics analysis and data visualisation. The data presented in this chapter of the thesis largely forms sections of a published manuscript (Joshi et al., 2019).

3.2 Results

3.2.1 Identification of TCRs enriched in tumour compared to non-tumour

Based on availability of RNA from multiple tumour regions, 60 patients within the TRACERx 100 cohort and 12 additional TRACERx post-100 patients were selected (Figure 3.1). Taken together, the TCR repertoires (both α-chain and β-chain) from 220 tumour regions, 64 matched non-tumour lung tissue samples and 56 PBMC samples taken at the time of primary surgery were sequenced. In addition, the TCR repertoires of 14 follow-up PBMC samples were sequenced (median time to follow-up was just under two years). The selection of patients covered a large range of mutational and clinical profile (Figure 3.2), offering the opportunity to explore how characteristics of the TCR repertoire fit with different clinical features often used to describe NSCLC tumours.

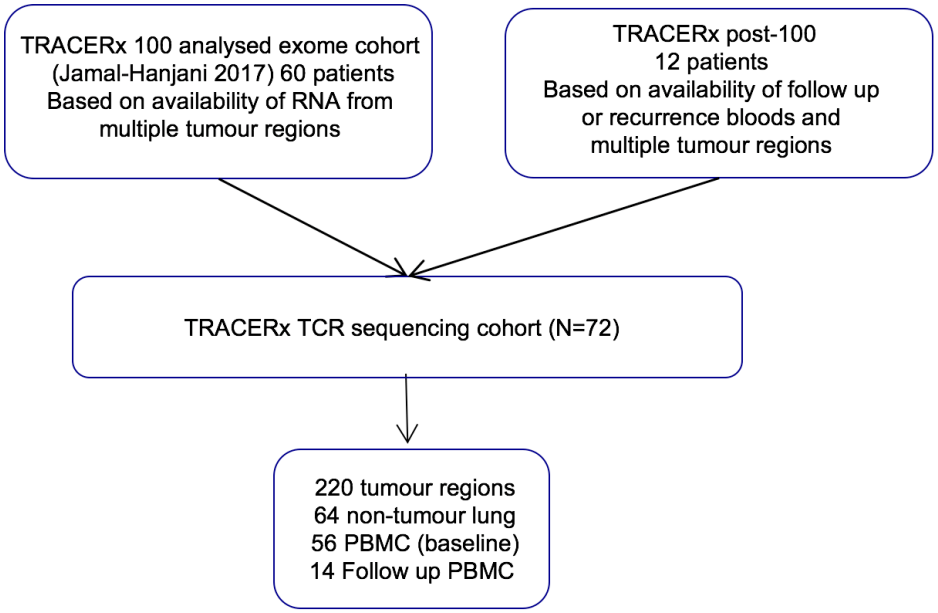

Figure 3.1 Lung TRACERx patient selection

CONSORT diagram showing the selection of lung TRACERx patients for TCR sequencing.

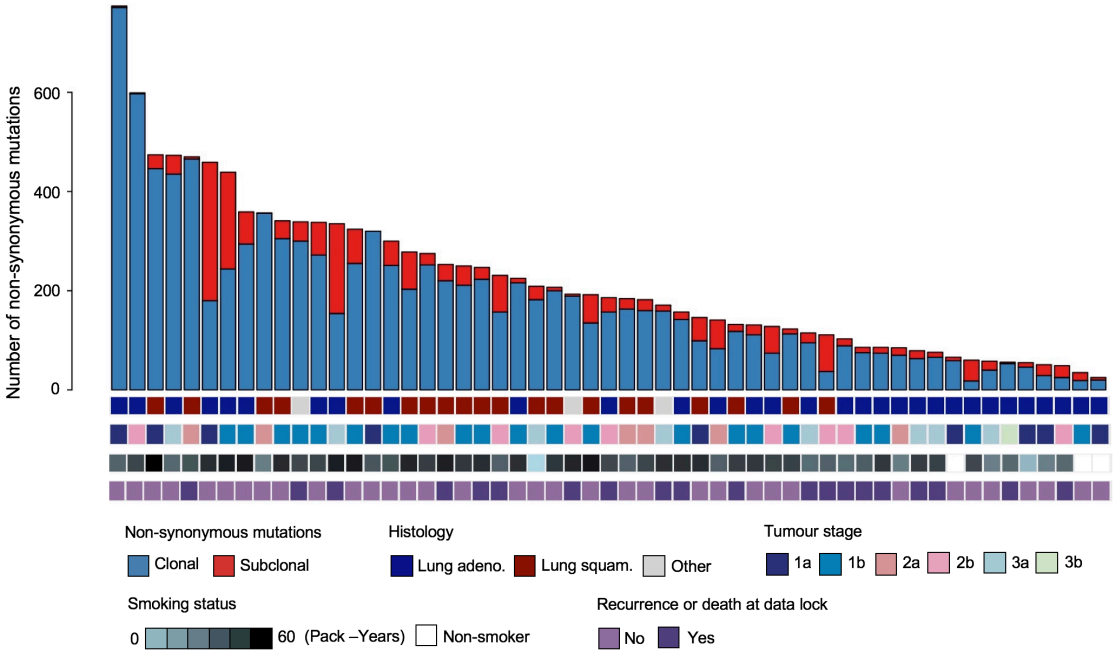

Figure 3.2 Description of Lung TRACERx patient cohort

The total number of non-synonymous mutations (clonal and subclonal) and patient clinical characteristics (histology, stage, smoking status and clinical outcome).

The median number of total and unique TCRs retrieved was higher for β-chain than for α-chain in tumour samples (Figure 3.3), probably reflecting the fact that there are more RNA molecules per cell for β sequences than for α sequences (Oakes et al., 2017). However, the total numbers of α- and β- chains were highly correlated, indicating that any metric derived in one compartment should be mirrored in the other. For that reason and for clarity purposes, if not otherwise specified, all data presented in this chapter and throughout the thesis was based on the β repertoire. In particular all analyses performed in this chapter were repeated successfully in the α repertoire and can be found in Joshi et al., 2019.

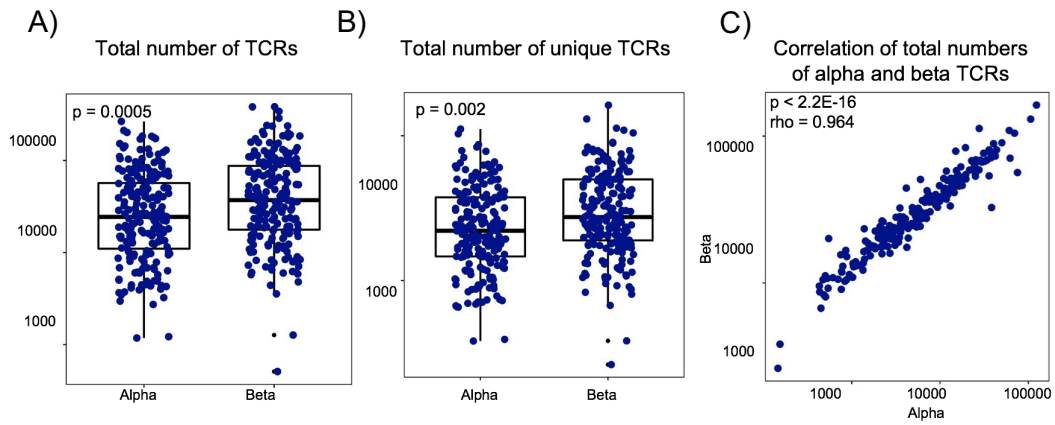

Figure 3.3 Total and unique number of intratumoural α-chain and β-chain TCRs sequenced, per tumour region

The total number of TCR α-chain and β-chain TCRs sequenced (A), the number of unique TCR sequences detected (B) and the correlation between the total number of TCR α-chain and β-chain segments sequenced (C) in multi-region tumours (n = 220). For the box plots displayed in (A) and (B) and all box plots in this thesis, the minimum and maximum values are indicated by the extreme points of the box plot; the median is indicated by the thick horizontal line; and the first and third quartiles are indicated by box edge.

Power laws distribution and their associated log-log representation have been previously described as a valuable way to model and visualise TCR distributions for circulating effector memory T cells in blood (Oakes et al., 2017). In order to capture potential immediate differences between tumour and non-tumour repertoire, log-log representations of both compartments were plotted for one patient (Figure 3.4). Surprisingly, no difference was observed between the two tissue types (the distribution parameters fitted to tumour and to non-tumour samples did not differ significantly at the cohort level: n = 64 patients, P = 0.16, two-sided Mann–Whitney test, see Joshi et al., 2019).

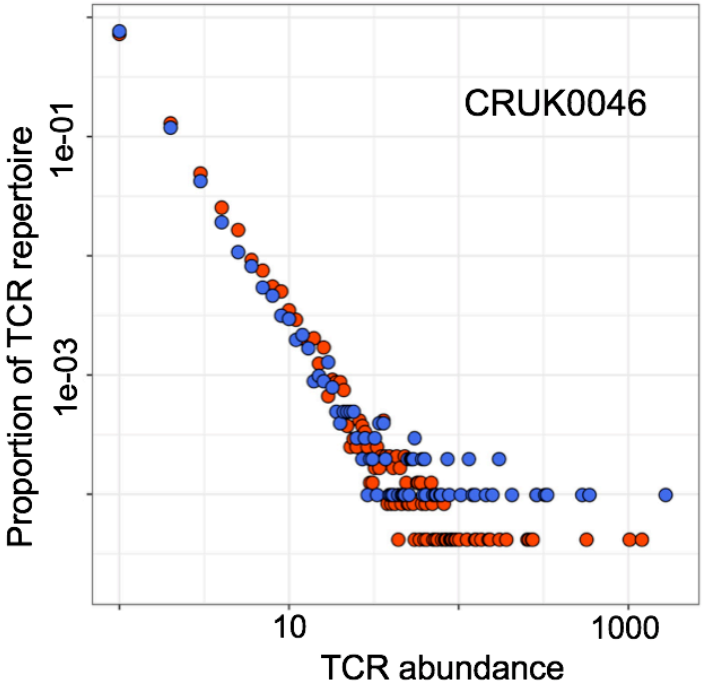

Figure 3.4 Log-log representation of intratumoural TCR frequency distribution

Representative log-log plot (from patient CRUK0046) tumour (red circles) and matched non-tumour repertoire (NTL; blue circles). The *x* axis represents TCR abundance (size of clone), and the *y* axis represents the proportion of the repertoire.

Similarly, when computing the clonality score (see Materials and Methods) to measure the diversity of the repertoire, no significant difference was observed between the two compartments (Figure 3.5).

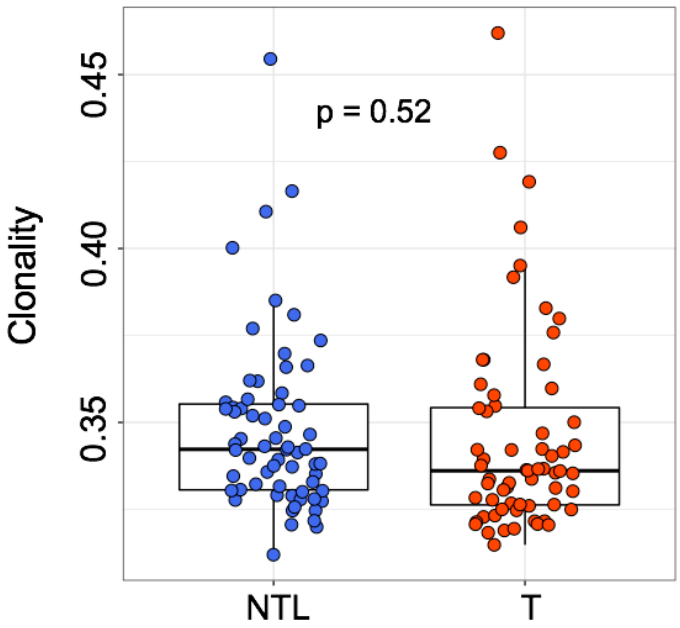

Figure 3.5 Tumour and non-tumour TCR repertoire clonality scores, per patient

The normalised clonality score (see Materials and Methods) of tumour regions (n = 72 patients, with multiple tumour regions pooled from an individual patient; red circles) and non-tumour lung samples (n = 64 patients; blue circles). The two-sided Mann–Whitney test *P* value is shown.

As no difference could be observed at the level of the entire repertoire, different frequency thresholds were applied in order to focus the analysis on the most expanded TCRs, which should intuitively identify more relevant specificities (Figure 3.6). Such TCRs could be identified at each threshold values, both in tumour and non-tumour lung.

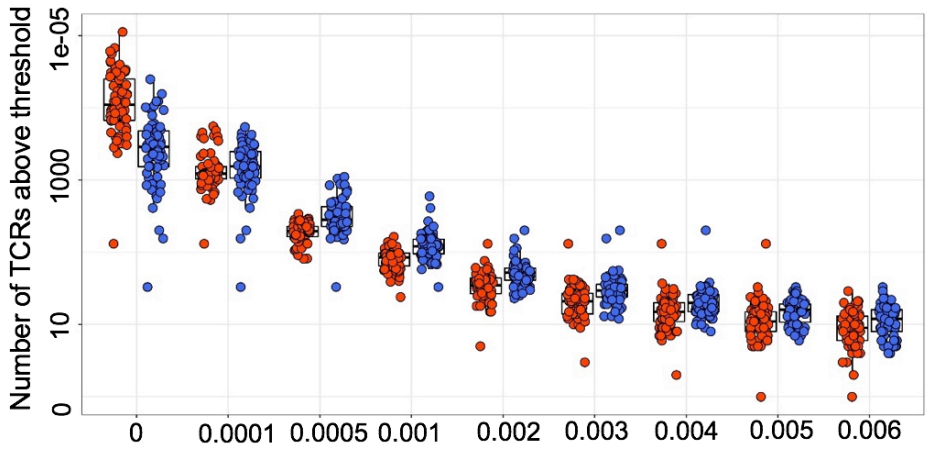

Figure 3.6 The intratumoural and non-tumour number of expanded TCRs per frequency threshold, per patient

The number of TCRs detected above a given frequency threshold is shown for tumour ($n = 72$ patients; red circles) and matched non-tumour lung ($n = 64$ patients; blue circles).

In order to assess whether such filtering was indeed retaining the most interesting part of the intratumoural repertoire, the composition of these groups of TCRs was further explored by labelling the TCRs observed at a frequency above 0.002 as "expanded". This cut-off value was selected arbitrarily on the basis of being high enough to filter out "bystander" clonotypes and low enough to select a sufficient number of TCRs for downstream analysis (a number of metrics are less robust if the input size is too small).

The TCRs expanded in the tumour and the ones expanded in the adjacent non-tumour tissue were both determined and their respective frequency in each compartment was compared. A small number of expanded TCRs had a similar frequency in both compartments i.e. corresponding to TCRs expanded both in the tumour sample and the matched non-tumour sample. However, the majority of expanded TCRs had a remarkably larger frequency in the tumour repertoire or in the non-tumour. In order to investigate these distributions more closely, the probability of each expanded TCR being tumour or non-tumour specific was computed (see Materials and Methods, Figure 3.7).

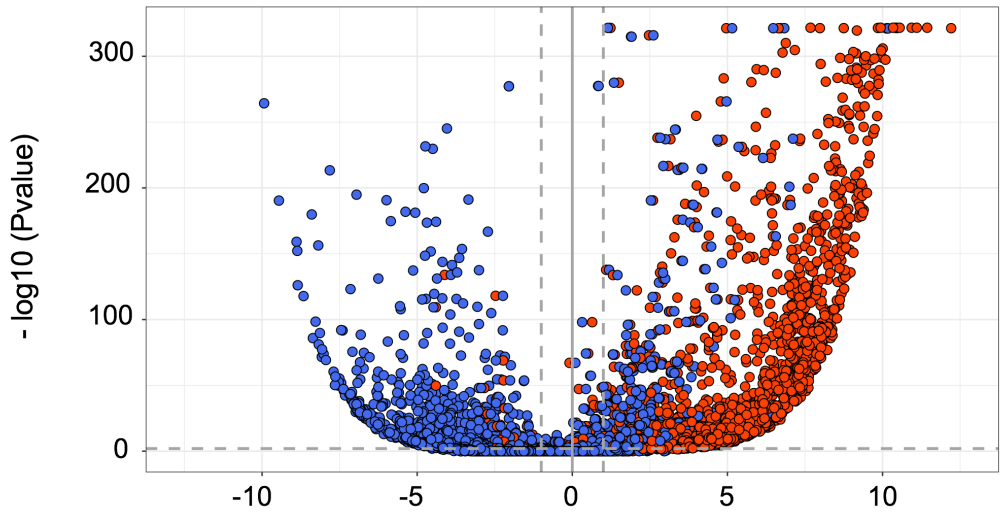

Figure 3.7 Differential expansion level of TCRs in tumour and non-tumour

A volcano plot showing the likelihood ($-\log_{10}$(*P* value)) of each expanded TCR being tumour or non-tumour specific, plotted against differential expression in tumour versus non-tumour lung. Blue circles represent TCRs expanded (>0.002) in non-tumour lung and red circles represent TCRs expanded in tumour. The horizontal dashed line corresponds to $P = 0.01$; the vertical dashed lines indicate a two-fold differential in expression between the tissues.

The vast majority of the intratumoural expanded TCRs was not found in the non-tumour tissue (mean proportion = 95%) whereas a much lower proportion of the TCRs expanded in the non-tumour lung was not found in the tumour tissue (mean proportion = 36%) (Figure 3.8). This is likely to be the effect of cancer specific antigenic stimulation occurring in the tumour versus expansion driven by common non-tumour lung antigens (e.g. respiratory viruses) occurring both in tumour and in the adjacent non-tumour tissue.

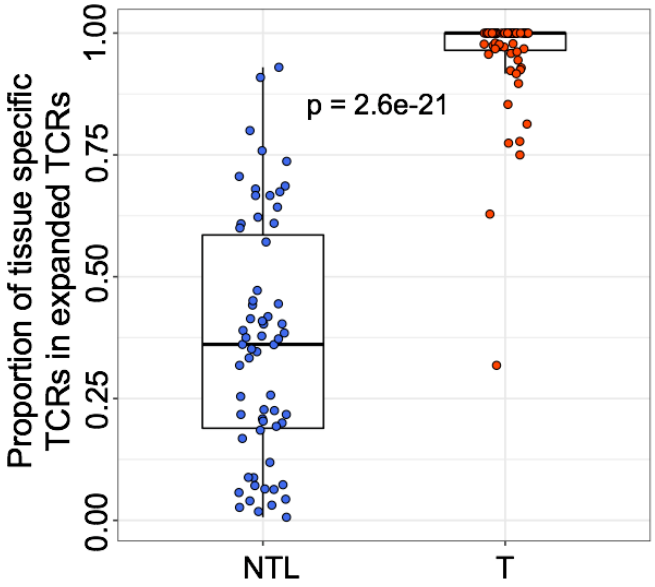

Figure 3.8 Proportion of tissue specific TCRs in expanded tumour and non-tumour TCRs

The proportion of expanded tumour or non-tumour lung TCRs that are specific to their respective tissue, defined on the volcano plot as TCRs that had adjusted $P < 0.05$ and a differential abundance of at least two between the tissues. The two-sided Mann–Whitney test P value is shown; $n = 72$ patients (tumour) and $n = 64$ patients (non-tumour lung).

These data show that applying such a frequency threshold on the intratumoural TCR repertoire was indeed enriching for a population of TCRs presenting features suggestive of tumour-specificity.

3.2.2 The number of TCRs expanded in the tumour correlates with the number of non-synonymous mutations

To explore this further, the matched genomic data was integrated with the TCR-Seq data by plotting the number of intratumoural expanded TCRs against the number of non-synonymous mutations (NSMs) when whole-exome sequencing was available (Figure 3.9). The number of expanded TCRs was significantly positively correlated with the number of non-synonymous mutations, which suggested that the distribution of expanded TCRs in the cohort was following the level of tumours' neo-antigenicity. This supported the hypothesis that specific intratumoural expansion of TCRs was driven, at least partially, by neoantigen derived stimulation.

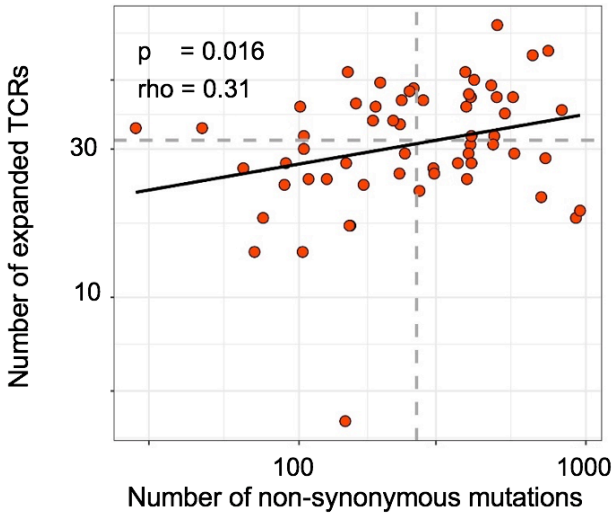

Figure 3.9 Correlation between intratumoural expanded TCRs and non-synonymous mutations

The correlation between the number of non-synonymous mutations and the number of unique intratumoural expanded TCRs (frequency ≥ 2/1,000) for each patient. The Spearman's rank correlation coefficient and *P* value are shown ($n = 59$ patients).

Interestingly, this association was robust over a range of cut-off values (Figure 3.10). Moreover, the same correlation was run for the number of TCRs expanded in the non-tumour lung as a control and no association was observed, agnostic of the cut-off retained. This was consistent with the hypothesis that the expansion observed in the tumour is driven by a cancer specific antigenic signal that is absent in the matched non-tumour tissue.

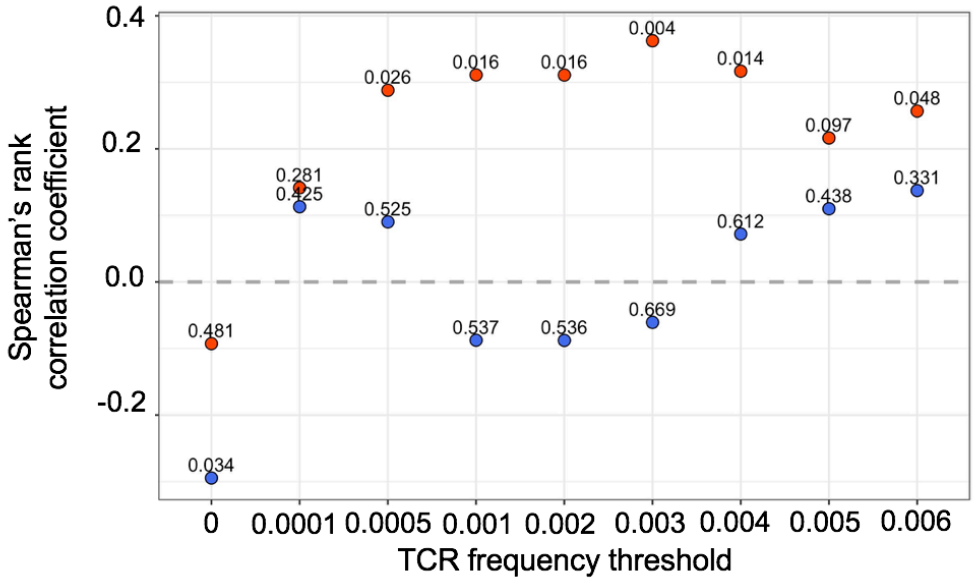

Figure 3.10 Summary of correlations between intratumoural expanded TCRs and non-synonymous mutations, per frequency threshold

The Spearman's rank correlation coefficient and *P* value (shown above each point) for the relationship between the number of non-synonymous mutations and the number of unique intratumoural ($n = 72$ patients; red circles) or non-tumour lung ($n = 64$ patients; blue circles). expanded sequences at different frequencies (ranging from all TCRs (threshold of zero) up to those found at a frequency of ≥8/1,000).

3.2.3 The distribution of expanded ubiquitous and expanded regional TCRs mirrors the genomic landscape of NSCLC

To characterise this observation further, the spatial heterogeneity of the association between expanded TCRs and mutational burden was evaluated. The frequency of expanded TCRs in each individual tumour region was computed for each patient and showed a large variety of distribution profiles (Figure 3.11).

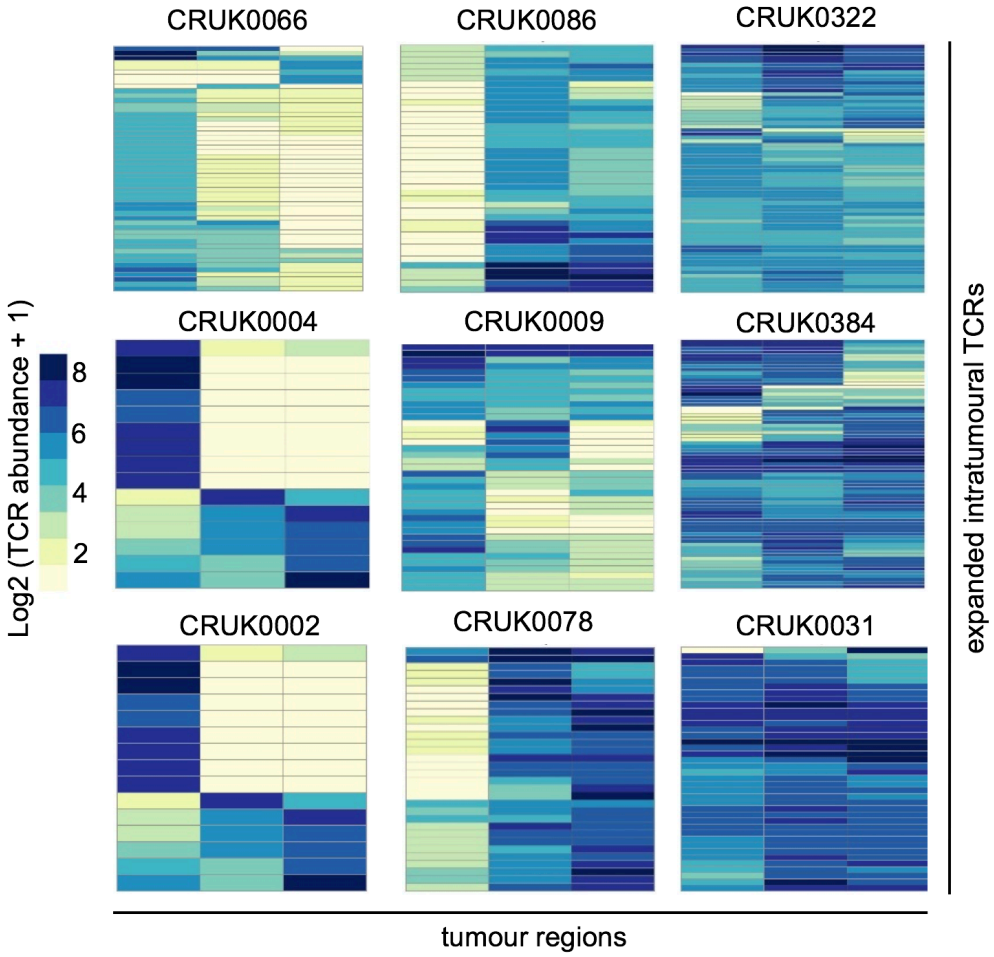

Figure 3.11 Representative intra-patient distributions of expanded TCRs across tumour regions

Heat maps showing the abundance (log_2 of 1 + the number of times each TCR was found) of expanded intratumoural TCRs (frequency $\geq$ 2/1,000) in different tumour regions for several patients. Patient ID is shown above each heat map. Each row represents one unique sequence, and each column represents one tumour region.

For some patients, expanded TCRs were measured at high expansion levels in each one of the available tumour regions whereas some patients displayed a much more heterogeneous pattern where most of the expanded TCRs were measured in some of the tumour regions but absent in others (Figure 3.11 and Figure 3.12).

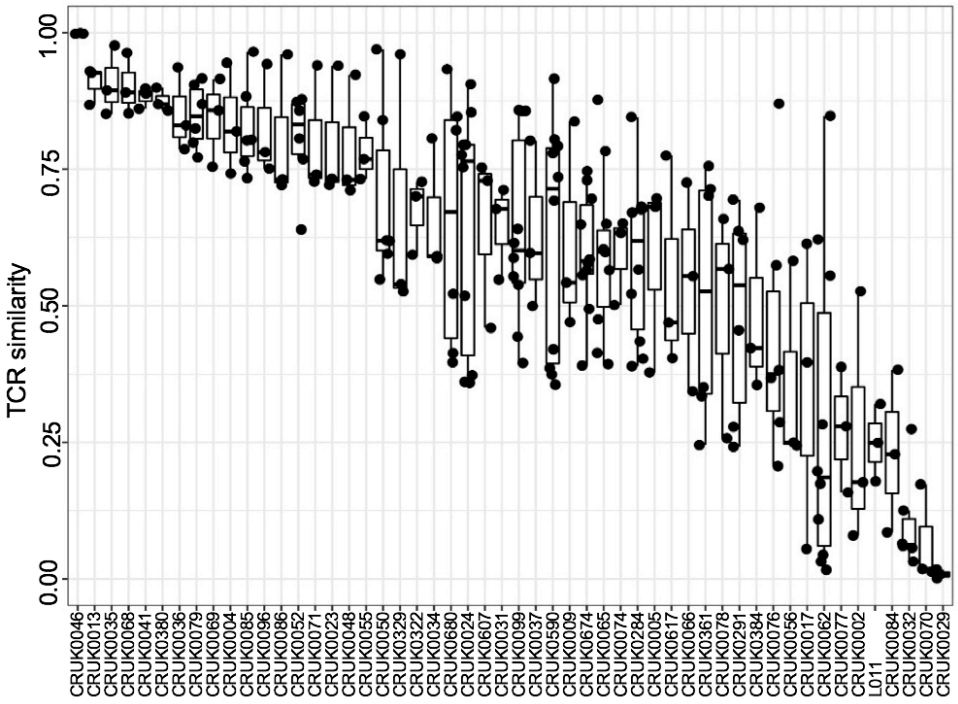

Figure 3.12 Intra-patient repertoire similarity

The pairwise intratumoural TCR repertoire similarity is shown for each patient (see Materials and Methods). Each circle represents a comparison between two regions from the same patient's tumour. Patients are ordered along the *x* axis by descending mean intratumoural TCR similarity ($n = 52$ patients).

Given the association observed between mutations and expanded TCRs and the previously well described genomic heterogeneity of NSCLC, the potential link between the TCR repertoire spatial heterogeneity and the mutational spatial heterogeneity was investigated.

In order to capture this hypothetical association, a diversity score was computed for each patient, both for the genomic compartment and the TCR repertoire compartment which showed that TCR diversity was correlated with genomic diversity (Figure 3.13).

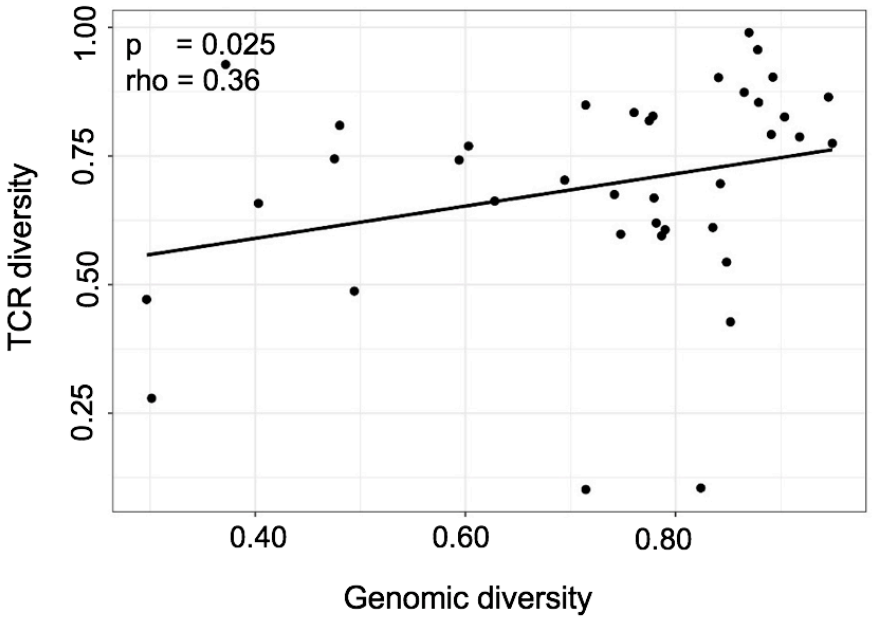

Figure 3.13 Correlation between TCR repertoire diversity and genomic diversity

TCR repertoire diversity plotted against genomic diversity for each patient. The Spearman's rank correlation coefficient and *P* value are shown; $n = 41$ patients.

To explore this observation further, for patients with at least three tumour regions, expanded TCRs were sub-classified (see Materials and Methods) as either "ubiquitous" (found in every tumour region) or "regional" (absent in at least one region). Mutations were sub-classified in the same way (which slightly differs from the widely used clonal/subclonal distinction but was better suited for this work, see Materials and Methods). Like ubiquitous mutations, expanded ubiquitous TCRs were more abundant than regional TCRs (resp. expanded regional mutations, Figure 3.14), which suggested that those numbers may be correlated to one another.

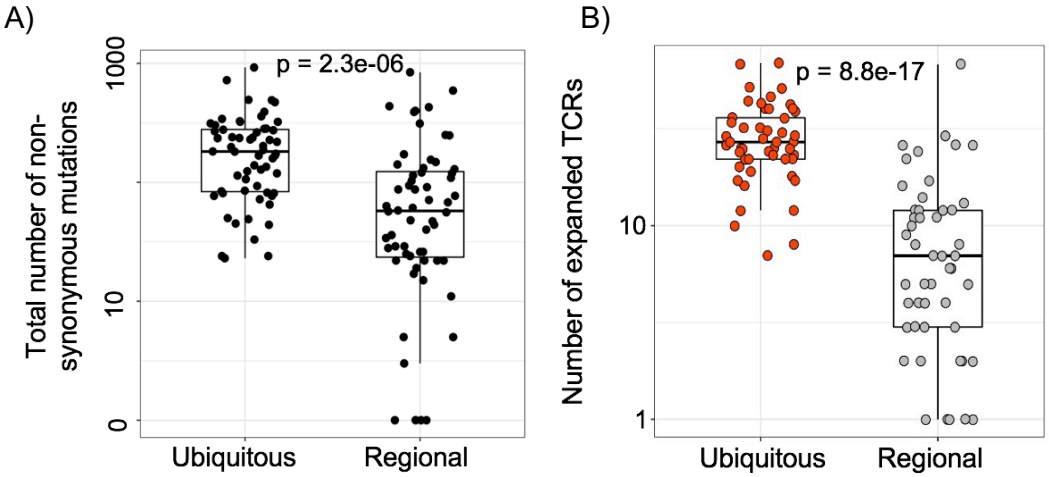

Figure 3.14 The number of non-synonymous mutations and the number of expanded TCRs, split between ubiquitous and regional

(A) The numbers of ubiquitous and regional non-synonymous mutations for each tumour region. The two-sided Mann–Whitney test *P* value is shown; $n = 60$ patients. (B) The numbers of expanded (frequency $\geq$ 2/1,000) ubiquitous (red circles) and regional (grey circles) sequences for each tumour region. The two-sided Mann–Whitney test *P* value is shown; $n = 52$ patients

The number of expanded ubiquitous and regional TCRs were both plotted against the number of regional and ubiquitous non-synonymous mutations (Figure 3.15). In multivariate regression, the number of expanded ubiquitous (resp. regional) TCRs was positively associated with the number of ubiquitous (resp. regional) non-synonymous mutations but not associated in any way with the number of regional (resp. ubiquitous) mutations.

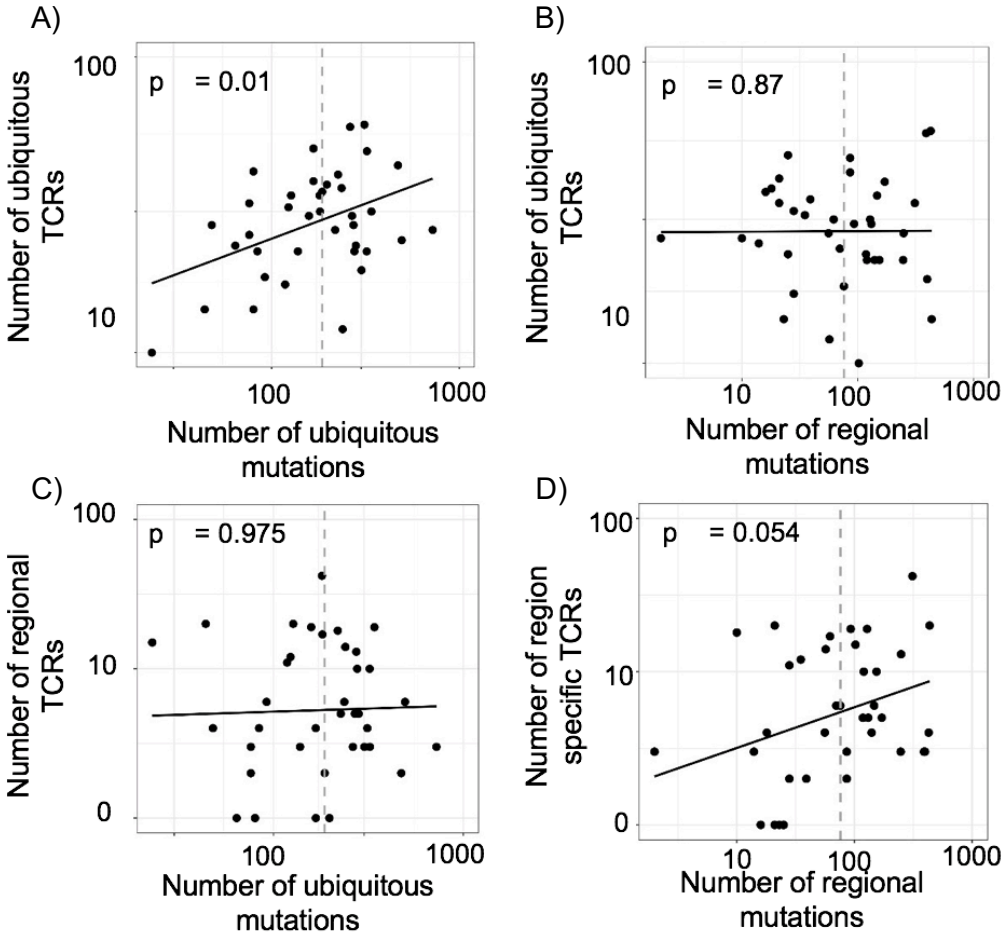

Figure 3.15 Correlation between the number of non-synonymous mutations and the number of expanded TCRs, split between ubiquitous and regional

The number of expanded ubiquitous (top) or regional (bottom) sequences plotted against the number of ubiquitous (left) or regional (right) nonsynonymous mutations for each tumour region. *P* values derived from multivariate regression are shown; dashed lines correspond to median values. $n = 42$ patients.

These data supported the hypothesis that tumour specific ubiquitous TCRs are likely to recognise ubiquitous mutations whereas regional mutations would only produce a spatially constrained antigenic signal, resulting in a spatially constrained expansion of TCRs and in turn leading to the presence of regional TCRs.

3.2.4 Clustering of amino-acid motifs identifies groups of TCRs similar to expanded TCRs

Further evidence to support the hypothesis that the expanded TCRs observed in the tumour were part of an antigen-specific response was sought in the amino acid content of expanded ubiquitous and regional TCRs. Patterns of amino acid motifs in sets of CDR3 β chains have previously been described as reflecting a common antigen specificity of TCRs (Dash et al., 2017; Glanville et al., 2017; Sun et al., 2017; N. Thomas et al., 2014). A custom clustering algorithm was built (Figure 3.16**,** see Materials and Methods) and firstly utilised to test whether expanded TCRs, hypothesised as enriched for tumour specificity, were indeed found inside clusters of similar CDR3 sequences. The algorithm was applied to each patient on the entire intratumoural repertoire and the structures identified were screened for the presence of expanded TCRs.

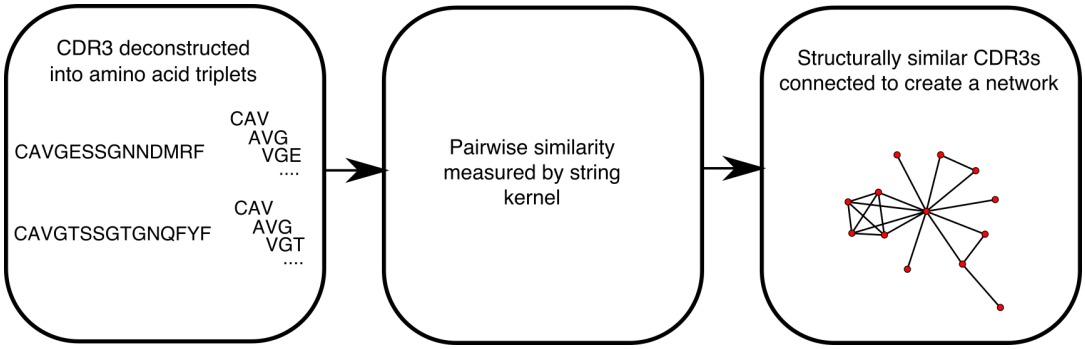

Figure 3.16 Representation of the CDR3 clustering algorithm

Diagram illustrating the CDR3 similarity network construction process. Individual CDR3s are deconstructed into overlapping series of contiguous amino acid triplets, and the pairwise similarity between two CDR3s is calculated as the normalised string (triplet) kernel. CDR3s that have a pairwise similarity of >0.82 are connected by an edge.

The method systematically identified more clusters of CDR3 chains around expanded TCRs than around random TCRs (Figure 3.17), likely due to the fact that the antigenic signal giving rise to the expanded TCRs also generated TCRs with similar CDR3 chains.

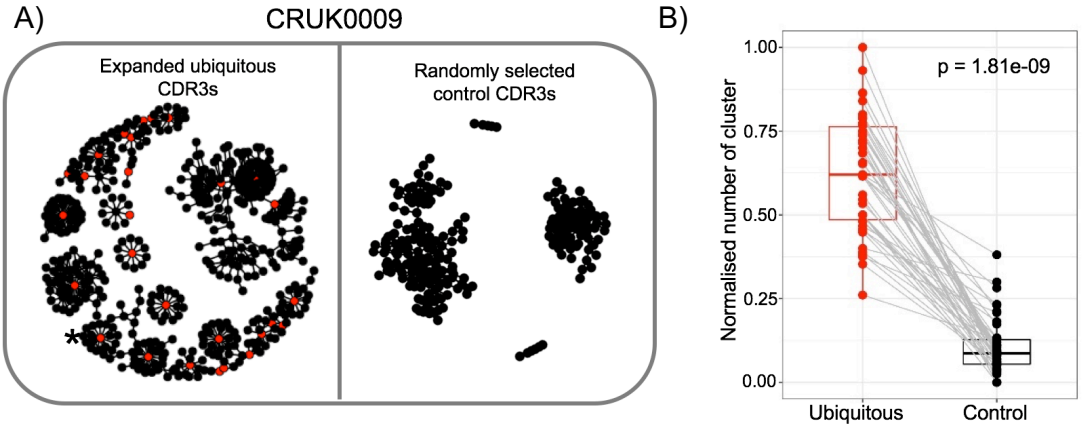

Figure 3.17 Summary of counts of networks around ubiquitous and control CDR3s

(A) Representative network diagrams of intratumoural CDR3 sequences for patient CRUK0009. Both panels show the network of CDR3 sequences that are connected to at least one other CDR3 within the tumour. Left, clustering around expanded intratumoural ubiquitous CDR3s (red circles). Right, clustering around a random sample of CDR3s from the same repertoire (same numbers as for the expanded ubiquitous CDR3s). (B) The clustering algorithm was run on all patients, and the number of clusters for the networks containing expanded ubiquitous and control randomly selected CDR3 sequences is shown. The one-sided Mann–Whitney test *P* value is shown; $n = 46$ patients.

3.2.5 The distribution of TCRs clustered around expanded ubiquitous or expanded regional TCRs follows identical spatial constraints

The composition of TCR clusters was additionally investigated in relation to the ubiquitous/regional labelling of expanded TCRs and their suspected link with ubiquitous/regional antigens. Each cluster was associated with a diversity score, measuring the spatial distribution of TCRs: a high diversity score corresponds to a cluster with an even representation of tumour regions and a low diversity score corresponds to a cluster skewed towards a specific subset of tumour regions. Then, two values were computed for each patient: the average of diversity scores of clusters containing a ubiquitous TCR and the average across clusters containing a regional TCR (Figure 3.18, see Materials and Methods). Clusters containing expanded regional TCRs presented significantly lower diversity scores, indicating that the CDR3 chains that are similar to expanded regional ones, likely recognising

the same antigen, are often found in the same region, whilst CDR3s that are similar to expanded ubiquitous ones are present across the entire tumour.

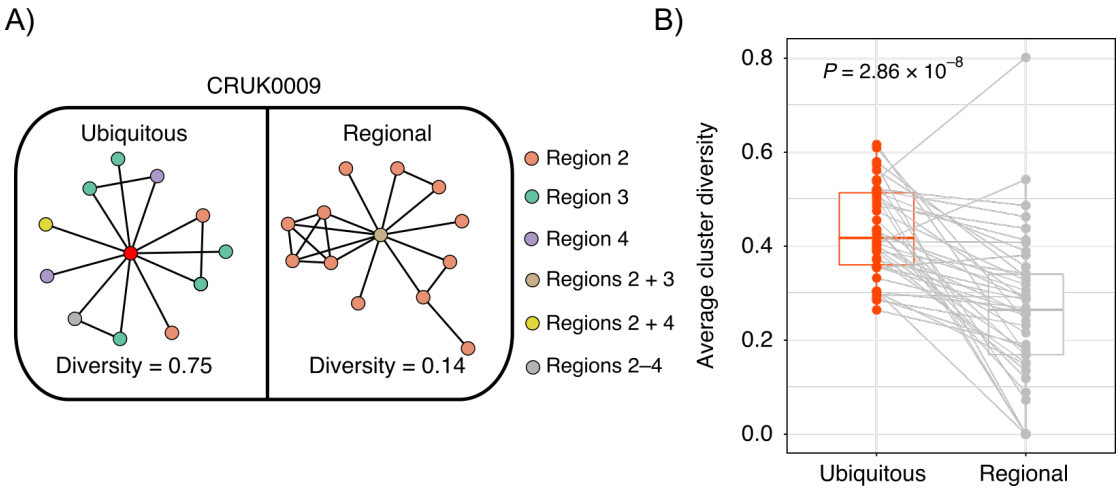

Figure 3.18 Summary of spatial composition of networks around ubiquitous and regional CDR3s

(A) Representative clustering around a ubiquitous or regional expanded TCR from CRUK0009, with nodes coloured according to the regions in which each TCR was found. **b** The average cluster diversity for clusters containing ubiquitous or regional expanded TCRs. The one-sided Mann–Whitney test *P* value is shown; $n = 42$ patients.

Together with the previous observations, the fact that spatially coherent clusters of related CDR3 sequences were identified was consistent with the hypothesis that these subsets are enriched for antigen-specific TCRs with similar spatial constraints.

3.2.6 RNA-Seq of FACS sorted subsets of $CD8^+$ T cells identifies expanded ubiquitous TCRs within a phenotype consistent with tissue resident tumour-antigen-reactive T cells

Three indirect transcriptomic approaches were leveraged to further characterise the antigenic specificity of expanded TCRs. Firstly, TILs from three patients were sorted, using flow cytometry, into $PD\text{-}1^+CD45RA^-CCR7^-CD57^-$ cells, a phenotype associated with cytotoxicity in cancer (Bengsch et al., 2018; Bengsch et al., 2010; A. C. Huang et al., 2017; Simoni et al., 2018) (labelled $PD\text{-}1^+$ cells) and all other $CD8^+$ T cells (labelled PD-1- cells) and subjected to RNA sequencing (Figure 3.19**,** see Materials and Methods). The resulting sequencing data was then screened for the presence

of expanded regional or ubiquitous TCRs as identified by matched TCR-Seq for these patients. Expanded ubiquitous and regional TCRs could be found both in the PD1$^+$ and PD1$^-$ sorted subsets. However, expanded ubiquitous TCRs could be found significantly more often in the PD1$^+$ subset than in the PD1$^-$ subset and the expanded regional TCRs seemed to follow the same pattern although it was difficult to conclude due to the low number of regional TCRs retrieved in these data.

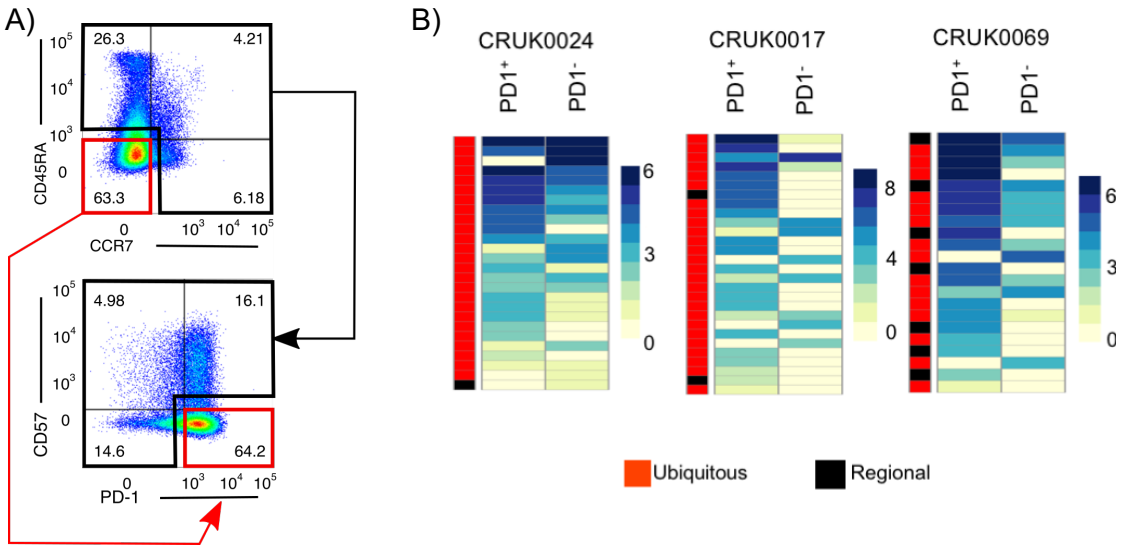

Figure 3.19 Distribution of expanded TCRs across FACS sorted PD-1$^+$ and PD1$^-$ CD8$^+$ T cells subgroups

(A) CD8$^+$ TILs from CRUK0024, CRUK0017 and CRUK0069 were sorted into two populations: CD45RA$^-$CCR7$^-$PD-1$^+$CD57$^-$ and all other CD8$^+$ TILs (referred to as PD-1$^+$ and PD- 1$^-$ subpopulations, respectively). The flow cytometry gating strategy for a representative patient is shown (pre-gated on live > singlets > CD3$^+$ > CD8$^+$ T cells). (B) RNA from sorted populations was extracted and sequenced, and the RNA-Seq data were mined for the presence of expanded ubiquitous and regional sequences. The heat maps show the number of times each expanded ubiquitous or regional TCR CDR3 sequence was found in the RNA-Seq data from PD-1$^+$ or PD-1$^-$ cells, (see Materials and Methods). The colour key gives the proportions scaled for each row, where each row represents a distinct expanded TCR sequence.

As mentioned above, given that lack of CD57 expression and high expression of PD-1 has been associated with tumour specific CD8 T cells, these data were coherent the analysis performed up to Section 3.2.5, suggesting enrichment for tumour specificity in these TCR subsets.

3.2.7 Single-Cell RNA-Seq of Neoantigen reactive CD8$^+$ T cells captures ubiquitous expanded TCRs

Secondly, CD8$^+$ TILs were sorted by fluorescent MHC multimers bound to a peptide encoded by a ubiquitous mutation in the *MTFR2* gene in order to isolate antigen specific T cells in patient L011. Single-cell RNA sequencing was then performed on the sorted cells (Figure 3.20, see Materials and Methods).

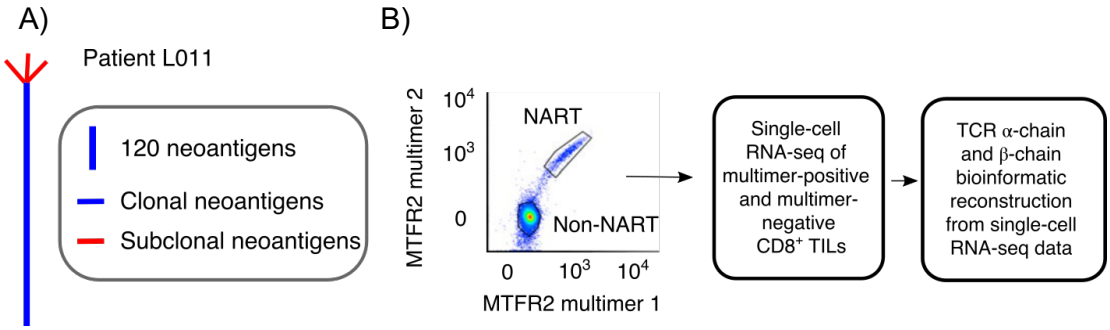

Figure 3.20 Isolation of neoantigen specific CD8$^+$ T cells in patient L011

(A) CD8$^+$ T cells reactive to a clonal neoantigen were isolated from a patient with a high clonal nonsynonymous mutational load (patient L011). Phylogenetic tree of tumour adapted from McGranahan et al., 2016. (B) Representative dot plot of the TIL peptide:multimer sorting strategy (left) and the workflow for single-cell RNA-Seq (right). CD8$^+$ TILs were screened for neoantigen- reactive T cells (NARTs) reactive to a peptide arising from the mutated *MTFR2* gene and sorted for single-cell RNA-Seq (pre-gated on lymphocyte > single cell > viable > CD3$^+$CD8$^+$ T cells)

Two families of TCRs were identified in the single cell data based on the sharing of a same α-chain ("CALRLGSGGSEKLV") or β-chain ("CASSPRTGGYEQY") and were subsequently compared to the multi-region TCR sequencing available for patient L011. Strikingly, the TCRs retrieved from the multimer positive sort were expanded and ubiquitously expressed in the tumour and were absent from the matched adjacent non-tumour tissue (Figure 3.21). Conversely, the TCRs retrieved from the multimer negative sort, displayed a more diverse distribution pattern across tumour regions. Some of them were also ubiquitously expanded in the TCR-Seq data, probably because they were specific to another ubiquitous antigen that not captured by this experiment.

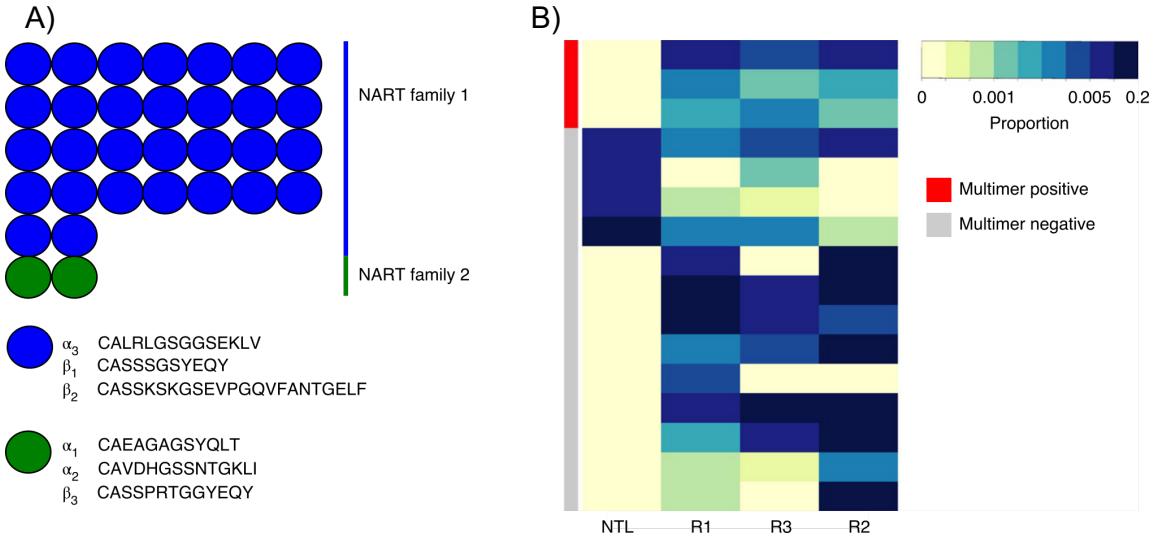

Figure 3.21 Frequency distribution of neoantigen specific TCRs across tumour and non-tumour samples for L011

(A) TCR α-chain and β-chain sequences were bioinformatically reconstructed from the single cells (see Materials and Methods). The neoantigen-reactive T cells comprised two families of cells with distinct CDR3 sequences and abundances. (B) Heat map showing the abundance of antigen-specific β-chain sequences (β_1, β_2 and β_3 from panel A) in tumour regions (R1–R3) and non-tumour lung (NTL) from multimer-positive (red) and multimer-negative (grey) cells within the bulk TCR sequencing data. Each row represents a unique TCR found in the multimer-positive or multimer-negative single cells.

To summarise, together with the previous observation that $PD1^+$ T cells are enriched for expanded ubiquitous TCRs, the fact that multimer positive T cells were also enriched for those was coherent with the hypothesis that this subset of TCRs identifies a population of cancer-specific T cells.

3.2.8 The number of expanded ubiquitous TCRs correlates with immune related gene signatures in matched RNA-Seq

Lastly, the bulk RNA-Seq data available was leveraged by computing the Danaher immune score for each tumour region (see Materials and Methods) and comparing it to the number of expanded regional and ubiquitous TCRs (Figure 3.22). Remarkably, the number of expanded ubiquitous TCRs was systematically positively correlated with several gene modules (amongst which the CD8 score, the NK score and the Th1 score showed a significant correlation) whereas the number of

expanded regional TCRs was systematically negatively correlated to the same gene modules, often in a statistically significant manner.

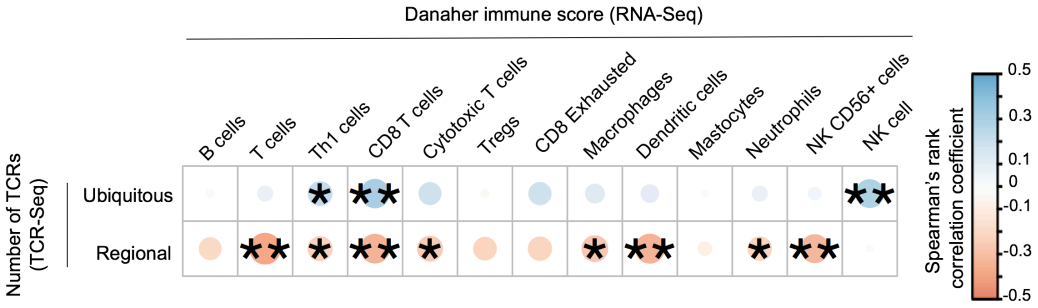

Figure 3.22 Summary of correlations between the number of expanded ubiquitous and regional TCRs and RNA derived immune scores

Correlation between the numbers of expanded intratumoural ubiquitous and regional TCR sequences and the transcriptional expression score (geometric mean) for a selection of Danaher immune scores. The area and colour of the circles correspond to the magnitude of the Spearman's rank correlation coefficient. *$P < 0.05$; **$P < 0.01$ after Bonferroni correction; $n = 32$ patients.

3.2.9 Expanded intratumoural TCRs can be found in matched PBMC samples at the time of resection and at follow-up

The associations between expanded TCR numbers and the genomic and transcriptomic data support the hypothesis that ubiquitous and regional TCRs identify a key population of cancer specific T cells in TCR-Seq data. For this work to be potentially utilised in future clinical applications, it would be highly valuable to be able to access those T cells through blood, without having to rely on multi-region intratumoural sampling. To this end, expanded ubiquitous and regional TCRs were searched for in matched blood samples and the corresponding proportions were computed (Figure 3.23). For both groups of TCRs, values were highly heterogeneous across the cohorts, ranging from 0 (none of the TCRs could be detected in blood) to 1 (all TCRs could be detected in blood).

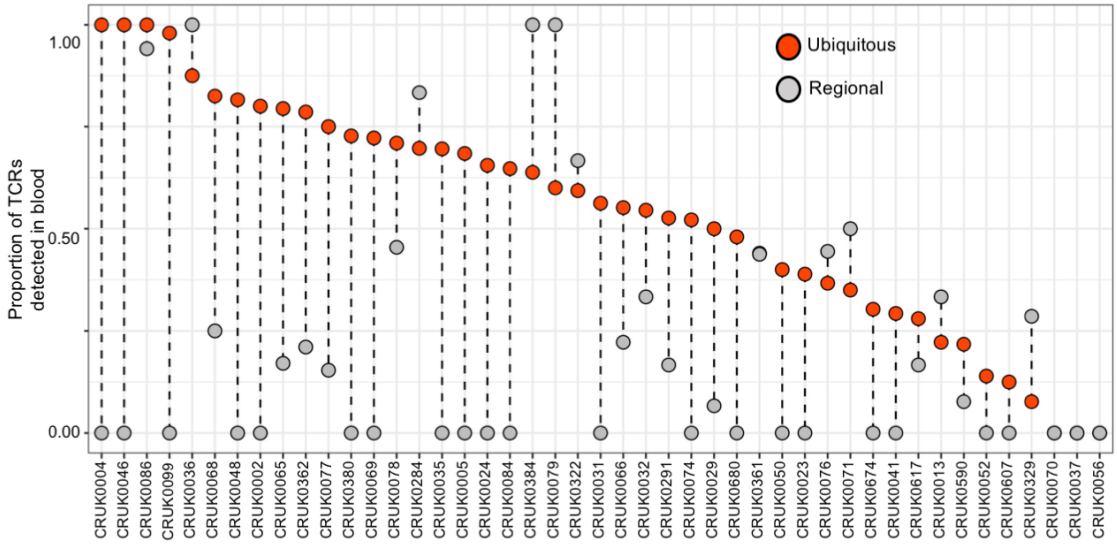

Figure 3.23 Proportion of expanded intratumoural ubiquitous and regional TCRs detected in peripheral blood, per patient

The proportion of expanded intratumoural ubiquitous and regional TCRs detected in peripheral blood from the same patient at the time of primary NSCLC surgery, with patients ordered by descending proportion; $n = 45$ patients.

Expanded ubiquitous TCRs displayed much greater detection rates in blood than regional ones. In addition, when looking at those TCRs that can be found in the blood, expanded ubiquitous TCRs had significantly higher expansion levels in matched blood (Figure 3.24).

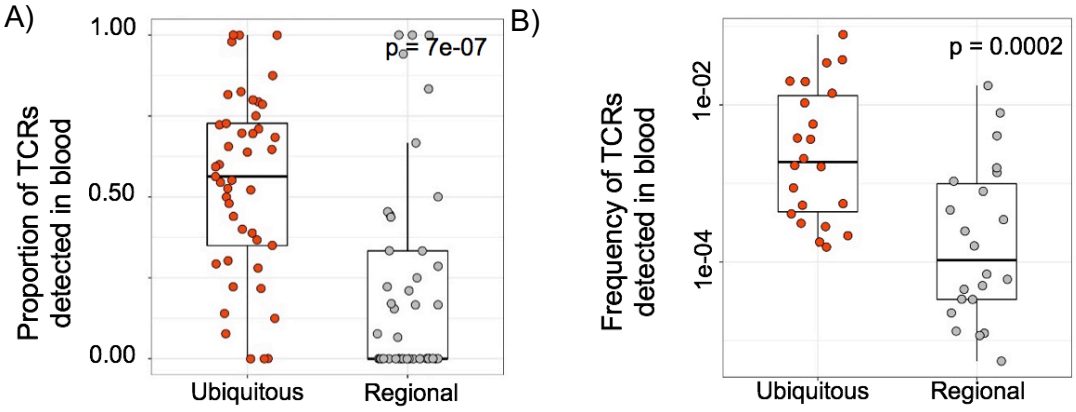

Figure 3.24 Summary of proportion and frequency of intratumoural expanded ubiquitous and regional TCRs in matched peripheral blood

(A) Summary of the proportion of expanded intratumoural ubiquitous (red circles) and expanded intratumoural regional (grey circles) TCRs detected within the blood

for all patients (the one-sided Mann–Whitney test *P* value is shown; $n = 45$ patients). (B) The frequency (number of TCR sequences detected, as a proportion of the total number of TCRs) of expanded intratumoural ubiquitous (red circles) and regional (grey circles) TCRs in the peripheral blood at the time of NSCLC surgery (the one-sided Mann–Whitney test *P* value is shown; $n = 42$ patients and 22 patients for ubiquitous and regional, respectively). Only TCRs actually detected in blood were used for this analysis.

To summarise, when compared to regional TCRs, not only were ubiquitous TCRs found more often in the blood but they were also found at higher expansion levels. This might be the result of the antigenic signal driving expansion of ubiquitous TCRs being stronger than the regional one because it is present across the entire tumour, leading to higher levels of emigration into the circulation.

The weak overlap between tumour and blood for the regional TCRs might also mean that the T cells that express regional TCRs have a lower replenishment capacity, which could lead to an impaired protection capacity and could explain the correlation observed in matched bulk RNA-Seq described in the previous section.

To explore this observation further, the same analysis was repeated in the 14 follow-up blood samples that were available (Figure 3.25). Once again, ubiquitous and regional TCRs behaved differently: intratumoural ubiquitous TCRs were found less often in follow-up bloods than in baseline bloods whereas no difference was observed for regional TCRs. The drop in ubiquitous TCRs frequency may be driven by the surgical debulking of the tumour which is accompanied by the loss of the antigenic signal. On the other hand, the lack of drop within the regional TCR compartment is coherent with the previous observations and in line with the hypothesis that these TCRs identify a population of T cells that are less relevant than the ubiquitous one. In addition, as a control, the same analysis was repeated for TCRs expanded in the non-tumour lung compartment and showed that, similarly to regional TCRs, no longitudinal drop in frequency was observed in the blood following surgery.

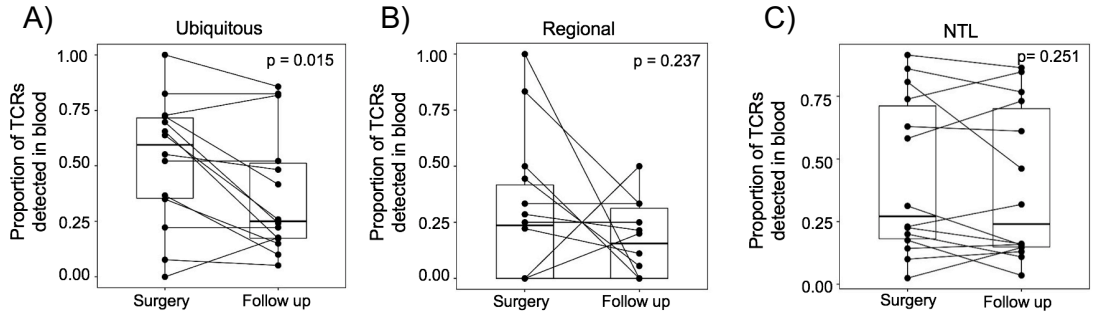

Figure 3.25 Summary of proportion and frequency of intratumoural expanded ubiquitous and regional TCRs in matched follow-up peripheral blood

The proportion of expanded intratumoural ubiquitous (A), regional (B) and non-tumour (C) TCRs that were detected in the blood at the time of primary NSCLC surgery and at routine follow-up (The one-sided Mann–Whitney test *P* values are shown; $n = 14$ individual patients).

Finally, to assess whether the frequency drop of expanded ubiquitous TCRs in the blood post-surgery was permanent or reversible, a longitudinal analysis of blood samples was performed for four patients for whom three separate time-points were available. Interestingly, a variety of possible temporal trajectories was observed: for CRUK0013 and CRUK0329 for instance, some ubiquitous TCRs disappeared at follow-up but reappeared at disease recurrence, some totally disappeared, some were found at all time-points and some were only detectable at recurrence, perhaps mirroring the parallel evolution of the tumoural genomic clonal composition and the corresponding antigenic landscape.

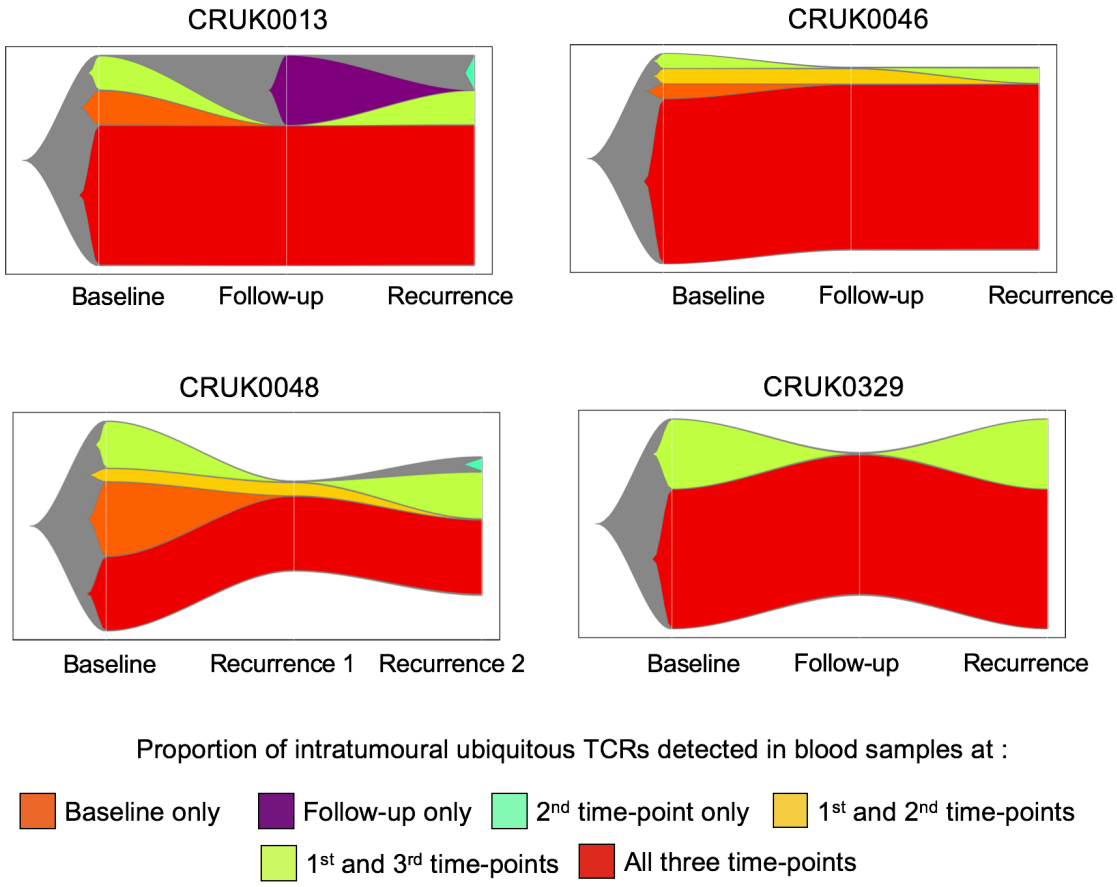

Figure 3.26 Representation of longitudinal evolution of intratumoural expanded ubiquitous TCRs prevalence in blood samples

The relative proportion of expanded intratumoural ubiquitous TCRs with different patterns of occurrence in peripheral blood taken at three longitudinal time-points for patients CRUK0013, CRUK0046, CRUK0048 and CRUK0329. Expanded intratumoural ubiquitous TCRs found at all three time points are shown in red.

3.3 Chapter discussion

The results presented in this chapter are consistent with a model in which TILs comprise a mixture of bystander T cells, attracted by the inflammatory response in the TME, and tumour-specific T cells, recognising tumour antigens (Bengsch et al., 2018; Ganesan et al., 2017; Simoni et al., 2018). In the context of treatment-naive NSCLC, applying a threshold on TCR frequency to label TCRs as “expanded” identified a population correlating with TMB, a primary antigenic source in lung cancer previously described to give rise to immunogenic neoantigens (McGranahan et al., 2016).

These clonotypes were highly enriched in tumour samples when compared to adjacent non-tumour tissue and could be further subdivided into ubiquitous and regional, depending on their multi-region frequency distribution within a patient's tumour. Just as the number and respective proportions of ubiquitous and regional mutations hugely varied between patients in NSCLC (Jamal-Hanjani et al., 2017), so did ubiquitous and regional TCRs. In addition, they correlated with one another, suggesting that the spatial distribution of expanded TCRs is deterministic and is driven by the corresponding shape of the mutational landscape.

The hypothetical antigenic spatial constraints that drive the distribution heterogeneity of expanded TCRs was further explored through clustering of similar CDR3 chains, which have been shown to identify antigen specific TCR subgroups (Glanville et al., 2017; N. Thomas et al., 2014). In this setting, the composition of subgroups containing ubiquitous or regional TCRs was coherent with their inferred specificity and the expected distribution of TCRs specific to the same ubiquitous or regional neoantigen.

Integrating these data with bulk RNA-Seq, and single-cell RNA-Seq of neoantigen reactive T cells comforted the hypothesis that expanded ubiquitous TCRs identify a population of tumour-specific cytotoxic T cells. Finally, these clonotypes could be detected in matched blood samples at the time of tumour resection but also at subsequent time-points when available which indicates that these specificities are stable over time and could be leveraged by targeted therapies.

This study has several limitations. Firstly, although the method described here enables to identify a T cell population likely to be enriched for tumour specificity, none of the metrics computed were found to correlate with clinical outcome. This observation was puzzling given that TMB in NSCLC is predictive of a longer time to disease recurrence (Ghorani et al., 2020) and response to anti-PD-1 checkpoint blockade therapy (Rizvi et al., 2015). More work is needed to characterise these TCRs but the limited flow cytometry data presented here suggests that they may belong to dysfunctional T cells, which would be in line with recent findings showing that such phenotypes are typically enriched for highly expanded TCRs in lung (Guo et al., 2018) and melanoma (H. Li et al., 2019). This would explain why they do not

provide protection and would mean that clinically leveraging these specificities would require the ability to identify expanded ubiquitous clonotypes and reverse their phenotype.

Secondly, the exceptional multi-region sampling that TRACERx offers was fully leveraged to identify expanded ubiquitous TCRs. Most studies or clinical settings do not have the resources to sample multiple regions per tumour which limits comparisons to other TCR data sets. It would be valuable to explore ways to define ubiquitous and regional TCRs that do not rely on systematic multi-region sampling.

Lastly, although the bioinformatics analysis strongly supports that expanded ubiquitous and regional TCRs are specific to ubiquitous and regional mutations, it remains to be formerly proven. The data presented on tetramer sorted T cells for one patient are convincing but relies on a single patient. Also, non-synonymous mutations were exclusively used to represent the antigenic signal of NSCLC, which is a simplistic view. Other antigenic sources have been described in lung cancer (Litchfield et al., 2021) and should be investigated in this model.

Chapter 4. Determinants of anti-PD1 response and resistance in clear cell renal cell carcinoma

4.1 Introduction

In contrast to NSCLC and some other solid cancers, TMB does not associate with better prognosis in RCC in the treatment naive setting nor does it in the context of checkpoint blockade therapy (Braun et al., 2020; McDermott et al., 2018; Motzer, Robbins, et al., 2020). Several other sources of antigen and their link to clinical response to CPI have been studied, including peptides derived from frameshift insertions and deletions (Turajlic et al., 2017), but the antigenic landscape of RCC remains elusive.

Interestingly however, RCC sits amongst the most immune-infiltrated cancers (Ricketts et al., 2018; Rooney, Shukla, Wu, Getz, & Hacohen, 2015), which indicates that, although it remains to be extensively characterised, a strong antigenic signal seems to be present and attracting T cells to the TME. As TCR sequencing analysis offers a valuable opportunity to study the specificity of the immune response without knowledge of the antigen, it appears to be a relevant tool to apply to the understanding of the biology of RCC.

ADAPTeR is a phase II, single-arm, open-label study of nivolumab in treatment-naive metastatic ccRCC. Whole-exome sequencing (WES); RNA sequencing; TCR repertoire sequencing; multiplex immunofluorescence, flow cytometry and single-cell RNA sequencing were performed across longitudinal, multi-region fresh tissue samples from each patient. 15 patients were enrolled from October 2015 to June 2018 and clinical benefit was defined as partial response or stable disease (see Material and Methods) for more than six months until disease progression (hereon termed ‘responder’; five patients). “pre-treatment” and “post-treatment” refer to the baseline and week nine time-point, respectively.

In this chapter, I firstly identify in RNA-Seq data a gene expression profile consistent with specific activation of T cells through TCR signalling in responding patients. Secondly, in matched TCR-Seq data, I identify a population of TCRs with features

suggestive of tumour specificity that are maintained upon treatment and associate with clinical response. Lastly, in a matched data set of single cell RNA-Seq, I show that specific structures of the TCR repertoire in a responding patient correspond to a subset of T cells that preferentially bound to nivolumab.

The ADAPTeR project is a collaborative effort involving multiple research groups. Details of individual contributions to the work presented in this chapter are as follows. Sample acquisition, preparation and sequencing were jointly done by Lewis Au (oncologist at the Royal Marsden, clinical research training fellow in the Turajlic group), Nicos Fotiadis (Consultant Interventional Radiologist at the Royal Marsden), the ADAPTeR trials team, Emine Hatipoglu (oncologist at the Royal Marsden, PhD student in the Quezada group) and Imran Uddin (single cell genomics technician, single cell genomics facility, UCL, formerly lab technician in the Chain group). Processing of raw FASTQ sequence data from TCR repertoire sequencing and single cell TCR/RNA-Seq was performed by Tahel Ronel (research fellow at the UCL Cancer Institute Cancer Immunotherapy Accelerator, previously post-doctoral fellow the Chain group) and Gordon Beattie (single cell genomics research fellow, single cell genomics facility, UCL), respectively. I performed the processing of raw FASTQ sequence data from RNA sequencing and all downstream bioinformatics analysis and data visualisation for both RNA-Seq and TCR-Seq data. The data presented in this chapter of the thesis forms sections of a submitted manuscript currently under review.

4.2 Results

4.2.1 Nivolumab induces upregulation of immune related transcripts

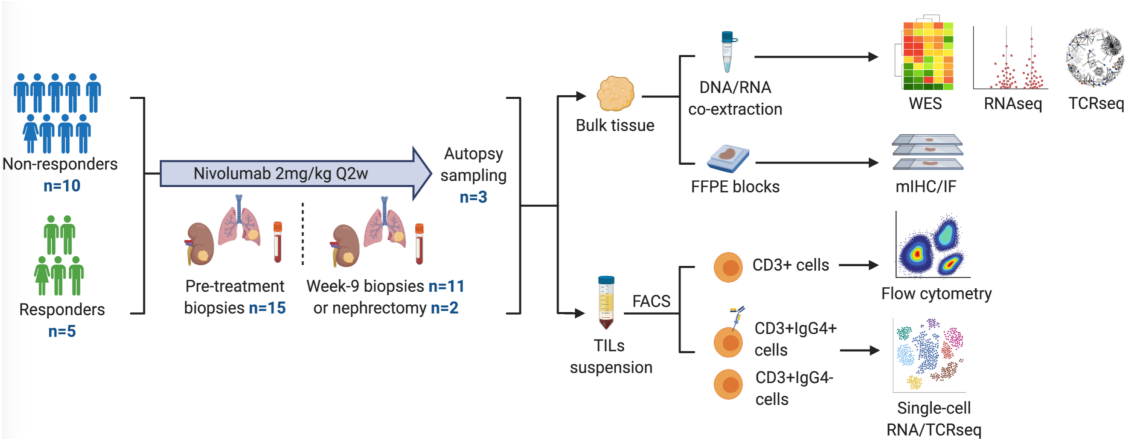

Figure 4.1 Description of ADAPTeR patient cohort

Patient study identification, best response to nivolumab by RECIST criteria, tumour sampling time-point, and number (n) of samples subjected to varied profiling techniques are shown. TIL := tumour-infiltrating lymphocyte; FACS := fluorescence-activated cell sorting; WES := whole-exome sequencing; mIHC = multiplexed immunohistochemistry; IF := immunofluorescence.

In order to detect pre-treatment transcriptomic differences between responders and non-responders, differential gene expression analysis (DGEA) was performed on 36 samples across 14 patients (Figure 4.1, see Materials and Methods). In order to visualise the prevalence of immune related transcripts in the set of genes that were differentially expressed, the Danaher immune score (Danaher et al., 2017) genes were overlaid onto the 1761 (resp. 1621) transcripts that were found to be up-regulated (resp. downregulated) in responders compared to non-responders. The genes of the Danaher immune score that mapped to the DGEA were exclusively found in the up-regulated compartment, indicating that immune infiltration prior to treatment was associated with clinical response (Figure 4.2, A).

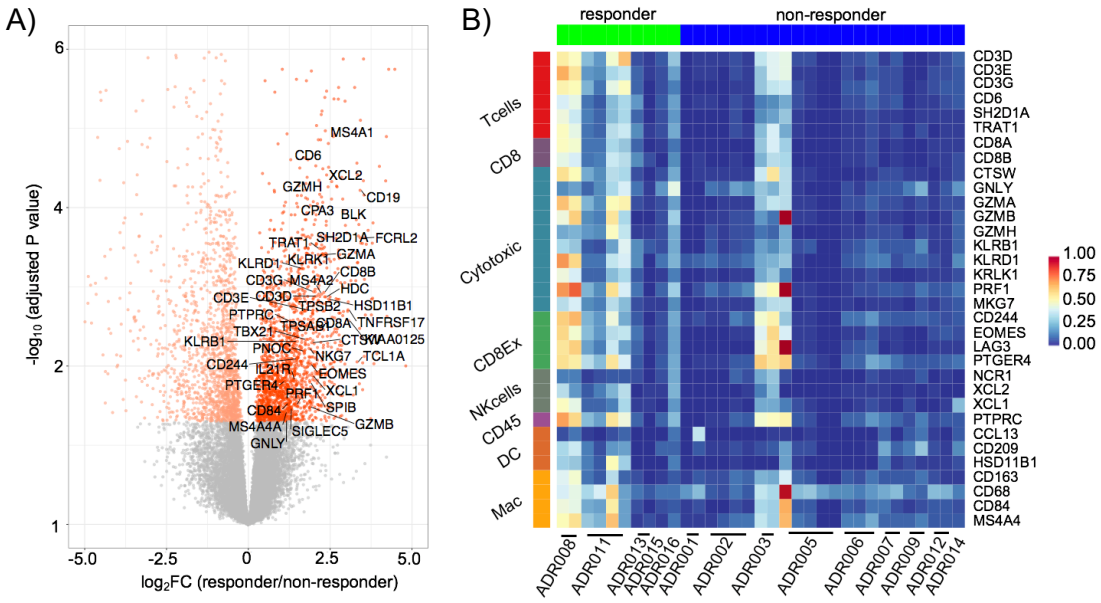

Figure 4.2 Responders versus non-responders pre-treatment differential gene expression analysis and transcriptomic immune deconvolution

(A) Transcripts differentially regulated pre-treatment between responders and non-responders (n = 33 samples, 14 patients, negative binomial Wald test, Benjamini–Hochberg corrected *P* values). 3,382 transcripts were differentially regulated (FDR<0.05), the ones that overlap with the Danaher immune score gene list are labelled. (B) Heatmap showing the relative expression (z scores) of genes from eight selected Danaher immune modules in pre-treatment samples.

When dissecting further the Danaher immune score (Figure 4.2, B), gene modules identifying lymphoid infiltration (“Tcells”, “CD8”, “Cytotoxic”, “CD8Ex”, “NKcells”) were preferentially enriched in responders compared to the myeloid ones (“DC”, “Mac”).

Next, to understand how checkpoint blockade therapy impacts the transcriptomic landscape that was observed at baseline, the same analysis was repeated in the 30 RNA samples (collected across 11 patients) available post-treatment (Figure 4.3).

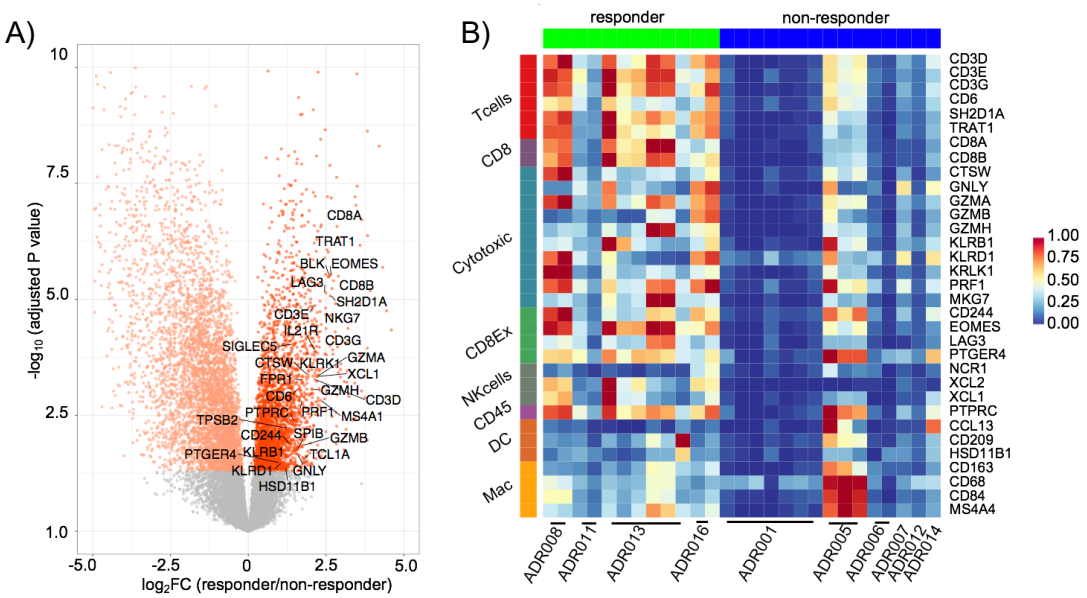

Figure 4.3 Responders versus non-responders post-treatment differential gene expression analysis and transcriptomic immune deconvolution

(A) Transcripts differentially regulated post-treatment between responders and non-responders (n = 27 samples, 10 patients, negative binomial Wald test, Benjamini–Hochberg corrected *P* values). 7,975 transcripts were differentially regulated (FDR<0.05), the ones that overlap with the Danaher immune score gene list are labelled. (B) Heatmap showing the relative expression (z scores) of genes from eight Danaher immune modules in post-treatment samples.

Whilst the pattern of enrichment of immune-related transcripts in responders post-treatment was similar to the one observed pre-treatment, the magnitude of this signal was much greater (Figure 4.3, A). With 2,837 upregulated transcripts and 5,138 downregulated ones, the total number of transcripts differentially regulated between responders and non-responders was more than twice the amount of what was observed pre-treatment. In addition, the observation made pre-treatment that the Danaher immune score was exclusively mapping to the upregulated transcripts was repeated post-treatment and the level of expression of the corresponding genes was higher post-treatment than pre-treatment (Figure 4.3, B).

When summarising the average value of each Danaher immune module per patient and per time-point, most of the lymphocyte modules behaved similarly (Figure 4.4). For the pan-T cells module, for instance (which was the only one to differ significantly

between responders and non-responders at both time-points), treatment pushed both responders and non-responders towards higher scores. This resulted in a level of T cell infiltration in non-responders post-treatment equivalent to the one observed pre-treatment in responders. However, the lymphocyte immune scores post-treatment in responders were higher than those in non-responders, at either time-point.

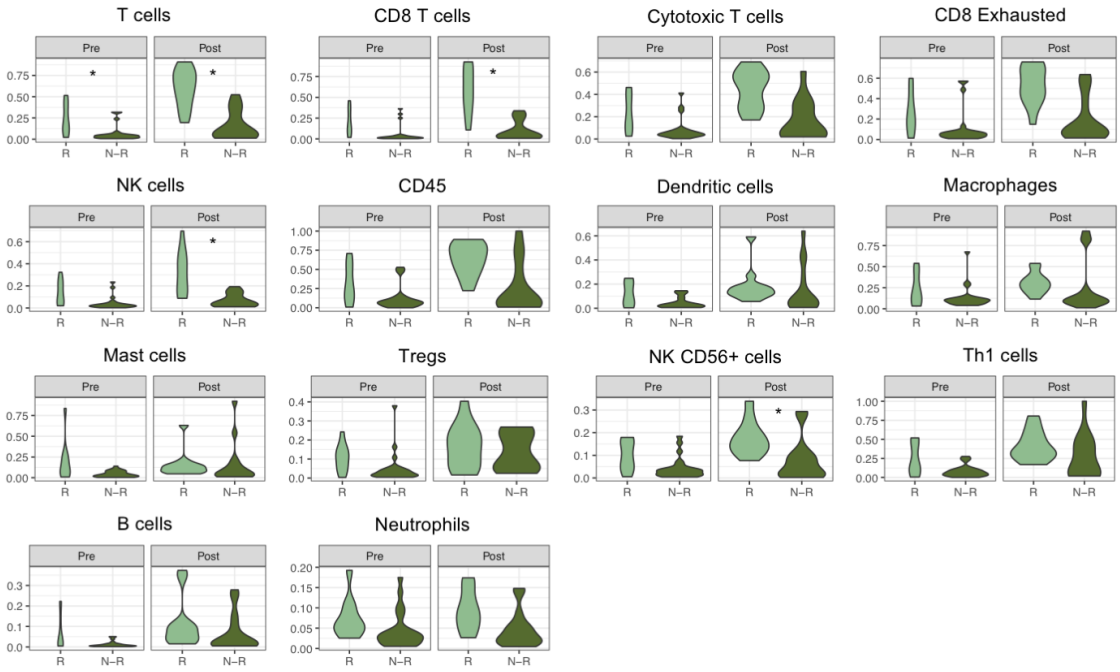

Figure 4.4 Danaher immune scores pre- and post-treatment comparing response groups

Summary of relative Danaher immune scores expression levels per response group and per time point. The two-sided Mann–Whitney test performed on one value per patient and per time-point (score averaged by median value across biopsies if several available at a given time point), significant *P* value are indicated (*: *P*<0.05; **: *P*<0.01). R – responders; N-R - non-responders. *n* = 14 patients and 10 patients for pre- and post-treatment, respectively.

4.2.2 Responders present exclusive upregulation of antigen recognition pathways

Next, to determine whether the association between immune infiltration and clinical response was simply due to the quantity of immune cells present in the TME, or if some other qualitative change that would have been undetectable with the method described above was playing an important role in these observations, Gene Set

Enrichment Analysis (GSEA) was run on the set of differentially expressed transcripts, at both time-points (Figure 4.5**,** see Materials and Methods).

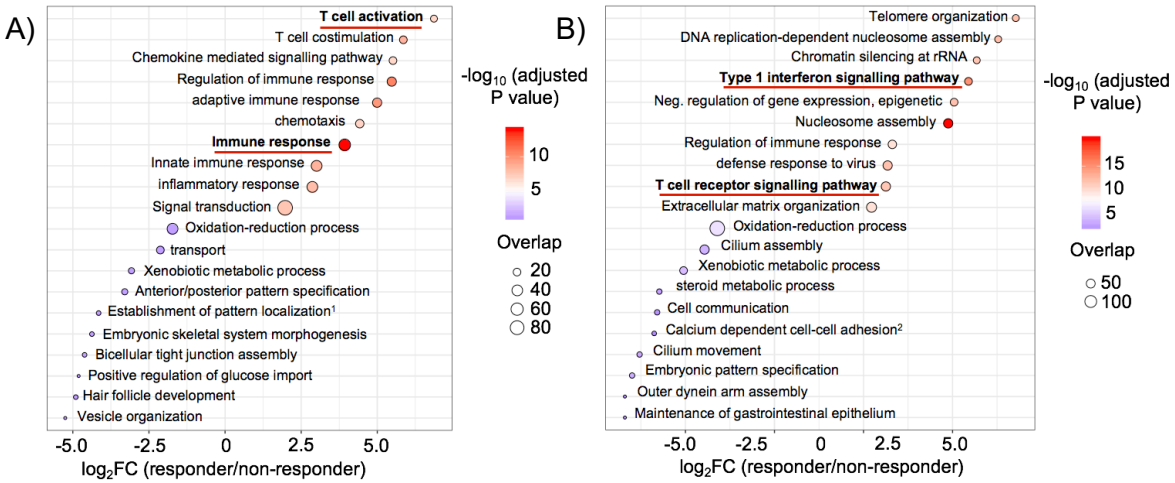

Figure 4.5 Gene set enrichment analysis on transcripts differentially regulated in responders at both time-points

GSEA analysis of genes preferentially up-regulated and down-regulated pre-treatment (A) or post-treatment (B) in responders. Overlap (n): number of significant genes from a pathway (hypergeometric test). *n* = 14 patients and 10 patients for pre- and post-treatment, respectively.

Although a number of pathways significantly enriched in responding patients were shared between pre- and post-treatment GSEA, there was a clear difference between time-points. Pathways enriched in pre-treatment responders describe a general enrichment for immune infiltration without any particular flavour of specificity at this point in time. However, post-treatment, enrichment of pathways such as “Type 1 interferon signalling pathway” or “T cell receptor signalling pathway” indicate a different quality of infiltration in responding patients, likely driven by antigenic stimulation.

4.2.3 Heterogeneity of the TCR repertoire in metastatic RCC

To further investigate the potential role of tumour specificity in the mechanism underlying response to Nivolumab in this context, the TCR repertoires from 63 tumour samples and 29 PBMC samples from 14 patients pre- and post-treatment were sequenced.

The analysis of repertoire similarity revealed a large range of heterogeneity within the cohort (Figure 4.6, see Materials and Methods), with a large variability of repertoire similarity values across samples from the same time-point, at the cohort level and at the patient level.

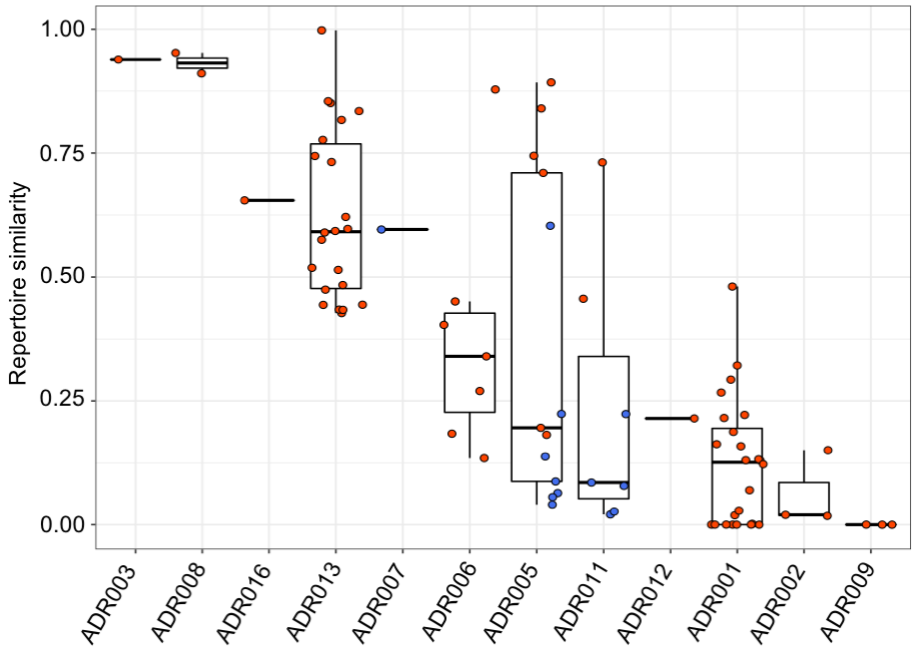

Figure 4.6 Intra-patient TCR repertoire heterogeneity

The pairwise intratumoural TCR repertoire similarity is shown for each patient. Each circle represents a comparison between two samples from the same patient and at the same time-point (*n* = 87 total comparisons from 12 patients). Red (resp. blue) circles indicate a pair of biopsies originating from the same site (resp. different metastatic sites).

The ubiquitous/regional dichotomy described in Chapter 3 could not be robustly repeated for this project for three main reasons. Firstly, the nature of the material collected for sequencing resulted in a significantly lower number of TCRs retrieved per individual sample. Only pooling the different samples available for the same patient and at the same time-point enabled to retrieve a median number of unique β-chain (3,644) per time point and per patient equivalent to what was observed in Chapter 3 (3,711). Secondly, applying the same selection criteria of a minimum of three samples per time-point and per patient used in Chapter 3 in order to robustly define ubiquitous and regional TCRs would result in a significant reduction of the cohort as only six patients would meet this criterion at both time-points. Lastly, as

opposed to Chapter 3's setting, multiple samples from the same patient and at the same time point in ADAPTeR were not always obtained from the same metastatic site, which adds difficulty to the interpretation of TCR repertoire spatial heterogeneity for that cohort.

Although heterogeneity could not be studied systematically at cohort level, some individuals could be studied in more details. ADR001 (non-responder) and ADR013 (responder) post-treatment nephrectomy samples met the criteria described above which enabled to explore spatial heterogeneity further for these two patients. Out of the seven (resp. six) post-treatment nephrectomy-derived tumour regions available for ADR001 (resp. ADR013), 5 were selected for each patient, based on similarity of total number of TCRs retrieved, for comparativeness and visualisation purpose. For both patients, each pair of tumour regions was compared with the cosine metric (Figure 4.7).

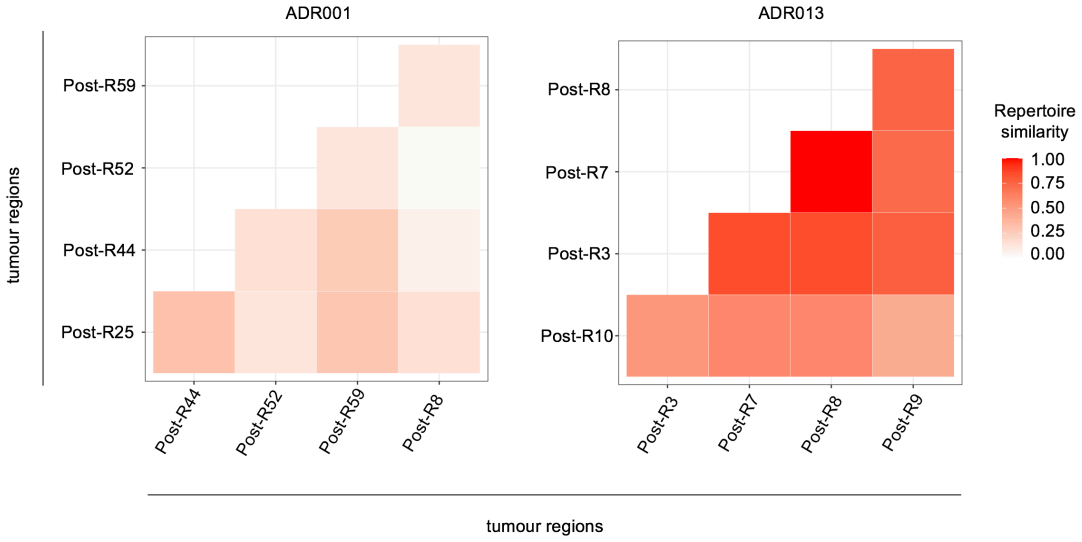

Figure 4.7 Spatial heterogeneity in post-treatment nephrectomy for ADR001 and ADR013

Heat maps showing the pairwise similarities of a selection of five biopsies in the post-treatment nephrectomy for ADR001 (left) and ADR013 (right).

Strikingly, pairwise similarity scores in ADR013 post-treatment samples reached an average value of 71% against 17% for ADR001. This indicated that the intratumoural

immune infiltrate in ADR013 post-nivolumab was more homogenous in terms of specificity than ADR001's one.

Although these observations are individual case studies, they emphasise the differences in heterogeneity in the TME that exist in this clinical scenario. This was coherent with the model derived from longitudinal analysis of combined regions from the cohort that will be detailed in the rest of the chapter.

4.2.4 Nivolumab treatment does not have a global effect on the intratumoural or the blood TCR repertoire

To test whether anti-PD1 treatment induces a global reshaping of the TCR repertoire in metastatic RCC, the clonality score for each tumour sample and each blood sample was computed (Figure 4.8**,** see Materials and Methods).

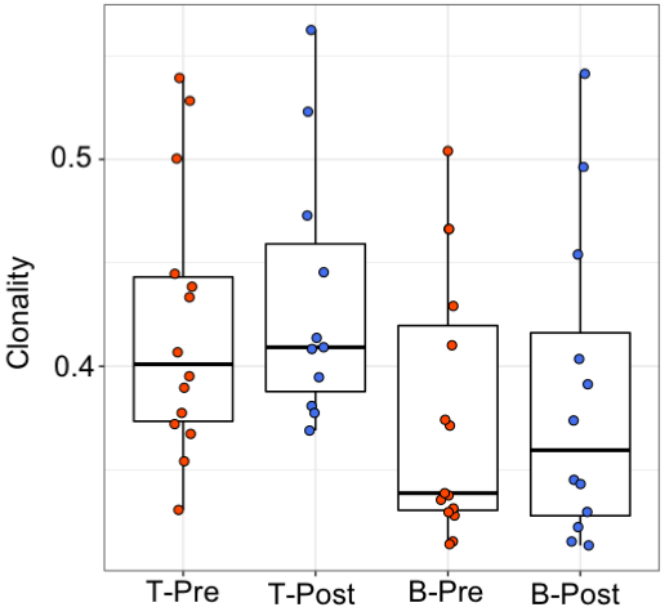

Figure 4.8 Intratumoural and blood pre- and post-treatment clonality score

Boxplots summarising the intratumoural (T-) and blood (B-) TCR repertoire clonality score pre-treatment (-Pre) and post-treatment (-Post). Tumour and blood clonality scores are shown for each patient at each time-point; *n* = 14 patients, 11 patients, 14 patients and 11 patients for T-Pre, T-Post, B-Pre and B-Post, respectively.

Unlike the significant increase in T cell infiltration measured in the RNA-Seq data as described in Section 4.2.1, the clonality score was not significantly higher pre- or post-treatment, neither intratumourally nor in the blood. In addition, the normalised count of TCRs that expand (significantly higher frequency post-treatment) was not different from the number of TCRs that contract (significantly lower frequency post-treatment neither) in the tumour, nor in the blood (Figure 4.9).

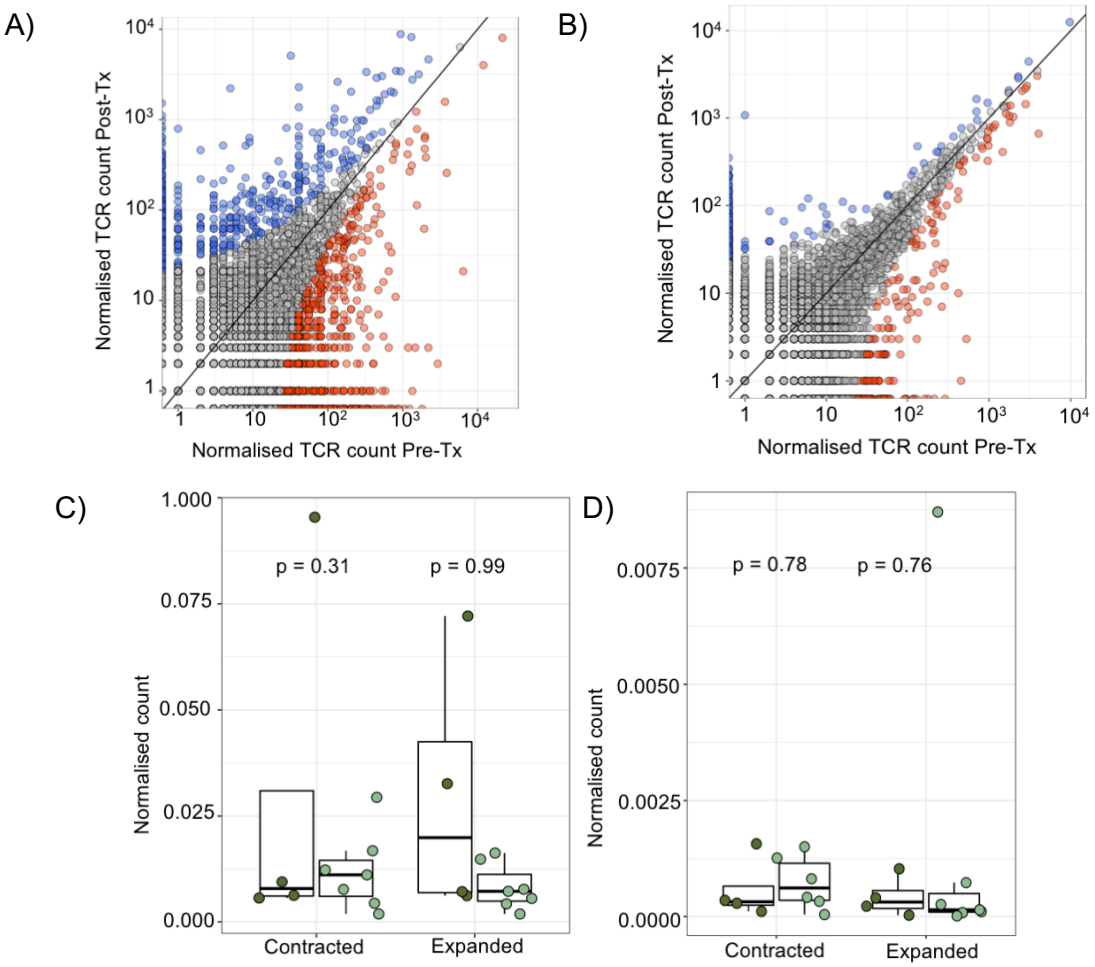

Figure 4.9 Intratumoural and blood longitudinal clonal expansion and contraction

Correlated TCR sizes in tumour (A) and blood (B) samples. Scatter plots of tumour and blood TCR abundance pre- and post-treatment are shown for all patients. TCRs are coloured by expansion/contraction status (see Materials and Methods). The number of intratumoural (C) and blood (D**)** TCRs labelled as expanded or contracted between time-points, per patient, normalised for the total number of TCRs tested. Two-sided Mann–Whitney tests *P* value shown; *n* = 11 patients and 12 patients for tumour and blood, respectively.

It must be noted however that the blood repertoire was overall much more stable than the intratumoural one, with a tenfold increase of the normalised number of expanded and contracted TCRs in the tumour compared to blood. This might indicate that the effects of anti-PD-1, or at least the ability to characterise it through TCR-Seq analysis, is of much greater magnitude in the TME, compared to the blood.

4.2.5 Longitudinal maintenance of expanded TCRs correlates with clinical response

As Nivolumab treatment did not seem to have a global effect on the overall repertoire clonality, clonality scores were calculated separately for responders and non-responders, both for tumour and blood (Figure 4.10).

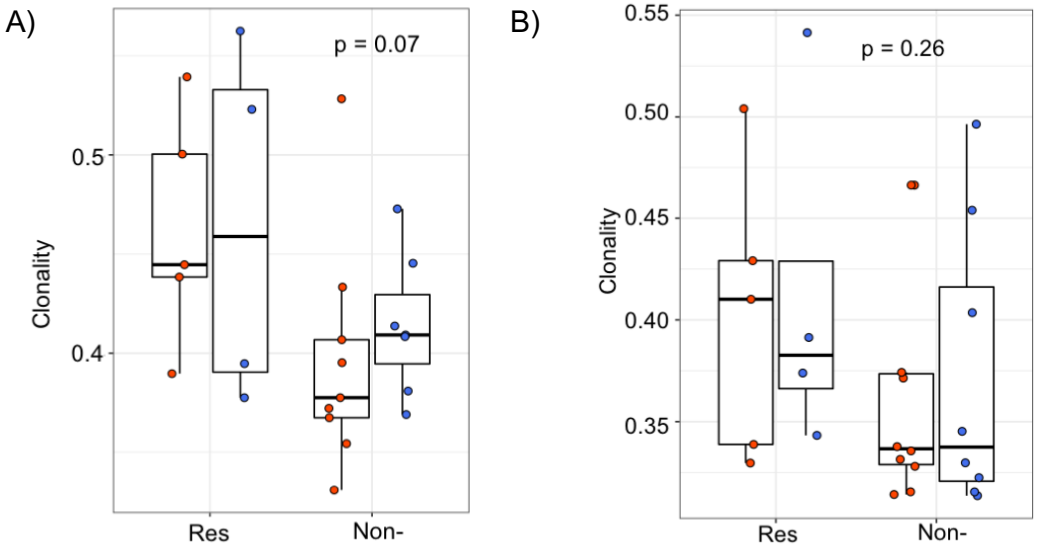

Figure 4.10 Intratumoural and blood pre- and post-treatment clonality score per response group

The intratumoural (A) and blood (B) TCR repertoire clonality score pre-treatment (red circles) and post-treatment (blue circles) is shown for each patient. Patients are split between responders (Res) and non-responders (Non-). Mixed-effect model *P* values shown; *n* = 14 patients, 11 patients, 14 patients and 11 patients for pre- and post-treatment tumours and pre- and post-treatment bloods, respectively.

Although it did not reach significance, mixed-effect modelling identified a trend towards higher clonality values in responders both pre- and post-treatment in the tumour, but not in the blood. In addition, the comparison of pre-treatment

intratumoural clonality values with the two-sided Mann-Whitney test reached statistical significance (P = 0.042). This observation suggested that the intratumoural TCR repertoire of responding patients contained more expanded clonotypes, both pre- and post-treatment. This was not observed in blood, suggesting that the process driving this signal in the tumour was either non-existent or too weak to be detected in the blood, coherent with what was observed in Section 4.2.4.

The high clonality values observed at both time-points in responders was raising the following question: were the TCRs driving high clonality pre- and post-treatment the same ones or not? To address this, similarity between individual patients' TCR repertoires at the two time-points was measured using the cosine metric (Figure 4.11, see Materials and Methods).

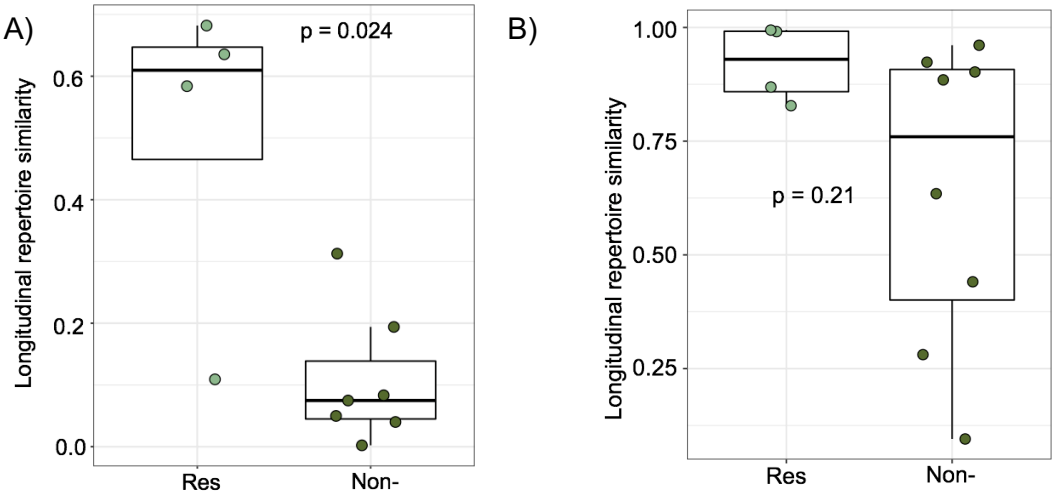

Figure 4.11 Intratumoural and blood longitudinal repertoire similarity score per response group

The intratumoural (A) and blood (B) similarity score between pre-treatment and post-treatment is shown for each patient. Patients are split between responders and non-responders. Responding patients exhibit greater intratumoural cosine score, with the two-sided Mann–Whitney test *P* value shown; n=11 patients for tumour and 12 patients for blood, respectively.

Strikingly, despite the small cohort size, responding patients displayed significantly higher intratumoural similarity scores than non-responders. This analysis was repeated in the blood samples, which did not display the same pattern across the entire cohort. Taken together, responding patients showed a trend for more

expanded TCRs than non-responders and these TCRs were more stable over time i.e. the clonal expansion pre-existing in the tumour was maintained upon nivolumab exposure.

In order to investigate this further, the frequency thresholds developed in the context of NSCLC (Chapter 3) were applied to both pre- and post-treatment samples of each patient in order to define expanded TCRs. To test whether focusing on expanded TCRs defined by such thresholds were indeed capturing the TCR population driving the differential clonality signal described above, the proportion of the repertoire occupied by TCRs labelled as expanded was correlated to the clonality score (Figure 4.12). It confirmed that the 0.002 expansion threshold (that previously proved useful to enrich for specificity whilst selecting a substantial number of TCRs compared to more stringent thresholds) was also relevant in this setting, identifying TCRs representative of the clonality signal.

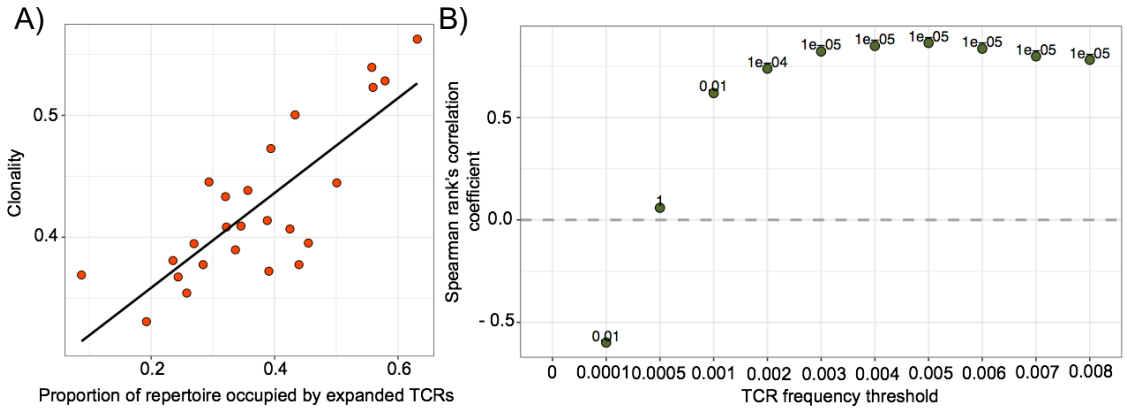

Figure 4.12 Relationship between the clonality score and the proportion of the intratumoural repertoire pre-treatment occupied by expanded TCRs

(A) The correlation between the proportion of the repertoire occupied by expanded TCRs defined by the 0.002 threshold and the clonality score, for each patient and each time point (B) The Spearman's rank correlation coefficient and *P* value (shown above each point; n=14 patients) for the relationship between the clonality score and the proportion of the intratumoural repertoire occupied by expanded TCRs defined by different frequency thresholds (ranging from all TCRs (threshold of zero) up to those found at a frequency of ≥8/1,000).

The greater stability of the intratumoural repertoire in responders described above (Figure 4.11) was further investigated from the perspective of the expanded TCRs

(Figure 4.13). When comparing the frequency of pre-treatment expanded TCR at both time-points, pre-treatment intratumoural expanded TCRs were found to be significantly less subject to clonal replacement in responders than in non-responders (see Materials and Methods). More precisely, the frequency of expanded TCRs pre-treatment was longitudinally stable in responders but decreased by more than 50% on average in non-responders. This suggested that stable stimulation of clonotypes occurred preferentially in responders' tumours.

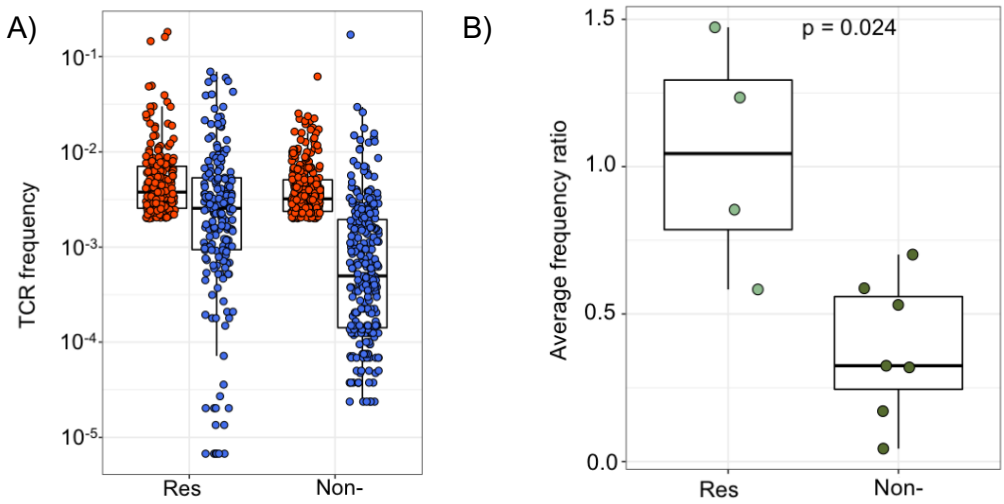

Figure 4.13 Longitudinal clonal replacement of pre-treatment expanded TCRs per response group

(A) The frequency distribution of the intratumoural expanded TCRs pre-treatment (red circles; *n* = 469 individual TCRs combined from 12 patients) and post-treatment (blue circles). Only TCRs that were detected post-treatment were included. (B) The arithmetic mean of Pre/Post frequency ratios of TCRs expanded pre-treatment, per patient (see Materials and Methods). Two-sided Mann–Whitney test *P* value shown.

4.2.6 The pattern of CDR3 sequences of maintained expanded TCRs display features of tumour specificity

Together with signals of antigen driven immune filtration detected in post-treatment RNA-Seq data as described in Section 4.2.2, the fact that expanded TCRs were preferentially maintained through treatment in responders strongly suggested that the T cell response differed significantly between responders and non-responders.

As discussed in Chapter 3, a recognised feature of the T cell response to specific antigen is the presence of groups of CDR3s with similar sequence features. The clustering algorithm based on amino acid triplet sharing (see Section 3.2.4) was applied to both pre- and post-treatment TCRs. The number of clusters identified was normalised per patient and per time-point (Figure 4.14, see Materials and Methods).

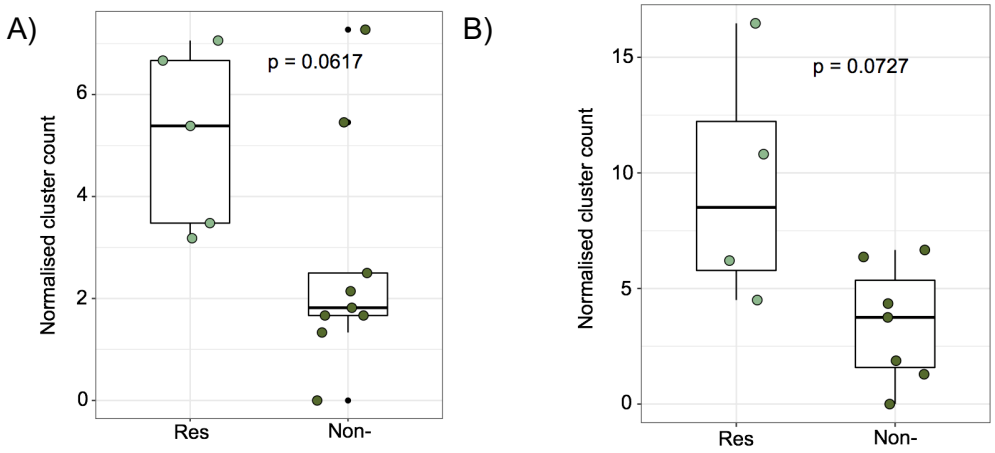

Figure 4.14 Normalised intratumoural cluster count pre- and post-treatment, per response group

The clustering algorithm was run on all patients, and the pre-treatment (A) and post-treatment (B) normalised number of clusters for the networks containing expanded sequences is shown. Two-sided Mann–Whitney test *P* value shown; *n* = 14 patients and 11 patients for pre- and post-treatment, respectively.

Although the differences did not reach statistical significance, this analysis identified a trend towards higher cluster counts in responders, at both time-points. Interestingly, the increase in cluster counts in both responders and non-responders and the higher number of clusters for responders resembled what was observed in terms of T-cell infiltration in Section 4.2.1.

Networks retrieved from ADR001 (non-responder) and ADR013 (responder) post-treatment nephrectomy samples were explored further (Figure 4.15). Remarkably, ADR013's post-treatment nephrectomy displayed a significantly more complex network structure than ADR001's post-treatment nephrectomy. This was in line with the observation that ADR013 displayed a more homogeneous intratumoural TCR repertoire at this time-point compared to ADR001 (see Section 4.2.3). Together,

these analyses support the hypothesis that a specific anti-tumour immune response amplified by exposure to nivolumab was present in ADR013 but not in ADR001.

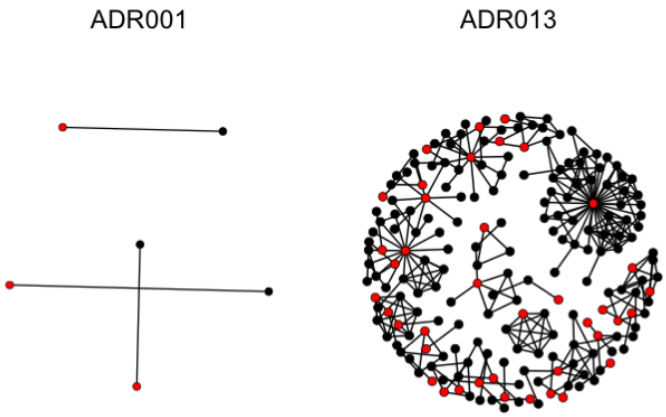

Figure 4.15 Representative intratumoural post-treatment cluster diagrams

Representative network diagrams of post-treatment intratumoural CDR3 β-chain sequences for patient ADR001 (left) and for patient ADR013 (right).

In conjunction with the maintenance of expanded TCRs, this analysis suggested that anti-PD1 treatment in metastatic RCC impacts the intratumoural immune infiltrate independently of the features of T cell infiltration, but highly benefits a T cell infiltrate already that is enriched for tumour specificity.

To examine this model more closely through the lens of the TCR repertoire, pre-treatment expanded TCRs were subdivided into “maintained” if they were also found to be expanded post-treatment and “replaced” if not. Then, for each patient, each maintained or replaced sequence was mapped to the pre-treatment clustering structure, as represented for patient ADR008 in Figure 4.16.

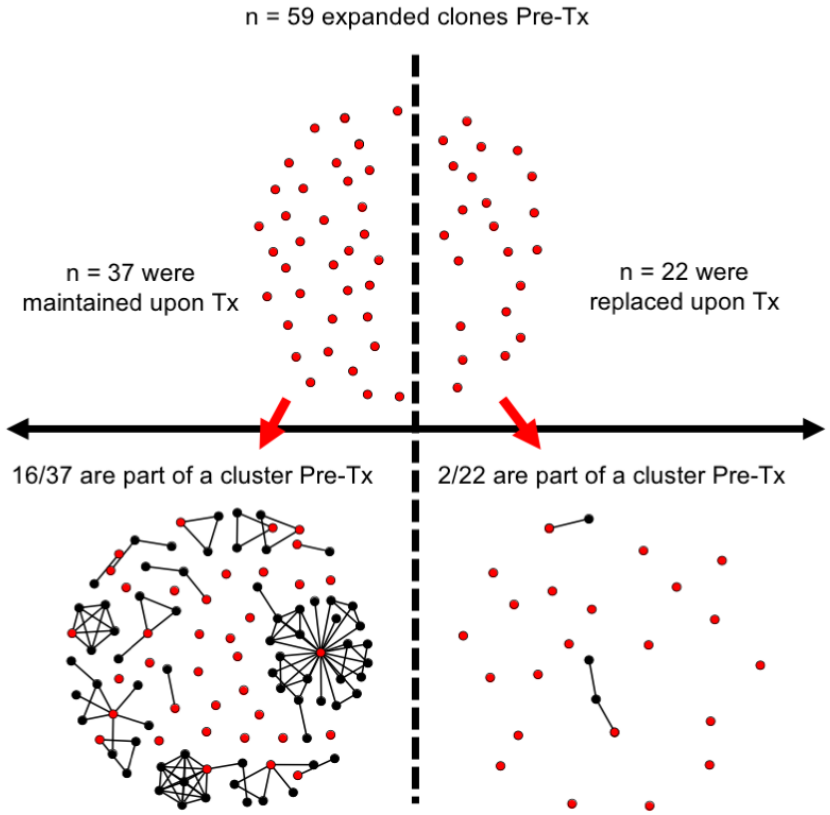

Figure 4.16 Representation of workflow analysis of maintained and replaced TCRs in pre-treatment clustering networks

Diagram illustrating pre-treatment clustering around maintained and replaced expanded TCRs for ADR008.

The relative proportion of the maintained or replaced pool of TCRs that were part of a CDR3 cluster was then computed for each patient independently of clinical response (Figure 4.17).

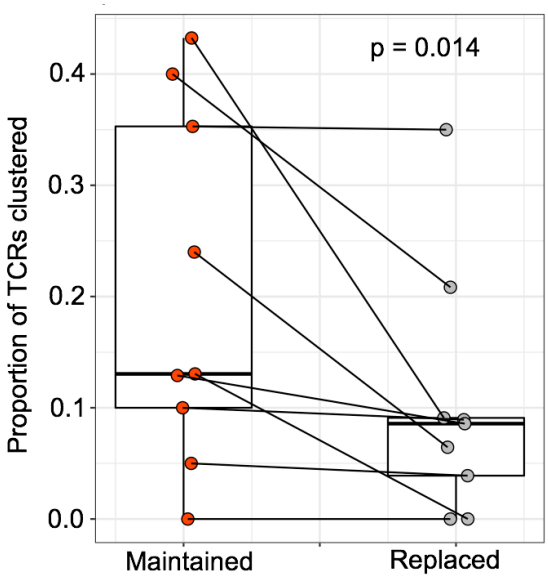

Figure 4.17 Proportion of maintained and replaced TCRs part of pre-treatment clustering networks, per patient

The proportion of pre-treatment expanded TCRs that are part of a cluster as depicted in Figure 4.16. TCRs were split between the ones that are also detected as expanded post-treatment and the ones that are not (respectively red circles and grey circles). Paired Wilcoxon signed-rank test *P* value shown.

Strikingly (as depicted for ADR008 in Figure 4.16) and without exception, the proportion of maintained TCRs that were part of a cluster of similar CDR3 sequences was significantly greater than the corresponding proportion of replaced TCRs.

Taken together, these data suggested that, in this cohort of patients, the clinical benefit of nivolumab preferentially associates with features consistent with the presence of tumour-specific T cells both from the perspective of TCR-Seq and RNA-Seq. The state of the TME pre-treatment presented prognostic clinical value and supported a model in which anti-PD1 benefits patients for whom tumours already contain strong features of tumour specific T cells prior to treatment. The data then suggested that treatment facilitates expansion of both pre-existing and novel tumour specific TCRs, as a function of the strength of the antigenic signal, which was detected to be much higher in responders via TCR-Seq analysis.

4.2.7 Paired single-cell RNA and TCR sequencing demonstrates enrichment of Nivolumab-bound cells in CDR3 clustering networks

Next, paired single-cell RNA- and TCR-Seq (scRNA/TCR-Seq) in drug-bound ($IgG4^+$ $CD3^+$) and unbound ($IgG4^-CD3^+$) immune cells was performed for ADR001 (non-responder) and ADR013 (responder) (see Materials and Methods). Individual cells were annotated with their corresponding VDJ information, which could then be related to the bulk TCR-Seq data discussed until now in this chapter. The transcriptomic analysis of this 10X dataset was performed by Gordon Beattie and will not be discussed in this thesis. In summary, it revealed that drug-bound cells had similar profiles in both patients, upregulating pro-inflammatory cytokine/chemokine genes and T-cell activation pathways. It also confirmed enrichment of expanded TCRs in ADR013 compared to ADR001, which were more prevalent in drug-bound cells compared to non-drug-bound cells and presented higher levels of cytotoxicity-associated gene expression (see Au et al., 2021 for more details).

Remarkably, many of the TCRs detected in single cells were present within the CDR3 amino acid triplet clusters for post-treatment repertoires of ADR001 and ADR013 as observed in the bulk-TCR data (displayed in Figure 4.15). We could classify each TCR in the networks into $IgG4^+$ or $IgG4^-$ based on the drug-binding status derived from the single cell data (if both $IgG4^+$ or $IgG4^-$ cells were identified by the same TCR, the most frequent label was retained). The clustering observed for ADR001 was very limited, preventing a definitive conclusion about the cluster-associated drug binding status. However, strikingly, a vast majority (89%) of post-treatment expanded clustered TCRs, both pre-existing and novel, were found to be drug-bound in ADR013 (Figure 4.18).

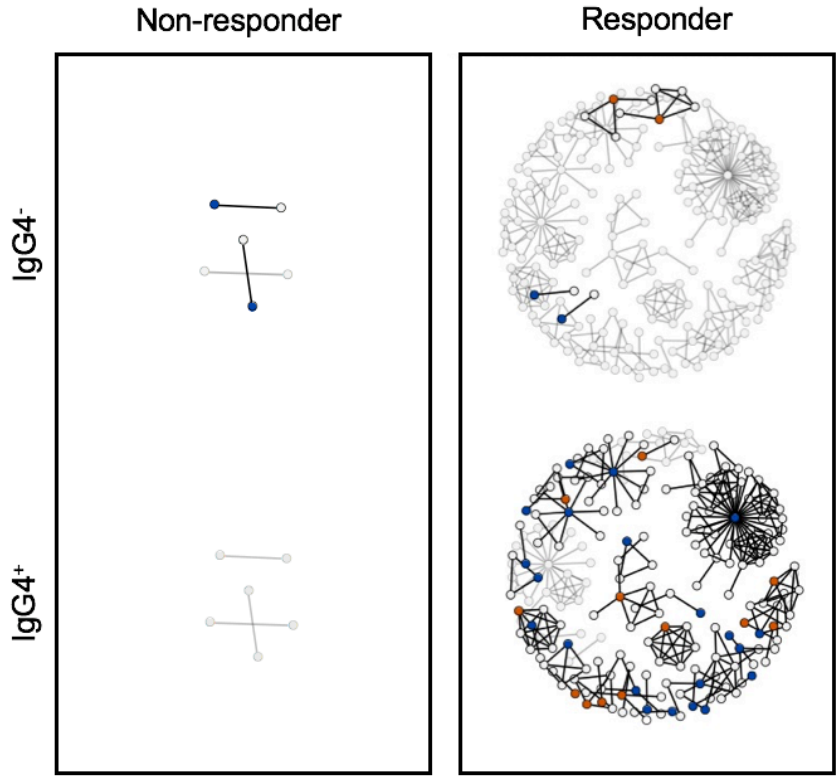

Figure 4.18 Drug-binding status of CDR3s within the network diagrams of post-treatment intratumoural sequences for ADR001 and ADR013

Clustering was performed within the bulk TCR-Seq data around expanded intratumoural TCRs, subdivided between TCRs that were expanded in the post-treatment repertoire exclusively (blue circles) and TCRs that were also expanded pre-treatment (orange circles). The network shows clusters for which at least one CDR3 was also detected in the scTCR-Seq. The network was then split between TCRs that were mapping to a majority of $IgG4^-$ cells (top panel) or a majority of $IgG4^+$ cells (bottom panel) in the single-cell data. The light grey circles represent the rest of clustering network derived from bulk post-treatment tissue, as shown in Figure 4.15.

This data shows that among the TCRs clustered by their similarity in CDR3s, which are expected to be enriched for tumour specific TCRs as demonstrated in this chapter and in Chapter 3, the majority are TCRs expressed by T cells bound to anti-PD-1 drug. Although the single cell analysis could only be carried out on two patients, the data provides further evidence for the hypothetical relationship between tumour specific T cell infiltration and expansion and the clinical impact of nivolumab in metastatic RCC.

4.2.8 Post-mortem detection of pre-treatment expanded maintained TCRs associates with non-progressive lesions

Finally, dynamics of maintained expanded TCRs could be explored in late metastatic events for one patient. Post-mortem sampling and subsequent TCR-Seq was performed for ADR005, who was co-recruited into the PEACE study in addition to ADAPTeR (see Materials and Methods). Primary tumour and lung metastases maintained a partial response to treatment until death, whereas new brain, bone and thoraco-nodal metastases emerged on treatment. To test if this differential response status observed in this patient could be linked to the population of expanded maintained TCRs that correlated with response during lifetime (see Section 4.2.5), the post-mortem repertoires were overlapped with the pool of five TCRs that were expanded both pre- and post-treatment (Figure 4.19).

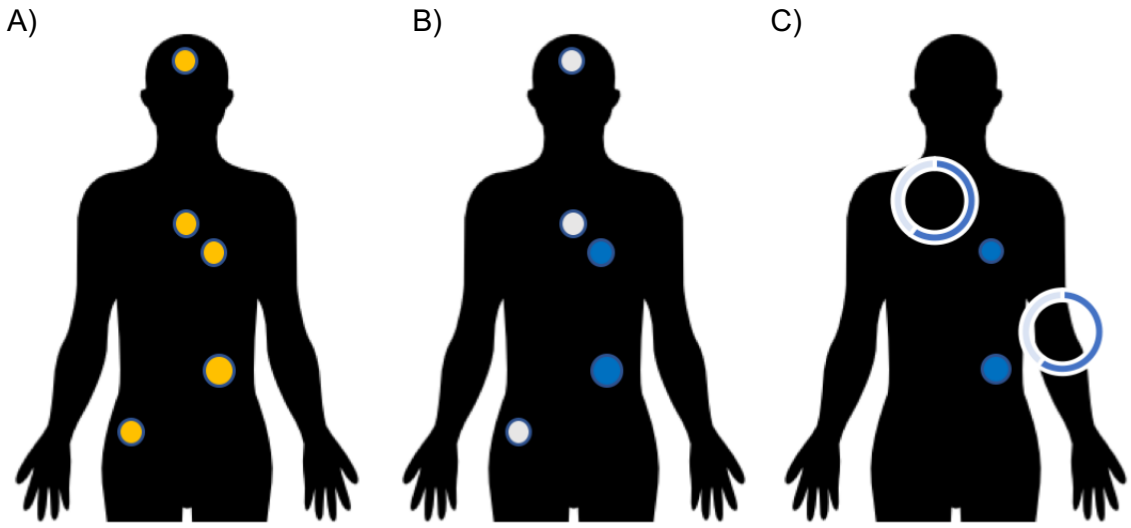

Figure 4.19 Post-mortem detection of expanded maintained TCRs for patient ADR005

(A) PEACE samples with available TCR-Seq data for patient ADR005. (B) Representation of non-progressive sites (blue circles) and progressive sites (grey circles). (C) The proportion of TCRs that were expanded both pre-treatment and post-treatment during life (n=5) detected in each post-mortem sample, only samples where the detection rate is greater than 0 are displayed. 3/5 were detected in the lung metastatic and 1/5, 1/5, 2/5 and 3/5 were detected in region 1, region 2, region 3 and region 4 of the primary site, respectively.

Strikingly, out of these five expanded maintained TCRs, three (resp. none) could be detected in the non-progressive disease sites (resp. progressive disease sites).

These persistent TCRs may be tracking antigens shared between metastatic lesions and the primary tumour and could be exerting control on the progression of these lesions upon anti-PD1 treatment. On the other hand, these TCRs might be absent from progressive lesions because the antigenic landscape evolved too much, enabling these sites to escape the immune response, even under nivolumab.

4.3 Chapter discussion

The results presented in this chapter shed light on the link between response to nivolumab and TCR repertoire distribution dynamics in metastatic renal cell carcinoma. More specifically, they suggest that the mechanism of PD-1 action may be to maintain pre-existing T cell clones, rather than to act primarily by driving T cell clonal replacement, as it has recently been claimed to be the central mechanism of anti-PD-1 treatment (Yost et al., 2019) and anti-PD-L1 treatment (T. D. Wu et al., 2020).

Firstly, high level insights into the immune infiltrate through bulk-RNA sequencing analysis revealed that nivolumab induced increased immune infiltration across the whole cohort. Interestingly, both the composition and the quantity of the infiltrate seemed to be different depending on clinical response status, with responders displaying a higher immune infiltration, both pre- and post-treatment, together with signs of specific antigen-dependent T cell activation in the TME post-treatment.

Supporting this hypothesis further, the TCR repertoire analysis revealed higher clonality values and larger amounts of CDR3 similarity networks in responders, both pre- and post-treatment, suggesting enrichment for tumour-specific T cells in these patients compared to non-responders. Longitudinal data revealed that the ability to maintain pre-existing TCRs throughout treatment associated significantly with clinical response.

A potential explanation could be that responders possess an immunogenic antigenic landscape, able to trigger and maintain cancer-specific clonotypes, whereas non-responders present a weaker antigenic signal to the immune system. Indeed, CDR3 similarity network analysis revealed that, in all patients, the expanded clonotypes

that were part of such networks were more likely to be maintained than other expanded clonotypes.

Paired single-cell RNA and single-cell TCR sequencing on two patients also supported this hypothesis, showing that TCRs within post-treatment CDR3 similarity networks corresponded to T-cells that had bound to the drug. This observation suggested that nivolumab benefits T cells expressing TCRs with features of antigen exposure, that have most likely expanded in the vicinity of a tumour providing a strong antigenic signal.

This work focused mostly on pre-treatment expanded clones that are maintained throughout treatment because of their association with clinical response in the ADAPTeR cohort. Clonal replacement was also observed, as reported in other studies, but did not associate with response.

Taken together, these data suggest that the maintenance of expanded TCRs throughout treatment and up until death might be a measurable consequence of a more important underlying biological feature: the quality and the strength of the antigenic signal driving the T cell response. Paradoxically, the nature of this signal in RCC remains elusive and inconsistent across studies (Abou Alaiwi et al., 2020; Braun et al., 2020; Braun et al., 2019; Litchfield et al., 2020; McDermott et al., 2018; Miao et al., 2018; Motzer, Banchereau, et al., 2020; Motzer, Robbins, et al., 2020; Rooney et al., 2015; Turajlic et al., 2017), compared to NSCLC where NSMs have been characterised as a rich antigenic source, as described in the previous chapter.

There are several limitations to this study. The results obtained in this chapter are based on a small cohort and need to be confirmed in larger cohorts. In addition, it would be interesting to determine if they translate to other anti-PD1 treatments and even other checkpoint blockade inhibitors in RCC and other cancers.

Also, most observations presented in this chapter solely rely on bioinformatic analysis and require further functional validation. In particular, in the context of the poorly characterised antigenic landscape of RCC, the hypothetical tumour specificity of expanded clones needs to be addressed by reactivity assays between tumour

material and autologous T cell populations. Functional characterisation of the tumour specificity of the maintained or replaced TCRs as defined in this chapter would be highly informative. This could further be tested in the context of checkpoint blockade therapy in patient derived xenograft models.

Chapter 5. Identification and characterisation of cytotoxic T cells in B cell acute lymphoblastic leukaemia

5.1 Introduction

B cell acute lymphoblastic leukaemia (B-ALL) is the most common type of childhood cancer and although standard of care for these patients achieves survival rates over 80%, treatment is long and toxic, and 15-20% of patients eventually relapse, often with more resistant disease (Hunger & Mullighan, 2015).

Chimeric antigen receptor-modified (CAR) T cells and other recent adoptive cell therapies, relying on engineered T cells that recognise cell-surface antigens such as CD19, are opening a vast range of promising new treatments but are expensive and relapse still occurs in the context of these therapies. Leveraging endogenous T cell populations could provide alternative treatment options.

Cancer specific T cells endogenously present in B-ALL patients' bone marrow or PBMC have been detected in isolation (Weber et al., 2013; Zamora et al., 2019) but their behaviour and characterisation inside the host remains undescribed.

Repertoire sequencing data are generated routinely in the clinic for the diagnosis and follow-up of B-ALL patients. Both the MRD laboratory at the Great Ormond Street Hospital for Children (GOSH) and the Bristol Genetics Laboratory at Southmead Hospital (SH) implement the widely used BIOMED-2 kit (van Dongen et al., 2003) and systematically generate repertoire sequencing at diagnosis, post-treatment and at disease relapse. This data are primarily utilised to detect the malignant leukaemic clones and to use them as barcodes to measure minimal residual disease (MRD) post-treatment (O'Connor et al., 2018).

In this chapter, I demonstrate that despite the overwhelming abundance of blast cells in diagnostic samples, clinical repertoire sequencing samples contain rich

information on the non-leukaemic infiltration in B-ALL. In particular, I firstly demonstrate that productive beta CDR3 chains, distinct from leukaemic re-arrangements, can be identified in this data. Secondly, I show that these TCRs can be detected at diagnosis, post-treatment and at relapse and display features of antigen exposure, for some patients. Lastly, in matched single-cell RNA and TCR sequencing, I identify cytotoxic phenotypes that present features of TCR engagement with leukaemic cells.

This project is a collaborative effort involving multiple research groups. Details of individual contributions to the work presented in this chapter are as follows. Sample acquisition, preparation and sequencing of BIOMED-2 were jointly done by Jack Bartram, Stuart Adams, Gary Wright, Lauren Wise, Natalie Kent and Eleanor Watt at GOSH and John Moppett, Jeremy Hancock, Stephanie Wakeman and Paul Archer at SH. Acquisition, preparation, and sequencing of 10X single-cell samples were jointly done by Maria Vila De Mucha (PhD student in the Quezada group), David O'Connor (paediatric haematologist at GOSH, clinical research training fellow in the Mansour group) and Imran Uddin (single cell genomics technician, single cell genomics facility, UCL). I performed the processing of raw data from bulk repertoire sequencing and single cell 10X data and all downstream bioinformatics analysis and data visualisation for repertoire sequencing data and 10X data (RNA-Seq and TCR-Seq data).

5.2 Results

5.2.1 Malignant clones are distinct from healthy clones in repertoire sequencing data

A cohort of 149 diagnostic, 32 post-treatment and 16 relapse partially annotated repertoire sequencing samples of bone marrow biopsies from GOSH and SH was analysed (Figure 5.1). This cohort will hereon be referred to as the BIOMED-2 cohort. In this setting, "treatment" refers to the induction phase of ALL treatment which consists of four weeks of chemotherapy and steroids. The subsequent phases of the treatment (consolidation and maintenance) are not considered in this project. The median time to relapse was three and a half years.

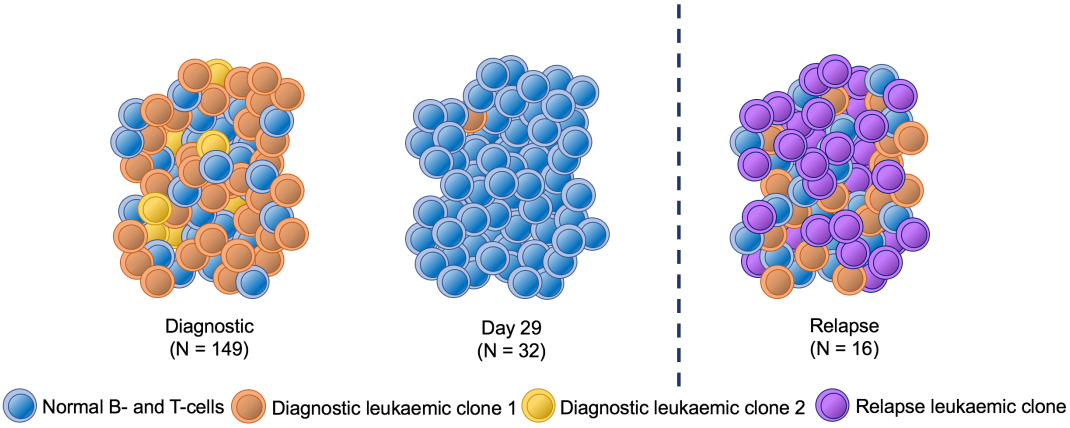

Figure 5.1 Representation of BIOMED-2 cohort

Bone marrow sampling time-point, and number (N) of samples subjected to repertoire sequencing are shown. 197 samples were evaluable: diagnostic (N=149); day 29 (post induction treatment, N=32); relapse (N=16). Seven additional bone marrow samples obtained from healthy individuals complete the cohort (not shown). N= 162 individuals

These samples were obtained through routine clinical collection and subjected to DNA based sequencing meant to monitor MRD. They were retrospectively analysed in the context of this project in order to investigate the hypothetical presence of tumour-reactive T cells in the bone marrow of childhood B-ALL patients. In addition, seven matched aged (median = six years old) healthy bone marrow samples, derived from excess marrow of donors (siblings in this case) in the context of allogeneic stem cell transplantation procedures, were included in the cohort and used as control in the analysis.

All samples were parsed through MiXCR to retrieve all BCR and TCR rearrangements, both productive ones and non-productive ones and log-log representations were screened in order to obtain a high-level visualisation of the B-ALL repertoire (Figure 5.2).

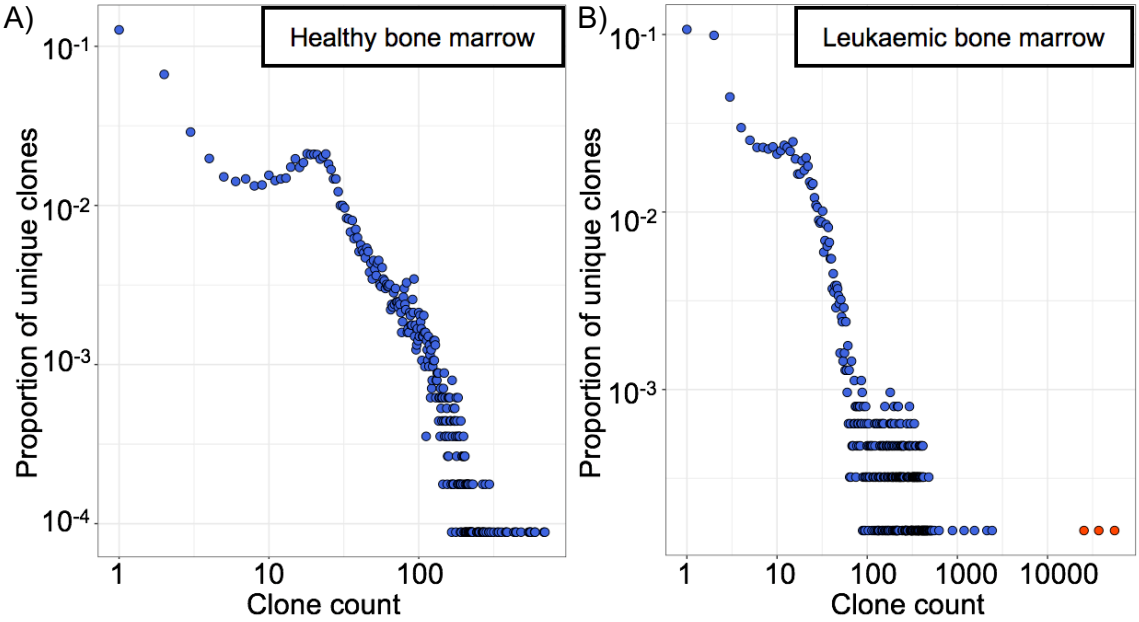

Figure 5.2 Log-log representation of healthy and leukaemic bone marrow repertoires

Representative log-log plots of bone marrow pooled BCR and TCR repertoires from healthy donor HD2 (A) and from patient PX319497 at diagnosis (B). Outlying clones are indicated by red circles (see Materials and Methods). The *x* axis represents TCR abundance (clone count), and the *y* axis represents the proportion of the repertoire (proportion of unique clones).

The shape of the bone marrow pooled BCR and TCR repertoires in diagnostic B-ALL samples was strikingly distinguishable from those of healthy bone marrows. Whereas the shape of the distribution in healthy bone marrow samples was a continuous curve as expected for log-log representation, the distribution in diagnostic B-ALL samples systematically comprised of an additional group of large clone counts, clearly separate from the main curve (red dots), hereon referred to as outlying clones.

Clinical data from these samples indicated that the median proportion of blast leukaemic cells in these samples was 93%, in line with what is described in the literature for such cohorts (Wright et al., 2019). In addition, reports on the genomic composition of this disease indicate that the clonal composition of B-ALL is quite restricted (Abdo et al., 2020), which explains why, usually, no more than five re-arrangements identified at diagnosis are registered to asses MRD post-treatment (Wright et al., 2019). In that context, the distribution revealed by log-log representation is not surprising and it can be safely assumed that outlying re-

arrangements are highly enriched for leukaemic re-arrangements whereas the rest of the distribution should be, in majority, leukaemia free.

Following this observation, a semi-automated outlier detection pipeline was run on each sample (see Materials and Methods), which was subsequently labelled as containing outlying rearrangements or not (Figure 5.3). As expected, both diagnostic and relapse samples systematically fell into the former category whereas health bone marrow samples fell into the latter. Post-treatment sample displayed a more heterogeneous profile reflecting the fact that not all patients reached satisfying MRD levels at day 29.

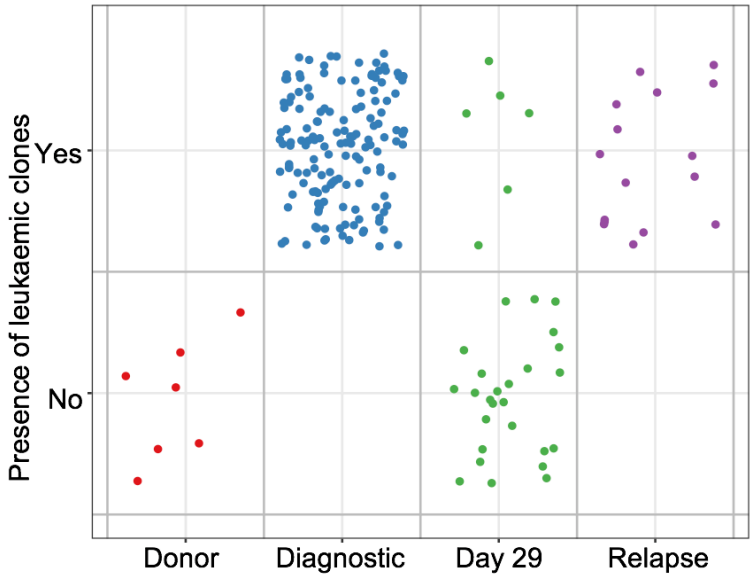

Figure 5.3 Representation of leukaemic clone detection rate

The binary leukaemic clone(s) detection status (presence/absence, see Materials and Methods) of healthy donor (red), diagnostic (blue), day 29 (green) and relapse (purple) bone marrow samples is shown for all patients.

5.2.2 Healthy beta TCR clones compose a rich and diverse repertoire at diagnosis

Next, for each sample, the non-leukaemic part of the repertoire was filtered for productive beta chain TCRs (not containing "_" or "*" in the MiXCR reconstructed amino acid sequence) starting with a Cysteine ("C") only. Alpha chain TCR and BCR rearrangements were not considered further, unless stated otherwise. Samples that contained less that 100 unique productive beta chain TCRs were excluded from

downstream analysis. The clonality score of the beta repertoire was computed for each patient, at each available time-point, together with healthy bone marrow controls (see Materials and Methods, Figure 5.4).

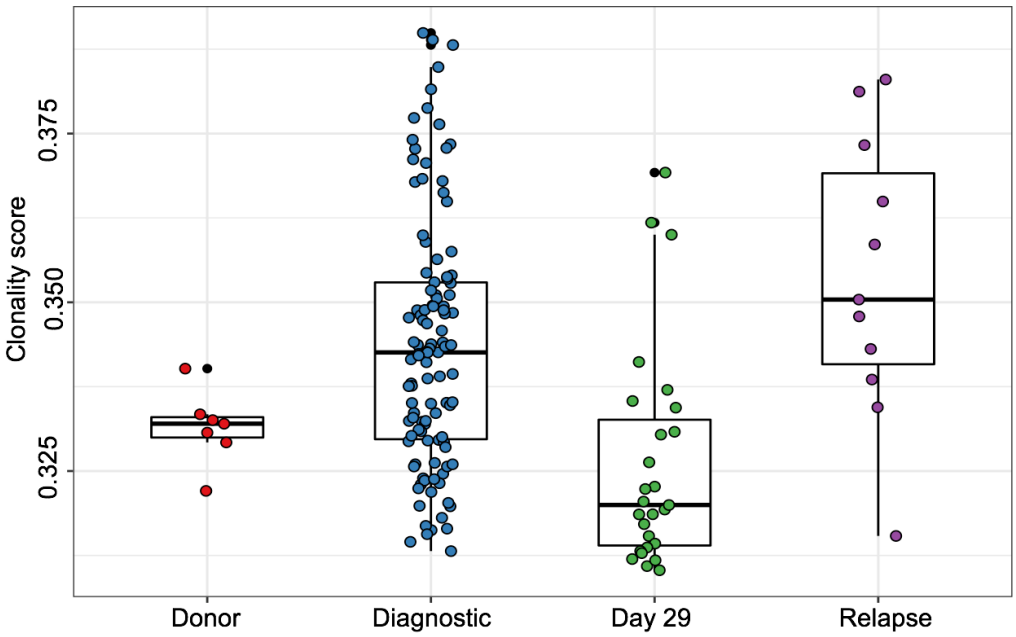

Figure 5.4 Healthy and leukaemic bone marrow TCR repertoire clonality score

The clonality score of healthy donor (red, N=7), diagnostic (blue, N=115), day 29 (green, N=28) and relapse (purple, N=14) bone marrow samples is shown (N=132 individuals).

Strikingly, the clonality scores of healthy bone marrow samples were very homogeneous with only limited variation observed between individuals. On the contrary, patient derived samples displayed a large range of clonality values at all time-points. In addition, clonality scores at diagnosis and at relapse were significantly higher than scores in patient derived day 29 post-treatment samples (P=0.00005 and P=0.0009, respectively). This data indicated higher prevalence of expanded productive beta TCRs in samples with high leukaemic content when compared with samples with low or null leukaemic content.

Similarly to the observations previously described for NSCLC and metastatic RCC in Chapter 3 and Chapter 4, the high frequency TCRs in these samples could potentially identify leukaemia specific T cells. The large variability of clonality values observed at diagnosis and at relapse could then be explained by a gradient of

immunogenicity of leukaemias across the cohort, which would be coherent with the variability of the corresponding genomic profiles in these patients.

To investigate this further, the GOSH patients were split into three main genomic categories when data was available. Patients retained for this analysis displayed either the *TEL-AML* fusion, the *MLL* fusion or a non-diploid genome (hyper ploidy, hypo ploidy or general haploidy detected). No association between genotype and clonality scores was observed (Figure 5.5), suggesting that this classification does not segregate more or less immunogenic leukaemias, in the context of the hypothetical presence of cancer specific T cells in these samples, identified by high clonality scores.

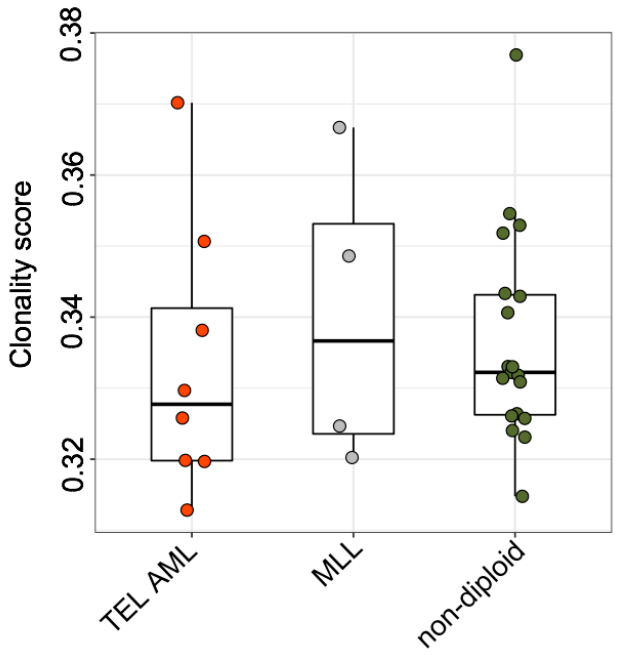

Figure 5.5 Bone marrow TCR repertoire clonality score per genotype

The clonality score of *TEL-AML* (red, N=8), *MLL* (blue, N=4) and non-diploid (green, N=19) diagnostic bone marrow samples is shown for each patient

5.2.3 Longitudinal analysis reveals repertoire sharing between diagnosis and relapse

Following this observation, the hypothetical presence of leukaemia-specific TCRs was investigated in an antigen-agnostic manner by longitudinally comparing the beta repertoires and outlying rearrangements (identified as described in Section 5.2.1)

between diagnosis and relapse. For some patients, the pool of outlying clones at diagnosis was distinct from the pool of outlying clones at relapse but for others, one or more re-arrangements were shared between time-points (Figure 5.6). This suggested that the relapse disease was driven by the same primary leukaemic clone in the latter group of patients, whilst it was most likely driven by a novel clone selected by Darwinian evolution in the former group of patients.

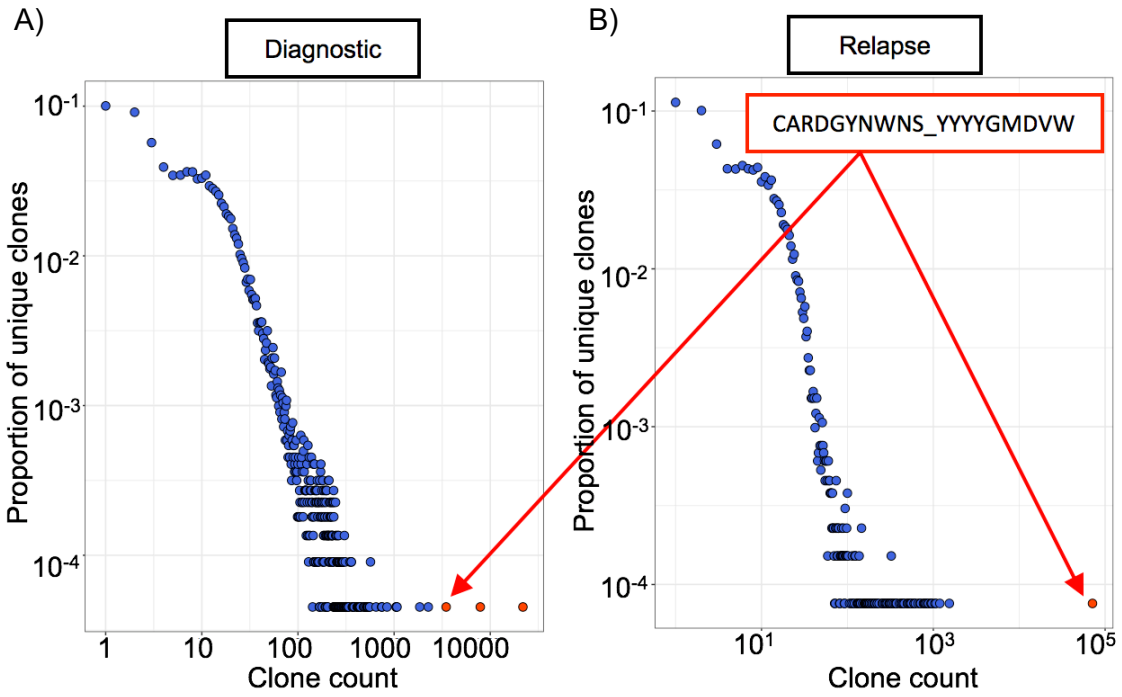

Figure 5.6 Log-log representation of matched diagnostic and relapse repertoires

Representative log-log plots of bone marrow pooled BCR and TCR repertoires from patient 883429 diagnostic (A) and relapse (B) sample. Outlying clones are indicated by red circles. The non-productive BCR re-arrangement "CARDGYNWNS_YYYYGMDVW", detected as leukaemic at both time-points, is shown. The *x* axis represents TCR abundance (clone count), and the *y* axis represents the proportion of the repertoire (proportion of unique clones).

Hypothetically, the sharing of one or more outlying clones between primary disease and relapse should translate to a longitudinally stable antigenic landscape which may consequently drive a longitudinally stable pool of leukaemia-specific TCRs. To investigate this hypothesis, primary and relapse non-leukaemic TCR repertoire were compared for the ten patients that had sufficient data at both time-points (see Materials and Methods, Figure 5.7).

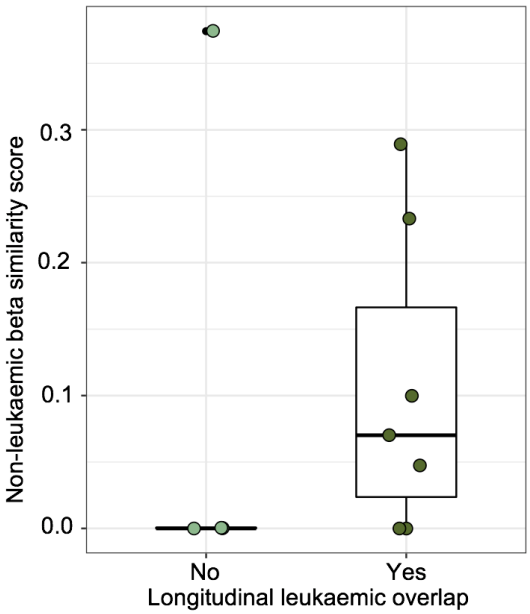

Figure 5.7 Leukaemic and non-leukaemic longitudinal similarity score

The similarity score between diagnostic and relapse non-leukaemic TCR repertoire is shown for each patient. Patients are split between leukaemic clone overlap group: Yes (resp. No) for patients with shared (resp. mutually exclusive) outlying clones between time-points.

Strikingly, four out of five patients for whom the outlying rearrangements, hypothesised as identifying leukaemia, were different between time-points displayed no overlap in the corresponding TCR repertoire. In contrast, five out of seven patients for whom at least one outlying re-arrangement at relapse that was already detected as such at diagnosis had non-zero overlap values in their corresponding TCR repertoire, which could indicate maintenance of cancer-specific TCRs.

Next, CDR3 amino acid clustering was performed on pooled diagnostic and relapse samples to test whether the longitudinal overlap could be extended to similar but not identical CDR3 chains (see Materials and Methods). Although clustering structures could not be observed for all patients, probably due to the low input size for some of them, interesting networks were found in several patients, including patient 883429 (Figure 5.8).

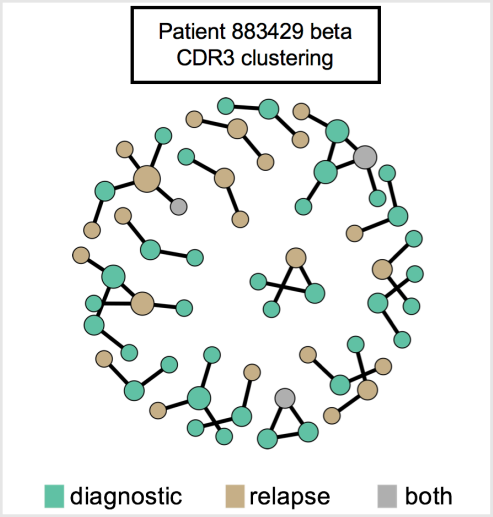

Figure 5.8 Pooled diagnostic and relapse CDR3 clustering representation

Representative network diagrams of bone marrow CDR3 sequences for patient 883429. Each circle represents a unique CDR3 found exclusively at diagnosis (green), exclusively at relapse (brown) or detected at both time points (grey). CDR3s that have a pairwise similarity of >0.82 are connected by an edge.

As previously explained in Chapter 3 and Chapter 4, such networks probably indicate groups of TCRs with identical specificity, which in this context are hypothesised to recognise leukaemia-specific antigens. In patient 883929, the relapse of identical leukaemic clones was not only associated with shared productive beta sequences between diagnosis and relapse (Figure 5.7) but also with a number of CDR3 similarity clusters, grouping together TCRs from both time-points.

Following the hypothesis formulated above, these observations could be interpreted as follows: at diagnosis, beta repertoires contain cancer specific TCRs. After intensive chemotherapy, both the normal and the malignant repertoire of the bone marrow is ablated. Some degree of memory immune repertoire is however stored and can be recruited in the instances where the leukaemia relapses, but only if the relapse is driven by the same leukaemic clone.

5.2.4 Matched single-cell RNA sequencing identifies CD8 T-cells with features of antigen experience in B-ALL samples

In order to provide more substantial evidence to the hypothesis generated by the analysis described above, single-cell RNA and TCR sequencing was performed for

three diagnostic samples from three different patients, together with a control healthy bone marrow sample. For each patient sample, T cells and leukaemic cells were separated and see remixed to a nine-to-one ratio in order to maximise the amount of sequencing dedicated to T cells whilst still keeping some amount of leukaemic cells. The aim was to investigate the hypothetical interactions between cytotoxic T cells and leukaemic cells that was expected in these samples (see Materials and Methods).

A total of 4,151 and 5,156 T cells were identified in the healthy sample and the B-ALL samples, respectively and 643 leukaemic cells. Both T cells and leukaemic cells were evenly spread between patients (mean=1,719; SD=129 for T cells and mean=214, SD=36 for leukaemic cells). Combining unsupervised clustering with expression profiles of CD19, CD3E, CD4, CD8A, FOXP3 and KLRB1 enabled to identify and separate the major cell types presents in the pooled single cell data set: $CD8^+$ T cells, $CD4^+$ T cells, $CD4^+$ T regulatory cells, innate-like T cells, gamma-delta T cells and leukaemic cells (Figure 5.9).

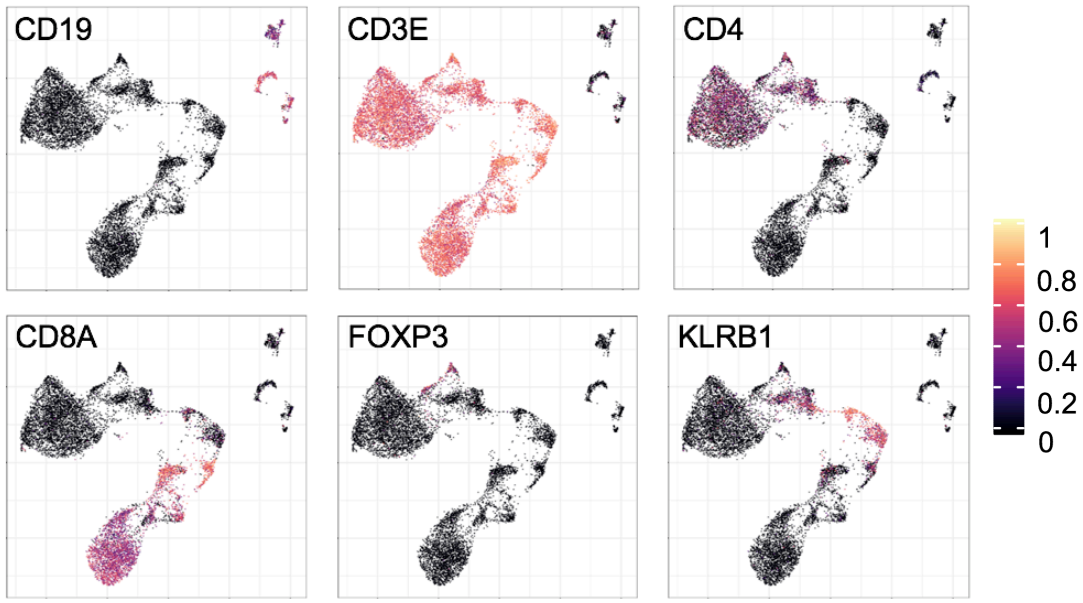

Figure 5.9 Pooled UMAP visualisation of T cells and leukaemic cells

UMAP (uniform manifold approximation and projection) visualisations of 9,438 cells from 3 diagnostic B-ALL bone marrow samples and one healthy donor sample. In each quadrant, the z-scored expression of a specific gene is shown.

Next, unsupervised clustering was performed on the T cells and the expression levels of a selection of 46 genes of interest were examined across the entire dataset which enabled to label each cluster (Figure 5.10).

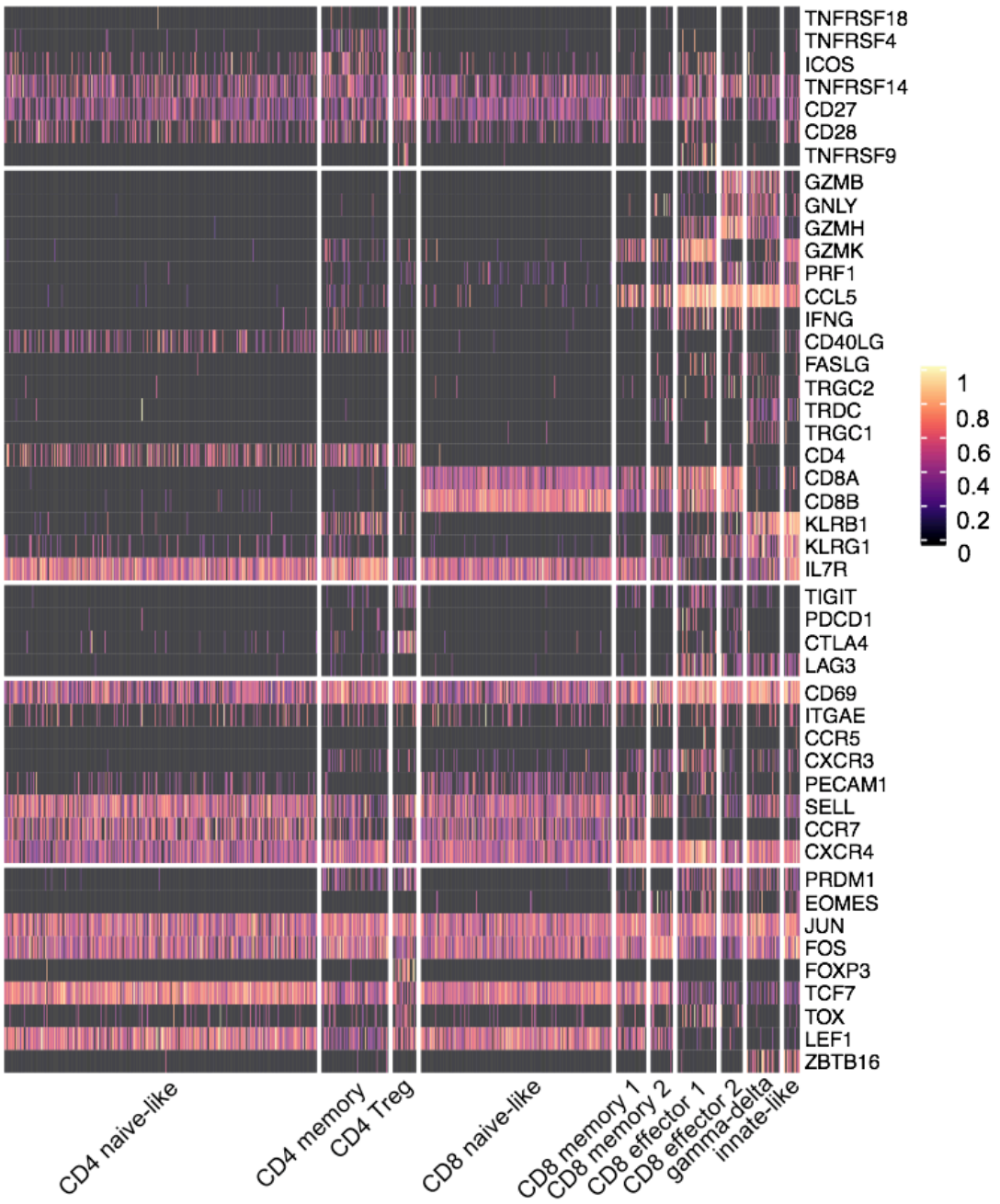

Figure 5.10 Gene expression profile of T cell clusters

Heatmap representation of z-scored expression of selected immune related genes across the different T cell clusters as defined, per cell.

Both the $CD4^+$ and the $CD8^+$ compartments contained a highly prevalent naive-like population, characterised by the highest expression levels of CCR7, LEF1 and TCF7

and a lack of expression of any gene related to effector function. The CD4 compartment also comprised of a memory-like population displaying intermediate levels of KLRB1 and BLIMP-1 (PRDM1) (Crotty, Johnston, & Schoenberger, 2010) together with high expression of CCR7, LEF1 and TCF7 and a regulatory T cell population, with a distinguishable expression of FOXP3, TIGIT and CTLA4.

The $CD8^+$ compartment was more complex and contained two memory populations (intermediate levels of CCL5, GZMK) and two effector populations (exclusive expression of IFN-γ (IFNG) and perforin (PRF1) and PD-1 (PDCD1)), in addition to the naive cluster. The two memory clusters differed by their usage of TIGIT and GZMK, CD8 memory 1 displaying intermediate levels of both genes versus no expression in CD8 memory 2. CD8 effector 2 appeared like more differentiated than CD8 effector 1, with higher expression of GNLY. In addition, CD8 effectors 1 and 2 were distinguishable by mutually exclusive usage of GZMK and GMZH, respectively.

Finally, 2 clusters of CD4-, CD8- T cells were present: one innate-like cell population, characterised by the highest expression levels of KLRB1 together with the expression of the characteristic transcription factor ZBTB16 (Leruste et al., 2019); and one gamma delta T cell population, characterised by expression of TRDC and TRGC1/2 together with the absence of αβ TCR retrieval. Cluster labelling and UMAP distribution is summarised in Figure 5.11.

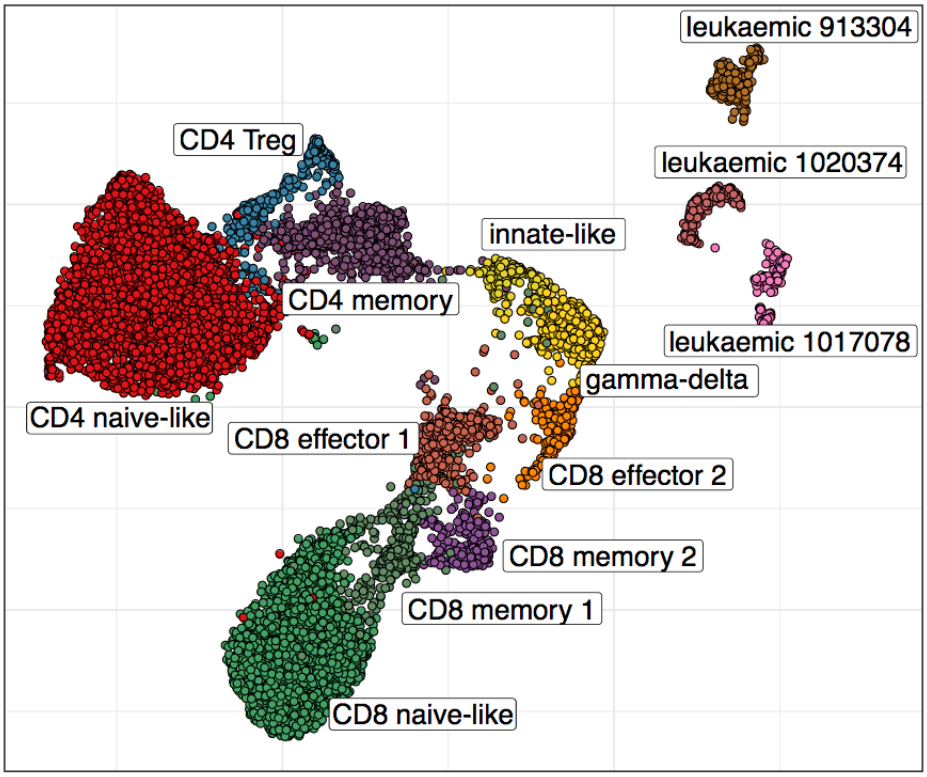

Figure 5.11 Pooled UMAP visualisation of T cells and leukaemic cells clusters

UMAP visualisation of 9,438 cells from three diagnostic B-ALL bone marrow samples and one healthy donor sample, and their repartition among the different identified clusters.

5.2.5 Cytotoxic CD8 T-cells are highly enriched in leukaemic bone marrows when compared to health control

Next, the pooled Seurat object of T cells was split between the three B-ALL patients and the healthy bone marrow sample to determine whether the distribution of the single-cell clusters identified above was different between T cells from leukaemic bone marrow and T cells from healthy bone marrow (Figure 5.12).

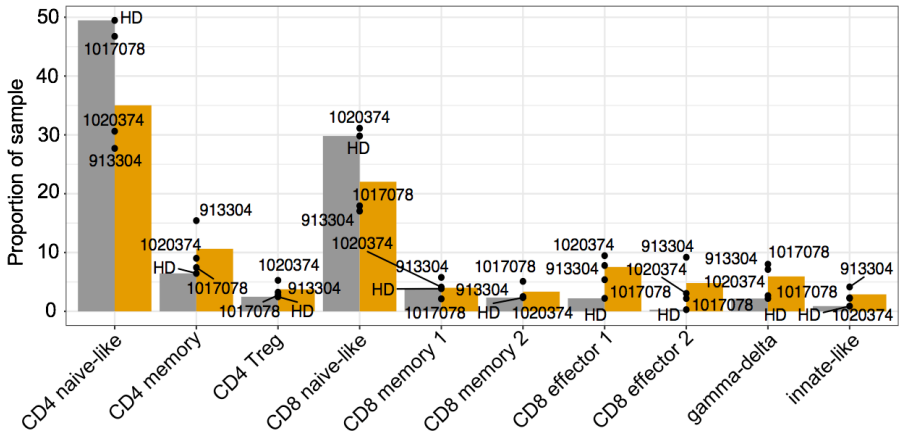

Figure 5.12 Prevalence of T cell clusters in B-ALL and healthy control samples

Bars displaying the relative cellular prevalence of T cell clusters in each individual sample, split between B-ALL patients (yellow) and healthy control (grey). Individual values are shown.

Naive-like CD4 and CD8 T cells were the most abundant subtypes throughout the data set, together accounting for more than 50% of all single cells, but there were even more abundant in the healthy bone marrow sample, where 80% of single cells were naive-like. Conversely, although much less abundant overall, the CD4 memory, CD8 effector 1&2, gamma-delta and innate-like clusters represented a systematically higher proportion of B-ALL samples than the healthy sample. Remarkably, the CD8 effector 2 cluster, which was identified as the most cytotoxic CD8 subtype (expression of GZMB, PRF1 and GNLY), was virtually absent from the healthy bone marrow and only found in leukaemic samples. No clear difference was observed in the rest of the clusters.

These observations are limited by the very small number of samples in this cohort and no statistical significance could be reached. To address this, a publically available data set of B-ALL and T-ALL bone marrow 10X sequencing data with healthy controls (Caron et al., 2020) was integrated to the Seurat object, hereon referred to as GSE132509. A total of three B-ALL samples, one T-ALL sample and three healthy bone marrow samples from GSE132509 was added to the analysis. UMAP coordinates and cluster annotations were obtained for this validation dataset via integration with the primary data (Figure 5.13, see Materials and Methods).

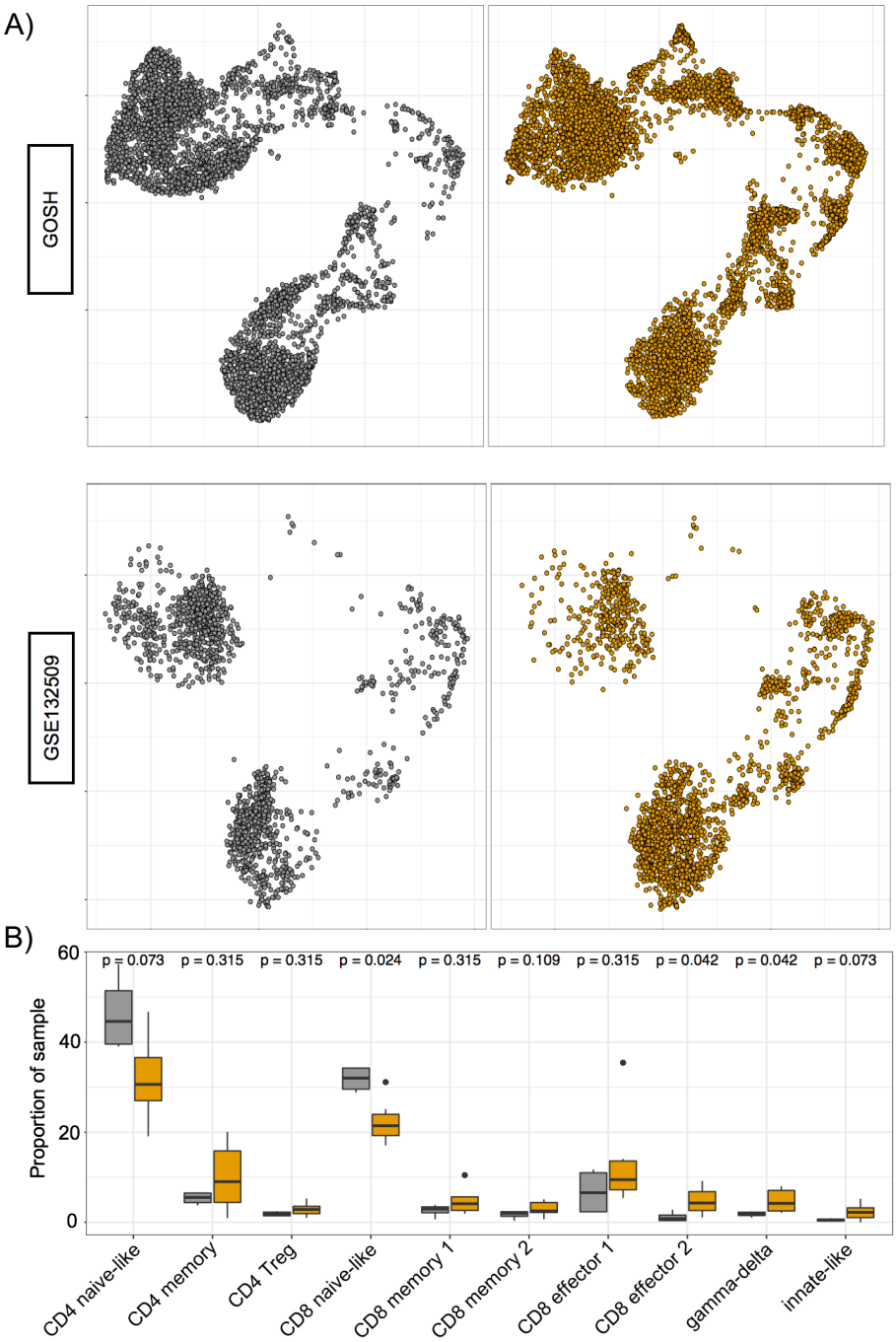

Figure 5.13 Prevalence of T cell clusters in primary and validation data sets

(A) UMAP visualisation of B-ALL bone marrow samples (right panels) and healthy donor samples (left panels) in the primary data set (top panels) and the validation data set (bottom panels). (B) Boxplots summarising the relative cellular prevalence of each T cell clusters in B-ALL patients (yellow) and healthy controls (grey). The two-sided Mann–Whitney test *P* values are shown.

The differential distribution of leukaemic samples versus healthy samples was remarkably stable across both data sets which enabled some of the observations made above to reach statistical significance when tested in the combined data set of

seven leukaemic samples and four healthy samples. More precisely, the CD8 effector 2 and the gamma-delta cluster (resp. the CD8 naive-like cluster) were significantly more (resp. less) abundant in the leukaemic samples when compared to healthy samples.

5.2.6 Single cell TCR-Seq reveals clonal expansion in cytotoxic T cell phenotypes

Next, for the original data set of three B-ALL samples and one healthy control, single beta CDR3 chains extracted from the matched VDJ kit sequencing data and were mapped to the transcriptomic data (see Materials and Methods). Individual CDR3 sequences were labelled as single (detected in only one cell) or expanded (detected in at least two cells, Figure 5.14).

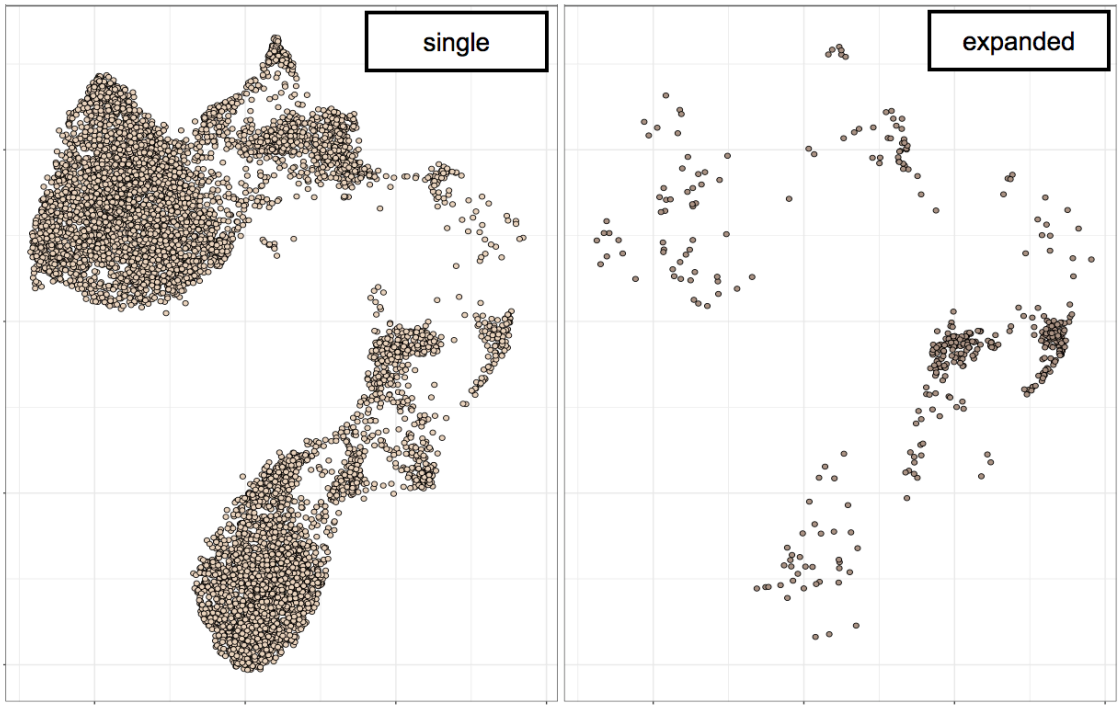

Figure 5.14 Pooled UMAP visualisation of singlet and doublet beta CDR3 chains

UMAP visualisation of 5,715 cells from 3 diagnostic B-ALL bone marrow samples and 4,006 cells from one healthy donor sample, split by TCR status. Cells were labelled as singlets if their TCR was seen only once in the data set and doublets if shared by at least one other cell.

Firstly, beta CDR3 chains were robustly identified in every cluster apart from the innate-like cluster and the gamma-delta cluster where the paucity of TCRs was in

line with the absence of CD8A or CD4 expression. Secondly, contrary to single TCRs which were evenly distributed across $CD4^+$ and $CD8^+$ clusters, expanded TCRs displayed a remarkably constrained profile and were almost exclusively detected in the two CD8 effector clusters. When looking at the distribution of expanded TCRs per sample, most of them were restricted to patient 913304, indicating either a sampling bias or a particularly strong clonal expansion occurring for this patient (Figure 5.15).

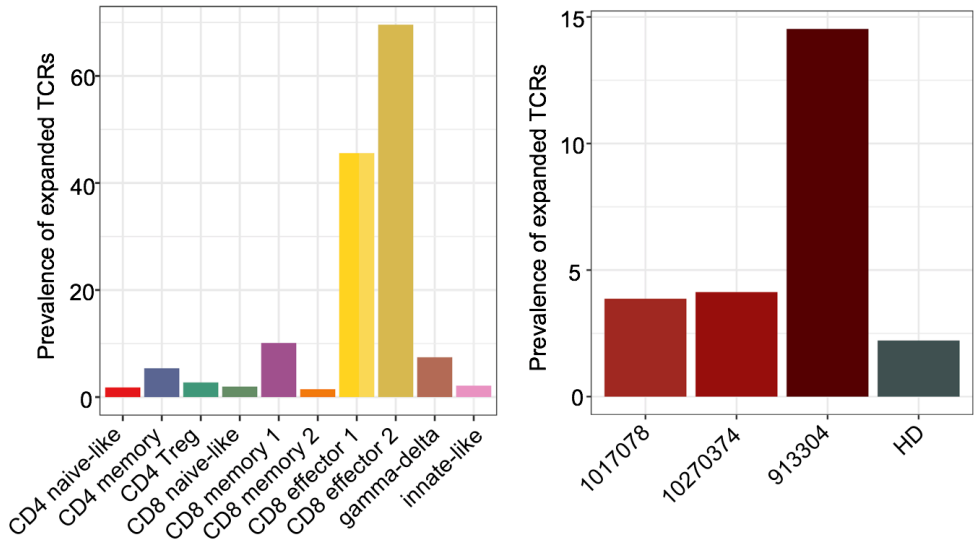

Figure 5.15 Prevalence of doublet CDR3 chains

Bars displaying the relative cellular prevalence of CDR3 doublets in each T cell clusters (left panel) and in each individual sample (right panel).

5.2.7 Engagement of cytotoxic T cells with leukaemic cells is supported by cell-cell interaction analysis

Together with a transcriptional profile displaying many features of cytotoxicity, the fact that both $CD8^+$ effector clusters contained a large proportion of expanded TCRs supported the hypothesis that these clusters contained T cells recognising leukaemic cells and engaged in cytotoxic action against those. To further investigate this model, co-expression of ligand-receptor pairs were computed between patient specific leukaemic clusters and each cluster of T cells using the cellphoneDB (Efremova et al., 2020) software (Figure 5.16, see Materials and Methods).

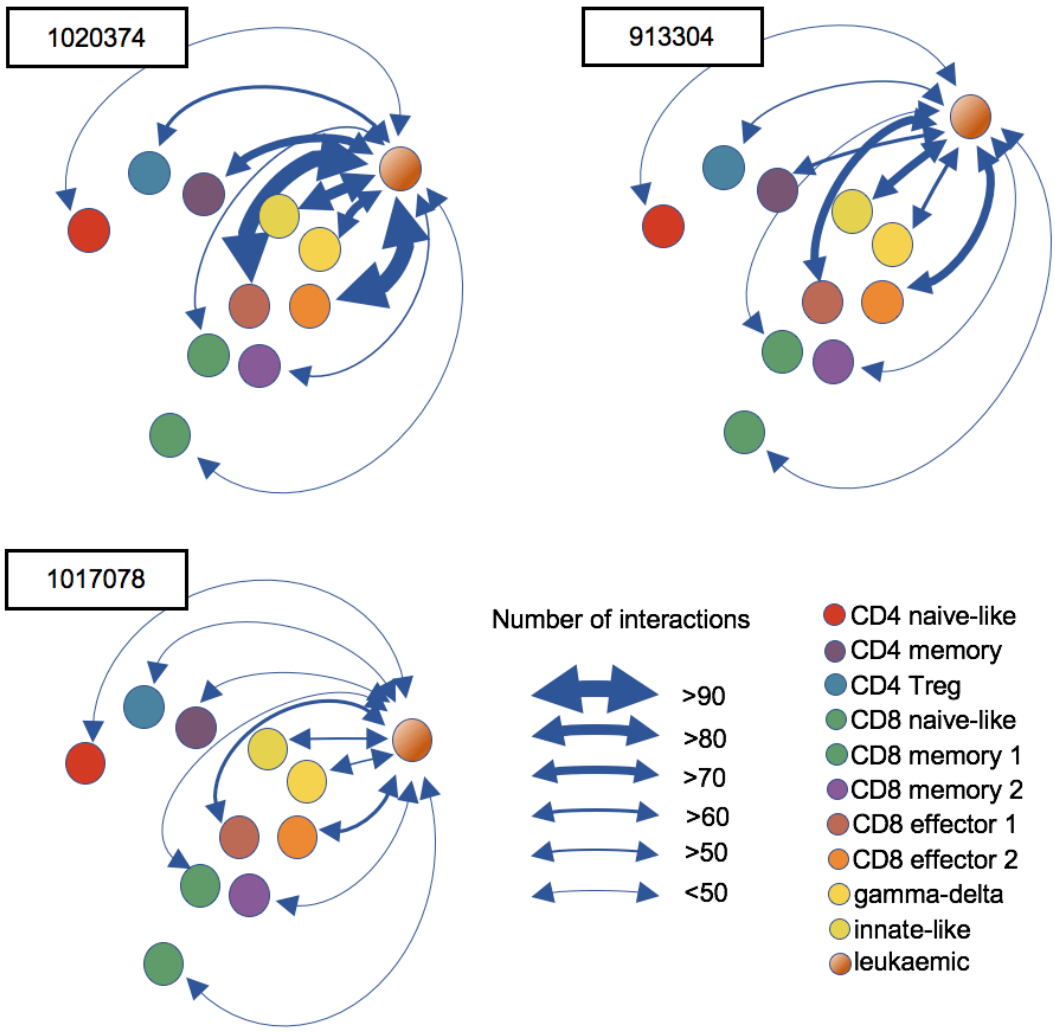

Figure 5.16 Interaction between T cell clusters and leukaemic cell clusters

Network diagram of ligand–receptor interactions for each pair of pooled T cell clusters and leukaemic clusters as determined by cellphoneDB (see Materials and Methods); Solid lines represent pathways between two populations, the width of each line is proportional to the number of pathways. Each quadrant represents one individual patient.

The number of interactions between T cell clusters and leukaemic clusters varied between patients, with a total of 675, 570 and 464 interactions detected for patient 1020374, 913304 and 1017078 respectively. This could perhaps indicate a gradient of immunogenicity of primary leukaemias, which was hypothesised in Section 5.2.2 following the observation that clonality scores were highly variable between patients at the time of diagnosis.

In addition, when split per T cell cluster, the distribution of the number of interactions was highly conserved across patients with the CD8 effector populations

systematically generating the greater number of interactions and the two naive populations, the smaller number (Table 5.1).

1020374	
CD4 naive-like	44
CD4 memory	75
CD4 Treg	66
CD8 naive-like	42
CD8 memory 1	56
CD8 memory 2	53
CD8 effector 1	90
CD8 effector 2	91
gamma-delta	76
innate-like	82

913304	
CD4 naive-like	39
CD4 memory	62
CD4 Treg	55
CD8 naive-like	38
CD8 memory 1	48
CD8 memory 2	45
CD8 effector 1	74
CD8 effector 2	73
gamma-delta	63
innate-like	73

1017078	
CD4 naive-like	31
CD4 memory	46
CD4 Treg	45
CD8 naive-like	31
CD8 memory 1	38
CD8 memory 2	37
CD8 effector 1	64
CD8 effector 2	62
gamma-delta	58
innate-like	52

Table 5.1 Number of interactions between T cell clusters and leukaemic cell clusters

The number of interactions between T cell clusters and leukaemic cell clusters, split per patient. Each row displays the number of interactions between the corresponding T cell cluster and the leukaemic cells from the corresponding patient.

Together with the observation that CD8 effector clusters display a phenotype previously reported as tumour-specific and that they accounted for the vast majority of expanded clones (see Section 5.2.6), this data suggests that these populations are actively engaged with leukaemic cells.

5.3 Chapter discussion

The results presented in this chapter reveal that the non-leukaemic TCR repertoire of the bone marrow in childhood B-ALL patients contains rich information on the potential endogenous specificities of T cells against cancer.

Firstly, retrospective analysis of clinical repertoire sequencing samples enabled to define the bone marrow beta TCR repertoire of B-ALL patients at diagnosis, post induction treatment and at relapse. Straightforward metrics demonstrated that the

TCR repertoire of B-ALL patients was significantly different from a control cohort of matched aged healthy individuals.

The latter group displayed a very homogeneous distribution profile with limited variation of the clonality score between individuals. In B-ALL patients however, clonality scores were higher overall and highly variable across the cohort and time-points. In addition, the re-appearance of the primary leukaemic clone at relapse was associated with the re-appearance of the same beta TCRs, suggesting the TCR repertoire could be stable in time when the antigenic landscape of the corresponding leukaemia was sufficiently stable.

Following the hypothesis that this signal in the TCR repertoire data was driven by antigenic signal, single-cell RNA-Seq and TCR-Seq of T-cells and leukaemic cells was performed. In this subsequent dataset, the presence of expanded T cell clones in specific subsets of effector CD8 T cells was clear and significantly enriched in B-ALL patients when compared to health controls. In addition, cell-cell interactions were detected between these subsets of effector CD8 T cells and leukaemic cells, suggesting that they were participating in an active, endogenous and specific anti-tumour immune response.

Taken together, these data support the idea that the bone marrow of B-ALL patients contains cancer-specific T-cells and additionally demonstrate that cytotoxic activity of T cells is endogenously occurring in situ. Together with the recent demonstration that neoantigen-reactive T cells can be isolated from the bone marrow of B-ALL patients (Zamora et al., 2019), this analysis supports the hypothesis that actionable cancer-specific T cells are present in the TME.

This study has several limitations. Firstly, similarly to what was discussed at the end of Chapter 3 and Chapter 4, functional work is required to fully prove the specificity and activity of the T-cells identified. Such work could include reactivity assays between tumour material and autologous T cell populations or transduction of TCRs hypothesised as-cancer specific into T cells and subsequent in vitro or in vivo (PDX models) cancer challenge.

Secondly, the current view is very limited by a T-cell exclusive description and further analysis of matched scRNA-Seq or flow cytometry of the rest of the non-leukaemic cell populations that interact with T-cells is needed. In particular, matched analysis of the myeloid compartment will help to understand what limits the cytotoxic activity of endogenous cytotoxic T-cells in B-ALL.

Lastly, further characterisation of TCR repertoire of B-ALL patients in the context of the current novel immunotherapy approaches, which are less destructive than the treatment received by the patients included in this analysis, would be key to understand how to therapeutically leverage the endogenous specificities identified here.

Chapter 6. Discussion and future directions

6.1 Introduction

Although the number of published cancer research studies generating TCR-Seq data is growing exponentially (Figure 6.1), we are still at the dawn of optimal usage and interpretation of these data. More generally, there is a consensus around the fact that next generation sequencing in cancer research has reached a critical point where, beyond the analysis of single-omics data analysis, it is the integration of multi-omics data that will yield novel insights (Chakraborty, Hosen, Ahmed, & Shekhar, 2018).

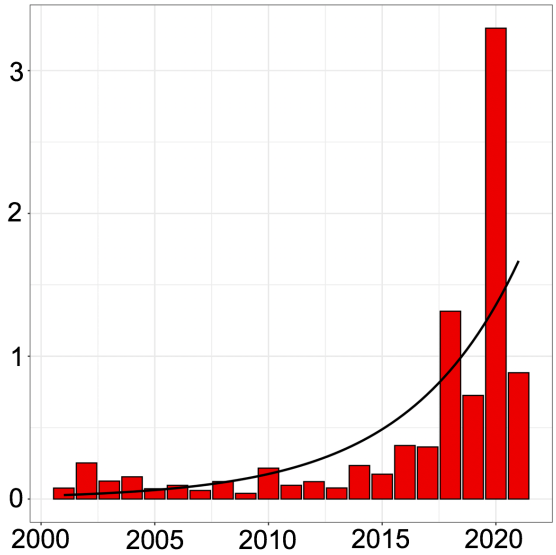

Figure 6.1 Normalised number of TCR cancer research publication since 2000

The normalised number of TCR cancer research publications per year. The y axis is showing the number of Google scholar results returned by the query #[allintitle: t cell receptor cancer] divided the number of Google scholar results returned by the query #[allintitle: cancer], multiplied by 1000. Each bar corresponds to a rolling yearly window for the Google scholar search, starting at [2000-2001]. These data were retrieved on the 12th of April 2021. The R function *geom_smooth* was used to compute and plot the logistic regression curve.

This thesis shows how TCR-Seq driven –omics data integration analysis can be performed in a range of cancer types and technological contexts to extract key features of the intratumoural TCR repertoire that are consistent with antigen

exposure, the cornerstone of the anti-tumour cytotoxic action of T cells (see Section 1.2.3).

More precisely, the fundamental immuno-oncology research questions that this thesis raises are the following (see Section 1.6):

1. Can cancer-specific TCRs be identified and distinguished from non-cancer-specific TCRs?
2. Can the TCR be used to link T cell specificity to T cell phenotype?
3. What features of the intratumoural TCR repertoire relate to clinical features of cancer?

In lung cancer, non-synonymous mutations are a well described source of tumour specificity, giving rise to tumour-specific T cells in the tumour and in the periphery, which have been identified and characterised in multiple instances. As well as providing further evidence for the existence of neoantigen-specific T cells via the analysis presented in Chapter 3, TCR-Seq data can now provide insights that go beyond this point. The analyses performed in the context of the multi-region, multi-omics TRACERx study highlighted the complex yet structured spatial heterogeneity of the TCR repertoire in NSCLC.

In renal cancer, in contrast to lung, the sources of antigen for the tumour-specific T cell response are still mainly unknown. The study presented in Chapter 4 did not benefit from the same spatial resolution or the same statistical power that the lung cancer cohort had. Despite these limitations, the renal cohort offered a unique longitudinal sampling strategy to investigate the causes and/or the effects of response to anti-PD1. Taken together, this means that the approach to the analysis of TCR-Seq data in this cohort had to be adapted. In this instance, the analysis led to novel insights into the relative importance of replacement and maintenance of tumour-specific T cells in mediating the therapeutic action of anti-PD-1 antibodies.

Finally, in childhood B-ALL, repertoire sequencing data is generated routinely in the clinic but, paradoxically, little is known about the cancer specific T cells in the bone marrow. Indeed, repertoire data is acquired in order to identify the malignant clones

and to longitudinally track their persistence (MRD>0) or disappearance (MRD=0). Previous studies demonstrated that specific T cells could be sorted from the bone marrow or the blood of B-ALL patients, suggesting that a larger cancer-specific T cell population exists and could be characterised by dedicated TCR-Seq analysis. Following on from the work presented in Chapter 3 and Chapter 4, the idea that such childhood B-ALL clinical cohorts of repertoire sequencing could be re-analysed emerged and was explored in Chapter 5. The results suggested that a cancer-specific T cell population might be identifiable through TCR-Seq and primed a dedicated single-cell sequencing experiment to characterise this population.

6.2 Current limitations and future perspectives

6.2.1 The intratumoural TCR repertoire in lung cancer

As mentioned above, the granularity achieved in the lung cancer cohort is very advanced but this study will still benefit from current and future work carried out in this cohort and/or separate cohorts. Firstly, despite indirect evidence, the specificity of ubiquitous and regional TCRs is not formally demonstrated. Single-cell TCR-Seq is currently being carried out by the Chain group in order to retrieve the α or the β chain matching the corresponding β or α chain identified as expanded ubiquitous or regional. Doing so will enable to reconstruct full αβ TCRs hypothesised as specific to ubiquitous or regional mutations as described in Chapter 3. The patients selected for this single-cell experiment are also enrolled in a patient-derived xenograft model study which means that, when αβ TCRs of interest are identified, they could be engineered into T cells and the T cells subsequently injected into these animals to study their ability to recognise and control tumour. This elaborate set of experiments could enable to fully validate the hypothesis put forward in Chapter 3.

In addition, the main caveat of this project is that the multi-region sampling setting that TRACERx offers is not translated into the clinical setting is most instances. Although it is hard to visualise how the ubiquitous/reginal dichotomy could be reproduced outside this setting, future work leveraging the overlap with the blood compartment or maybe more advanced machine learning approaches could assess how much information can be inferred from limited single sample data.

6.2.2 The intratumoural TCR repertoire in renal cancer

The findings presented in Chapter 4 rely on a small number of patients and despite reaching statistical significance for some of them, would highly benefit from being repeated in larger cohorts. In particular, a small number of patients are outlying the main distribution for certain metrics and it would be important to determine if they represent a larger group of ccRCC patients that has distinctive features or if they are simple outliers that would not scale up in numbers in a larger cohort.

The second main limitation of this project is the inability to identify the antigenic sources driving the TCR expansions observed. Although this caveat is not specific to this study but applies broadly to renal studies, it remains a very important piece of the puzzle that Chapter 4 is attempting to put together. Recent data suggest that our failure to identify cancer-specific antigens in renal cancer is mostly due to technological limitations as some putative antigen sources are today easier to characterise than others. For instance, although this could not be repeated in this study, insertions and deletions have been identified as a potential greater source of immunogenicity that NSMs in renal cancer (Turajlic et al., 2017). Future work carried out by groups that aim to describe novel alternative sources of cancer-specific antigens, such as the Litchfield group at UCL, is needed. This will help understand what determines maintenance or replacement of expanded clones upon nivolumab treatment and potentially inform better predictors of response.

6.2.3 The intratumoural TCR repertoire in childhood leukaemia

The childhood leukaemia study presented in Chapter 5 has two main components: a large cohort of relatively poor-quality bulk data and a small cohort of high quality single-cell data.

Two axes could be considered to validate the findings in the first cohort. Firstly, $CD3^+$ cells could be purified from these samples and submitted to dedicated TCR-Seq. That way, the large fraction of repertoire data occupied by malignant clones in the primary samples would be available for a deeper description of the non-malignant TCR repertoire. However, given the limited availability of these clinical samples for re-sequencing, this option is not realistic. Another axis would be to pursue the re-

purposing approach and leverage the large amount of bulk RNA-Seq publically available for childhood B-ALL patients, such as the St Jude Cloud (McLeod et al., 2021). Tools such as TRUST (B. Li et al., 2017) version 4 claim to be able to reconstruct TCRs from bulk RNA-Seq data sets and could be applied to generate additional low quality but large cohorts of repertoire data of childhood B-ALL patients' bone marrow samples.

Regarding the second part of the study, focusing on a small number of patients, it is absolutely critical that further experiments are conducted in order to push the model forward. In addition to repeating the observations made in a larger cohort, expanding these findings to the larger immune population, including the myeloid compartment, would be insightful. Also, functional work providing further evidence of the hypothetical tumour specificity of the TCRs identified in this data set are needed. To this end, the three patients selected for single-cell RNA-Seq were also selected for a tumour reactivity assay currently being conducted between the Chain group and the Quezada group. In addition, those same patients were additionally selected for the patient derived xenograft program, which could be leveraged to study this hypothesis further, in vivo.

6.3 Summary

Taken together, the work presented in this thesis demonstrates that TCR-Seq data analysis can be used to identify and characterise features of the intratumoural TCR repertoire associated with antigen exposure. These features are consistent with antigen exposure but are not formally linked to tumour specificity. Instead, they are used in an antigen-agnostic manner throughout the thesis to characterise the potentially reactive intratumoural TCR repertoire and how it relates to particular T cell phenotypes and clinical features.

It is important to note that the analyses performed do not distinguish tumour-specific TCRs from non-tumour-specific ones on the basis of individual TCR characteristics. On the contrary, in most instances, the relevant metrics emerge from a consideration of the global TCR repertoire at a given time-point. The identification of tumour-reactive TCRs in an antigen-agnostic manner is considered a strength of this

approach, since the antigen landscape in most cancers is complex, dynamic and often poorly described.

As the number of data sets grows (Figure 6.1), it seems critical that standardised and unified gold-standard analytical tools emerge, suitable to the wider community of non-computational cancer immunologists. I hope that this thesis, which puts forward a model of best practice for the identification of features associated with antigen exposure in tumour-derived TCR-Seq data, will benefit the community and contribute to such approaches being developed and widely applied.

Reference List

Abdo, C., Thonier, F., Simonin, M., Kaltenbach, S., Valduga, J., Petit, A., . . . Macintyre, E. (2020). Caution encouraged in next-generation sequencing immunogenetic analyses in acute lymphoblastic leukemia. *Blood, 136*(9), 1105-1107. doi:10.1182/blood.2020005613

Abou Alaiwi, S., Nassar, A. H., Xie, W., Bakouny, Z., Berchuck, J. E., Braun, D. A., . . . Choueiri, T. K. (2020). Mammalian SWI/SNF Complex Genomic Alterations and Immune Checkpoint Blockade in Solid Tumors. *Cancer Immunol Res, 8*(8), 1075-1084. doi:10.1158/2326-6066.CIR-19-0866

Ademokun, A., Wu, Y. C., Martin, V., Mitra, R., Sack, U., Baxendale, H., . . . Dunn-Walters, D. K. (2011). Vaccination-induced changes in human B-cell repertoire and pneumococcal IgM and IgA antibody at different ages. *Aging Cell, 10*(6), 922-930. doi:10.1111/j.1474-9726.2011.00732.x

Aguirre-Ghiso, J. A. (2007). Models, mechanisms and clinical evidence for cancer dormancy. *Nat Rev Cancer, 7*(11), 834-846. doi:10.1038/nrc2256

Algarra, I., Cabrera, T., & Garrido, F. (2000). The HLA crossroad in tumor immunology. *Hum Immunol, 61*(1), 65-73. doi:10.1016/s0198-8859(99)00156-1

Allam, A., & Kabelitz, D. (2006). TCR trans-rearrangements: biological significance in antigen recognition vs the role as lymphoma biomarker. *J Immunol, 176*(10), 5707-5712. doi:10.4049/jimmunol.176.10.5707

Alt, F. W., Rathbun, G., Oltz, E., Taccioli, G., & Shinkai, Y. (1992). Function and control of recombination-activating gene activity. *Ann N Y Acad Sci, 651*, 277-294. doi:10.1111/j.1749-6632.1992.tb24626.x

Amaria, R. N., Reddy, S. M., Tawbi, H. A., Davies, M. A., Ross, M. I., Glitza, I. C., . . . Wargo, J. A. (2018). Neoadjuvant immune checkpoint blockade in high-risk resectable melanoma. *Nat Med, 24*(11), 1649-1654. doi:10.1038/s41591-018-0197-1

Anderson, G., Partington, K. M., & Jenkinson, E. J. (1998). Differential effects of peptide diversity and stromal cell type in positive and negative selection in the thymus. *J Immunol, 161*(12), 6599-6603. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/9862687

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., . . . Sherlock, G. (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet, 25*(1), 25-29. doi:10.1038/75556

Ashton-Rickardt, P. G., Van Kaer, L., Schumacher, T. N., Ploegh, H. L., & Tonegawa, S. (1993). Peptide contributes to the specificity of positive selection of CD8+ T cells in the thymus. *Cell, 73*(5), 1041-1049. doi:10.1016/0092-8674(93)90281-t

Attaf, M., Huseby, E., & Sewell, A. K. (2015). αβ T cell receptors as predictors of health and disease. *Cell Mol Immunol, 12*(4), 391-399. doi:10.1038/cmi.2014.134

Bagaev, D. V., Vroomans, R. M. A., Samir, J., Stervbo, U., Rius, C., Dolton, G., . . . Shugay, M. (2020). VDJdb in 2019: database extension, new analysis infrastructure and a T-cell receptor motif compendium. *Nucleic Acids Res, 48*(D1), D1057-D1062. doi:10.1093/nar/gkz874

Bagaev, D. V., Zvyagin, I. V., Putintseva, E. V., Izraelson, M., Britanova, O. V., Chudakov, D. M., & Shugay, M. (2016). VDJviz: a versatile browser for immunogenomics data. *BMC Genomics, 17*(1), 453. doi:10.1186/s12864-016-2799-7

Bartram, J., Mountjoy, E., Brooks, T., Hancock, J., Williamson, H., Wright, G., . . . Hubank, M. (2016). Accurate Sample Assignment in a Multiplexed,

Ultrasensitive, High-Throughput Sequencing Assay for Minimal Residual Disease. *J Mol Diagn, 18*(4), 494-506. doi:10.1016/j.jmoldx.2016.02.008

Bengsch, B., Ohtani, T., Khan, O., Setty, M., Manne, S., O'Brien, S., . . . Wherry, E. J. (2018). Epigenomic-Guided Mass Cytometry Profiling Reveals Disease-Specific Features of Exhausted CD8 T Cells. *Immunity, 48*(5), 1029-1045 e1025. doi:10.1016/j.immuni.2018.04.026

Bengsch, B., Seigel, B., Ruhl, M., Timm, J., Kuntz, M., Blum, H. E., . . . Thimme, R. (2010). Coexpression of PD-1, 2B4, CD160 and KLRG1 on exhausted HCV-specific CD8+ T cells is linked to antigen recognition and T cell differentiation. *PLoS Pathog, 6*(6), e1000947. doi:10.1371/journal.ppat.1000947

Bentley, D. R., Balasubramanian, S., Swerdlow, H. P., Smith, G. P., Milton, J., Brown, C. G., . . . Smith, A. J. (2008). Accurate whole human genome sequencing using reversible terminator chemistry. *Nature, 456*(7218), 53-59. doi:10.1038/nature07517

Bertness, V., Kirsch, I., Hollis, G., Johnson, B., & Bunn, P. A., Jr. (1985). T-cell receptor gene rearrangements as clinical markers of human T-cell lymphomas. *N Engl J Med, 313*(9), 534-538. doi:10.1056/nejm198508293130902

Bevan, M. J. (1997). In thymic selection, peptide diversity gives and takes away. *Immunity, 7*(2), 175-178. doi:10.1016/s1074-7613(00)80520-8

Bolotin, D. A., Poslavsky, S., Mitrophanov, I., Shugay, M., Mamedov, I. Z., Putintseva, E. V., & Chudakov, D. M. (2015). MiXCR: software for comprehensive adaptive immunity profiling. *Nat Methods, 12*(5), 380-381. doi:10.1038/nmeth.3364

Borrman, T., Pierce, B. G., Vreven, T., Baker, B. M., & Weng, Z. (2020). High-throughput modeling and scoring of TCR-pMHC complexes to predict cross-reactive peptides. *Bioinformatics, 36*(22-23), 5377-5385. doi:10.1093/bioinformatics/btaa1050

Braun, D. A., Hou, Y., Bakouny, Z., Ficial, M., Sant' Angelo, M., Forman, J., . . . Choueiri, T. K. (2020). Interplay of somatic alterations and immune infiltration modulates response to PD-1 blockade in advanced clear cell renal cell carcinoma. *Nat Med, 26*(6), 909-918. doi:10.1038/s41591-020-0839-y

Braun, D. A., Ishii, Y., Walsh, A. M., Van Allen, E. M., Wu, C. J., Shukla, S. A., & Choueiri, T. K. (2019). Clinical Validation of PBRM1 Alterations as a Marker of Immune Checkpoint Inhibitor Response in Renal Cell Carcinoma. *JAMA Oncol*. doi:10.1001/jamaoncol.2019.3158

Brodin, P., Jojic, V., Gao, T., Bhattacharya, S., Angel, C. J., Furman, D., . . . Davis, M. M. (2015). Variation in the human immune system is largely driven by non-heritable influences. *Cell, 160*(1-2), 37-47. doi:10.1016/j.cell.2014.12.020

Burnet, M. (1957). Cancer; a biological approach. I. The processes of control. *Br Med J, 1*(5022), 779-786. doi:10.1136/bmj.1.5022.779

Caron, M., St-Onge, P., Sontag, T., Wang, Y. C., Richer, C., Ragoussis, I., . . . Bourque, G. (2020). Single-cell analysis of childhood leukemia reveals a link between developmental states and ribosomal protein expression as a source of intra-individual heterogeneity. *Sci Rep, 10*(1), 8079. doi:10.1038/s41598-020-64929-x

Cha, E., Klinger, M., Hou, Y., Cummings, C., Ribas, A., Faham, M., & Fong, L. (2014). Improved survival with T cell clonotype stability after anti-CTLA-4 treatment in cancer patients. *Sci Transl Med, 6*(238), 238ra270. doi:10.1126/scitranslmed.3008211

Chakraborty, S., Hosen, M. I., Ahmed, M., & Shekhar, H. U. (2018). Onco-Multi-OMICS Approach: A New Frontier in Cancer Research. *BioMed Research International, 2018*, 9836256. doi:10.1155/2018/9836256

Chapman, C. G., Yamaguchi, R., Tamura, K., Weidner, J., Imoto, S., Kwon, J., . . . Kiyotani, K. (2016). Characterization of T-cell Receptor Repertoire in Inflamed

Tissues of Patients with Crohn's Disease Through Deep Sequencing. *Inflamm Bowel Dis, 22*(6), 1275-1285. doi:10.1097/MIB.0000000000000752

Chen, D. S., & Mellman, I. (2013). Oncology meets immunology: the cancer-immunity cycle. *Immunity, 39*(1), 1-10. doi:10.1016/j.immuni.2013.07.012

Chen, D. S., & Mellman, I. (2017). Elements of cancer immunity and the cancer-immune set point. *Nature, 541*(7637), 321-330. doi:10.1038/nature21349

Chen, H., Ndhlovu, Z. M., Liu, D., Porter, L. C., Fang, J. W., Darko, S., . . . Walker, B. D. (2012). TCR clonotypes modulate the protective effect of HLA class I molecules in HIV-1 infection. *Nat Immunol, 13*(7), 691-700. doi:10.1038/ni.2342

Chen, M. F., Chen, P. T., Chen, W. C., Lu, M. S., Lin, P. Y., & Lee, K. D. (2016). The role of PD-L1 in the radiation response and prognosis for esophageal squamous cell carcinoma related to IL-6 and T-cell immunosuppression. *Oncotarget, 7*(7), 7913-7924. doi:10.18632/oncotarget.6861

Clemente, M. J., Przychodzen, B., Jerez, A., Dienes, B. E., Afable, M. G., Husseinzadeh, H., . . . Maciejewski, J. P. (2013). Deep sequencing of the T-cell receptor repertoire in CD8+ T-large granular lymphocyte leukemia identifies signature landscapes. *Blood, 122*(25), 4077-4085. doi:10.1182/blood-2013-05-506386

Crotty, S., Johnston, R. J., & Schoenberger, S. P. (2010). Effectors and memories: Bcl-6 and Blimp-1 in T and B lymphocyte differentiation. *Nat Immunol, 11*(2), 114-120. doi:10.1038/ni.1837

Csardi, G., & Nepusz, T. (2005). The Igraph Software Package for Complex Network Research. *InterJournal, Complex Systems*, 1695.

Danaher, P., Warren, S., Dennis, L., D'Amico, L., White, A., Disis, M. L., . . . Fling, S. P. (2017). Gene expression markers of Tumor Infiltrating Leukocytes. *J Immunother Cancer, 5*, 18. doi:10.1186/s40425-017-0215-8

Danahy, D. B., Berton, R. R., & Badovinac, V. P. (2020). Cutting Edge: Antitumor Immunity by Pathogen-Specific CD8 T Cells in the Absence of Cognate Antigen Recognition. *J Immunol, 204*(6), 1431-1435. doi:10.4049/jimmunol.1901172

Dash, P., Fiore-Gartland, A. J., Hertz, T., Wang, G. C., Sharma, S., Souquette, A., . . . Thomas, P. G. (2017). Quantifiable predictive features define epitope-specific T cell receptor repertoires. *Nature, 547*(7661), 89-93. doi:10.1038/nature22383

Davey, M. P., Meyer, M. M., & Bakke, A. C. (1994). T cell receptor V beta gene expression in monozygotic twins. Discordance in CD8 subset and in disease states. *J Immunol, 152*(1), 315-321.

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., . . . Gingeras, T. R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics, 29*(1), 15-21. doi:10.1093/bioinformatics/bts635

Dunn, G. P., Old, L. J., & Schreiber, R. D. (2004a). The immunobiology of cancer immunosurveillance and immunoediting. *Immunity, 21*(2), 137-148. doi:10.1016/j.immuni.2004.07.017

Dunn, G. P., Old, L. J., & Schreiber, R. D. (2004b). The three Es of cancer immunoediting. *Annu Rev Immunol, 22*, 329-360. doi:10.1146/annurev.immunol.22.012703.104803

Efremova, M., Vento-Tormo, M., Teichmann, S. A., & Vento-Tormo, R. (2020). CellPhoneDB: inferring cell-cell communication from combined expression of multi-subunit ligand-receptor complexes. *Nat Protoc, 15*(4), 1484-1506. doi:10.1038/s41596-020-0292-x

Ehrlich, P. (1908). *Ueber den jetzigen Stand der Karzinomforschung*.

Emerson, R. O., Sherwood, A. M., Rieder, M. J., Guenthoer, J., Williamson, D. W., Carlson, C. S., . . . Robins, H. S. (2013). High-throughput sequencing of T-cell receptors reveals a homogeneous repertoire of tumour-infiltrating lymphocytes in ovarian cancer. *J Pathol, 231*(4), 433-440. doi:10.1002/path.4260

Erkes, D. A., Smith, C. J., Wilski, N. A., Caldeira-Dantas, S., Mohgbeli, T., & Snyder, C. M. (2017). Virus-Specific CD8(+) T Cells Infiltrate Melanoma Lesions and Retain Function Independently of PD-1 Expression. *J Immunol, 198*(7), 2979-2988. doi:10.4049/jimmunol.1601064

Eyles, J., Puaux, A. L., Wang, X., Toh, B., Prakash, C., Hong, M., . . . Abastado, J. P. (2010). Tumor cells disseminate early, but immunosurveillance limits metastatic outgrowth, in a mouse model of melanoma. *J Clin Invest, 120*(6), 2030-2039. doi:10.1172/JCI42002

Fairfax, B. P., Taylor, C. A., Watson, R. A., Nassiri, I., Danielli, S., Fang, H., . . . Middleton, M. R. (2020). Peripheral CD8(+) T cell characteristics associated with durable responses to immune checkpoint blockade in patients with metastatic melanoma. *Nat Med, 26*(2), 193-199. doi:10.1038/s41591-019-0734-6

Fang, H., Knezevic, B., Burnham, K. L., & Knight, J. C. (2016). XGR software for enhanced interpretation of genomic summary data, illustrated by application to immunological traits. *Genome Med, 8*(1), 129. doi:10.1186/s13073-016-0384-y

Feng, L., Qian, H., Yu, X., Liu, K., Xiao, T., Zhang, C., . . . Zhang, K. (2017). Heterogeneity of tumor-infiltrating lymphocytes ascribed to local immune status rather than neoantigens by multi-omics analysis of glioblastoma multiforme. *Sci Rep, 7*(1), 6968. doi:10.1038/s41598-017-05538-z

Finn, O. J. (2008). Cancer immunology. *N Engl J Med, 358*(25), 2704-2715. doi:10.1056/NEJMra072739

Ganesan, A. P., Clarke, J., Wood, O., Garrido-Martin, E. M., Chee, S. J., Mellows, T., . . . Ottensmeier, C. H. (2017). Tissue-resident memory features are linked to the magnitude of cytotoxic T cell responses in human lung cancer. *Nat Immunol, 18*(8), 940-950. doi:10.1038/ni.3775

Gao, Y., Yang, W., Pan, M., Scully, E., Girardi, M., Augenlicht, L. H., . . . Yin, Z. (2003). Gamma delta T cells provide an early source of interferon gamma in tumor immunity. *J Exp Med, 198*(3), 433-442. doi:10.1084/jem.20030584

Gene Ontology, C. (2021). The Gene Ontology resource: enriching a GOld mine. *Nucleic Acids Res, 49*(D1), D325-D334. doi:10.1093/nar/gkaa1113

Gerlinger, M., Quezada, S. A., Peggs, K. S., Furness, A. J., Fisher, R., Marafioti, T., . . . Swanton, C. (2013). Ultra-deep T cell receptor sequencing reveals the complexity and intratumour heterogeneity of T cell clones in renal cell carcinomas. *J Pathol, 231*(4), 424-432. doi:10.1002/path.4284

Ghorani, E., Reading, J. L., Henry, J. Y., de Massy, M. R., Rosenthal, R., Turati, V., . . . Quezada, S. A. (2020). The T cell differentiation landscape is shaped by tumour mutations in lung cancer. *Nat Cancer, 1*(5), 546-561. doi:10.1038/s43018-020-0066-y

Girardi, M., Glusac, E., Filler, R. B., Roberts, S. J., Propperova, I., Lewis, J., . . . Hayday, A. C. (2003). The distinct contributions of murine T cell receptor (TCR)gammadelta+ and TCRalphabeta+ T cells to different stages of chemically induced skin cancer. *J Exp Med, 198*(5), 747-755. doi:10.1084/jem.20021282

Girardi, M., Oppenheim, D. E., Steele, C. R., Lewis, J. M., Glusac, E., Filler, R., . . . Hayday, A. C. (2001). Regulation of cutaneous malignancy by gammadelta T cells. *Science, 294*(5542), 605-609. doi:10.1126/science.1063916

Giudicelli, V., Chaume, D., & Lefranc, M. P. (2005). IMGT/GENE-DB: a comprehensive database for human and mouse immunoglobulin and T cell receptor genes. *Nucleic Acids Res, 33*(Database issue), D256-261. doi:10.1093/nar/gki010

Glanville, J., Huang, H., Nau, A., Hatton, O., Wagar, L. E., Rubelt, F., . . . Davis, M. M. (2017). Identifying specificity groups in the T cell receptor repertoire. *Nature, 547*(7661), 94-98. doi:10.1038/nature22976

Goronzy, J. J., Zettl, A., & Weyand, C. M. (1998). T cell receptor repertoire in rheumatoid arthritis. *Int Rev Immunol, 17*(5-6), 339-363. doi:10.3109/08830189809054410

Gros, A., Robbins, P. F., Yao, X., Li, Y. F., Turcotte, S., Tran, E., . . . Rosenberg, S. A. (2014). PD-1 identifies the patient-specific CD8(+) tumor-reactive repertoire infiltrating human tumors. *J Clin Invest, 124*(5), 2246-2259. doi:10.1172/JCI73639

Guo, X., Zhang, Y., Zheng, L., Zheng, C., Song, J., Zhang, Q., . . . Zhang, Z. (2018). Global characterization of T cells in non-small-cell lung cancer by single-cell sequencing. *Nat Med, 24*(7), 978-985. doi:10.1038/s41591-018-0045-3

Han, J., Duan, J., Bai, H., Wang, Y., Wan, R., Wang, X., . . . Wang, J. (2020). TCR Repertoire Diversity of Peripheral PD-1(+)CD8(+) T Cells Predicts Clinical Outcomes after Immunotherapy in Patients with Non-Small Cell Lung Cancer. *Cancer Immunol Res, 8*(1), 146-154. doi:10.1158/2326-6066.CIR-19-0398

Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W. M., Zheng, S., Butler, A., . . . Satija, R. (2020). Integrated analysis of multimodal single-cell data. *bioRxiv*, 2020.2010.2012.335331. doi:10.1101/2020.10.12.335331

Hausser, J., & Strimmer, K. (2009). Entropy Inference and the James-Stein Estimator, with Application to Nonlinear Gene Association Networks. *J. Mach. Learn. Res., 10*, 1469–1484.

Hayakawa, Y., Rovero, S., Forni, G., & Smyth, M. J. (2003). Alpha-galactosylceramide (KRN7000) suppression of chemical- and oncogene-dependent carcinogenesis. *Proc Natl Acad Sci U S A, 100*(16), 9464-9469. doi:10.1073/pnas.1630663100

Heather, J. M., Best, K., Oakes, T., Gray, E. R., Roe, J. K., Thomas, N., . . . Chain, B. (2015). Dynamic Perturbations of the T-Cell Receptor Repertoire in Chronic HIV Infection and following Antiretroviral Therapy. *Front Immunol, 6*, 644. doi:10.3389/fimmu.2015.00644

Heather, J. M., Ismail, M., Oakes, T., & Chain, B. (2018). High-throughput sequencing of the T-cell receptor repertoire: pitfalls and opportunities. *Brief Bioinform, 19*(4), 554-565. doi:10.1093/bib/bbw138

Hogan, S. A., Courtier, A., Cheng, P. F., Jaberg-Bentele, N. F., Goldinger, S. M., Manuel, M., . . . Levesque, M. P. (2019). Peripheral Blood TCR Repertoire Profiling May Facilitate Patient Stratification for Immunotherapy against Melanoma. *Cancer Immunol Res, 7*(1), 77-85. doi:10.1158/2326-6066.Cir-18-0136

Hogquist, K. A., Gavin, M. A., & Bevan, M. J. (1993). Positive selection of CD8+ T cells induced by major histocompatibility complex binding peptides in fetal thymic organ culture. *Journal of Experimental Medicine, 177*(5), 1469-1473. doi:10.1084/jem.177.5.1469

Huang, A. C., Postow, M. A., Orlowski, R. J., Mick, R., Bengsch, B., Manne, S., . . . Wherry, E. J. (2017). T-cell invigoration to tumour burden ratio associated with anti-PD-1 response. *Nature, 545*(7652), 60-65. doi:10.1038/nature22079

Huang, H., Wang, C., Rubelt, F., Scriba, T. J., & Davis, M. M. (2020). Analyzing the Mycobacterium tuberculosis immune response by T-cell receptor clustering with GLIPH2 and genome-wide antigen screening. *Nat Biotechnol, 38*(10), 1194-1202. doi:10.1038/s41587-020-0505-4

Hunger, S. P., & Mullighan, C. G. (2015). Acute Lymphoblastic Leukemia in Children. *N Engl J Med, 373*(16), 1541-1552. doi:10.1056/NEJMra1400972

Imai, K., Matsuyama, S., Miyake, S., Suga, K., & Nakachi, K. (2000). Natural cytotoxic activity of peripheral-blood lymphocytes and cancer incidence: an 11-year follow-up study of a general population. *Lancet, 356*(9244), 1795-1799. doi:10.1016/S0140-6736(00)03231-1

ImmunoMind, T. (2019). immunarch: An R Package for Painless Analysis of Large-Scale Immune Repertoire Data. In.

Jamal-Hanjani, M., Hackshaw, A., Ngai, Y., Shaw, J., Dive, C., Quezada, S., . . . Swanton, C. (2014). Tracking genomic cancer evolution for precision medicine: the lung TRACERx study. *PLoS Biol, 12*(7), e1001906. doi:10.1371/journal.pbio.1001906

Jamal-Hanjani, M., Wilson, G. A., McGranahan, N., Birkbak, N. J., Watkins, T. B. K., Veeriah, S., . . . Consortium, T. R. (2017). Tracking the Evolution of Non-Small-Cell Lung Cancer. *N Engl J Med, 376*(22), 2109-2121. doi:10.1056/NEJMoa1616288

Joshi, K., de Massy, M. R., Ismail, M., Reading, J. L., Uddin, I., Woolston, A., . . . Chain, B. (2019). Spatial heterogeneity of the T cell receptor repertoire reflects the mutational landscape in lung cancer. *Nat Med, 25*(10), 1549-1559. doi:10.1038/s41591-019-0592-2

Kaplinsky, J., & Arnaout, R. (2016). Robust estimates of overall immune-repertoire diversity from high-throughput measurements on samples. *Nature Communications, 7*(1), 11881. doi:10.1038/ncomms11881

Karatzoglou, A., Smola, A., Hornik, K., & Zeileis, A. (2004). kernlab - An S4 Package for Kernel Methods in R. *2004, 11*(9), 20. doi:10.18637/jss.v011.i09

Kirsch, I., Vignali, M., & Robins, H. (2015). T-cell receptor profiling in cancer. *Molecular Oncology, 9*(10), 2063-2070. doi:https://doi.org/10.1016/j.molonc.2015.09.003

Kloor, M., von Knebel Doeberitz, M., & Gebert, J. F. (2005). Molecular testing for microsatellite instability and its value in tumor characterization. *Expert Rev Mol Diagn, 5*(4), 599-611. doi:10.1586/14737159.5.4.599

Koebel, C. M., Vermi, W., Swann, J. B., Zerafa, N., Rodig, S. J., Old, L. J., . . . Schreiber, R. D. (2007). Adaptive immunity maintains occult cancer in an equilibrium state. *Nature, 450*(7171), 903-907. doi:10.1038/nature06309

Korsmeyer, S. J., Arnold, A., Bakhshi, A., Ravetch, J. V., Siebenlist, U., Hieter, P. A., . . . Waldmann, T. A. (1983). Immunoglobulin gene rearrangement and cell surface antigen expression in acute lymphocytic leukemias of T cell and B cell precursor origins. *J Clin Invest, 71*(2), 301-313. doi:10.1172/jci110770

Kuang, M., Cheng, J., Zhang, C., Feng, L., Xu, X., Zhang, Y., . . . Cheng, S. (2017). A novel signature for stratifying the molecular heterogeneity of the tissue-infiltrating T-cell receptor repertoire reflects gastric cancer prognosis. *Sci Rep, 7*(1), 7762. doi:10.1038/s41598-017-08289-z

Kvistborg, P., Philips, D., Kelderman, S., Hageman, L., Ottensmeier, C., Joseph-Pietras, D., . . . Schumacher, T. N. (2014). Anti-CTLA-4 therapy broadens the melanoma-reactive CD8+ T cell response. *Sci Transl Med, 6*(254), 254ra128. doi:10.1126/scitranslmed.3008918

Leruste, A., Tosello, J., Ramos, R. N., Tauziede-Espariat, A., Brohard, S., Han, Z. Y., . . . Bourdeaut, F. (2019). Clonally Expanded T Cells Reveal Immunogenicity of Rhabdoid Tumors. *Cancer Cell, 36*(6), 597-612 e598. doi:10.1016/j.ccell.2019.10.008

Li, B., & Dewey, C. N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics, 12*, 323. doi:10.1186/1471-2105-12-323

Li, B., Li, T., Wang, B., Dou, R., Zhang, J., Liu, J. S., & Liu, X. S. (2017). Ultrasensitive detection of TCR hypervariable-region sequences in solid-tissue RNA-seq data. *Nat Genet, 49*(4), 482-483. doi:10.1038/ng.3820

Li, H., van der Leun, A. M., Yofe, I., Lubling, Y., Gelbard-Solodkin, D., van Akkooi, A. C. J., . . . Amit, I. (2019). Dysfunctional CD8 T Cells Form a Proliferative,

Dynamically Regulated Compartment within Human Melanoma. *Cell, 176*(4), 775-789 e718. doi:10.1016/j.cell.2018.11.043

Li, H., Ye, C., Ji, G., Wu, X., Xiang, Z., Li, Y., . . . Han, J. (2012). Recombinatorial biases and convergent recombination determine interindividual TCRbeta sharing in murine thymocytes. *J Immunol, 189*(5), 2404-2413. doi:10.4049/jimmunol.1102087

Litchfield, K., Reading, J. L., Lim, E. L., Xu, H., Liu, P., Al-Bakir, M., . . . Swanton, C. (2020). Escape from nonsense-mediated decay associates with anti-tumor immunogenicity. *Nat Commun, 11*(1), 3800. doi:10.1038/s41467-020-17526-5

Litchfield, K., Reading, J. L., Puttick, C., Thakkar, K., Abbosh, C., Bentham, R., . . . Swanton, C. (2021). Meta-analysis of tumor- and T cell-intrinsic mechanisms of sensitization to checkpoint inhibition. *Cell, 184*(3), 596-614 e514. doi:10.1016/j.cell.2021.01.002

Logan, A. C., Gao, H., Wang, C., Sahaf, B., Jones, C. D., Marshall, E. L., . . . Miklos, D. B. (2011). High-throughput VDJ sequencing for quantification of minimal residual disease in chronic lymphocytic leukemia and immune reconstitution assessment. *Proc Natl Acad Sci U S A, 108*(52), 21194-21199. doi:10.1073/pnas.1118357109

Lopez-Nevot, M. A., Esteban, F., Ferron, A., Gutierrez, J., Oliva, M. R., Romero, C., . . . Garrido, F. (1989). HLA class I gene expression on human primary tumours and autologous metastases: demonstration of selective losses of HLA antigens on colorectal, gastric and laryngeal carcinomas. *Br J Cancer, 59*(2), 221-226. doi:10.1038/bjc.1989.45

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol, 15*(12), 550. doi:10.1186/s13059-014-0550-8

Lu, B., & Finn, O. J. (2008). T-cell death and cancer immune tolerance. *Cell Death Differ, 15*(1), 70-79. doi:10.1038/sj.cdd.4402274

Lucca, L. E., Axisa, P. P., Lu, B., Harnett, B., Jessel, S., Zhang, L., . . . Hafler, D. A. (2021). Circulating clonally expanded T cells reflect functions of tumor-infiltrating T cells. *J Exp Med, 218*(4). doi:10.1084/jem.20200921

Mahe, E., Pugh, T., & Kamel-Reid, S. (2018). T cell clonality assessment: past, present and future. *J Clin Pathol, 71*(3), 195-200. doi:10.1136/jclinpath-2017-204761

McDermott, D. F., Huseni, M. A., Atkins, M. B., Motzer, R. J., Rini, B. I., Escudier, B., . . . Powles, T. (2018). Clinical activity and molecular correlates of response to atezolizumab alone or in combination with bevacizumab versus sunitinib in renal cell carcinoma. *Nat Med, 24*(6), 749-757. doi:10.1038/s41591-018-0053-3

McGranahan, N., Furness, A. J., Rosenthal, R., Ramskov, S., Lyngaa, R., Saini, S. K., . . . Swanton, C. (2016). Clonal neoantigens elicit T cell immunoreactivity and sensitivity to immune checkpoint blockade. *Science, 351*(6280), 1463-1469. doi:10.1126/science.aaf1490

McGranahan, N., Rosenthal, R., Hiley, C. T., Rowan, A. J., Watkins, T. B. K., Wilson, G. A., . . . Consortium, T. R. (2017). Allele-Specific HLA Loss and Immune Escape in Lung Cancer Evolution. *Cell, 171*(6), 1259-1271 e1211. doi:10.1016/j.cell.2017.10.001

McLeod, C., Gout, A. M., Zhou, X., Thrasher, A., Rahbarinia, D., Brady, S. W., . . . Zhang, J. (2021). St. Jude Cloud: A Pediatric Cancer Genomic Data-Sharing Ecosystem. *Cancer Discov*. doi:10.1158/2159-8290.CD-20-1230

Messaoudi, I., Patiño, J. A. G., Dyall, R., LeMaoult, J., & Nikolich-Žugich, J. (2002). Direct Link Between mhc Polymorphism, T Cell Avidity, and Diversity in Immune Defense. *Science, 298*(5599), 1797-1800. doi:10.1126/science.1076064

Miao, D., Margolis, C. A., Gao, W., Voss, M. H., Li, W., Martini, D. J., . . . Van Allen, E. M. (2018). Genomic correlates of response to immune checkpoint therapies in clear cell renal cell carcinoma. *Science, 359*(6377), 801-806. doi:10.1126/science.aan5951

Miho, E., Yermanos, A., Weber, C. R., Berger, C. T., Reddy, S. T., & Greiff, V. (2018). Computational Strategies for Dissecting the High-Dimensional Complexity of Adaptive Immune Repertoires. *Front Immunol, 9*, 224. doi:10.3389/fimmu.2018.00224

Miles, J. J., Bulek, A. M., Cole, D. K., Gostick, E., Schauenburg, A. J., Dolton, G., . . . Sewell, A. K. (2010). Genetic and structural basis for selection of a ubiquitous T cell receptor deployed in Epstein-Barr virus infection. *PLoS Pathog, 6*(11), e1001198. doi:10.1371/journal.ppat.1001198

Mora, T., & Walczak, A. M. (2016). Quantifying lymphocyte receptor diversity. *bioRxiv*, 046870. doi:10.1101/046870

Mora, T., & Walczak, A. M. (2019). How many different clonotypes do immune repertoires contain? *bioRxiv*, 707885. doi:10.1101/707885

Moris, P., De Pauw, J., Postovskaya, A., Gielis, S., De Neuter, N., Bittremieux, W., . . . Meysman, P. (2020). Current challenges for unseen-epitope TCR interaction prediction and a new perspective derived from image classification. *Briefings in Bioinformatics*. doi:10.1093/bib/bbaa318

Motzer, R. J., Banchereau, R., Hamidi, H., Powles, T., McDermott, D., Atkins, M. B., . . . Rini, B. (2020). Molecular Subsets in Renal Cancer Determine Outcome to Checkpoint and Angiogenesis Blockade. *Cancer Cell, 38*(6), 803-817 e804. doi:10.1016/j.ccell.2020.10.011

Motzer, R. J., Robbins, P. B., Powles, T., Albiges, L., Haanen, J. B., Larkin, J., . . . Choueiri, T. K. (2020). Avelumab plus axitinib versus sunitinib in advanced renal cell carcinoma: biomarker analysis of the phase 3 JAVELIN Renal 101 trial. *Nat Med, 26*(11), 1733-1741. doi:10.1038/s41591-020-1044-8

Naidus, E., Bouquet, J., Oh, D. Y., Looney, T. J., Yang, H., Fong, L., . . . Zhang, L. (2021). Early changes in the circulating T cells are associated with clinical outcomes after PD-L1 blockade by durvalumab in advanced NSCLC patients. *Cancer Immunol Immunother*. doi:10.1007/s00262-020-02833-z

Nazarov, V. I., Pogorelyy, M. V., Komech, E. A., Zvyagin, I. V., Bolotin, D. A., Shugay, M., . . . Mamedov, I. Z. (2015). tcR: an R package for T cell receptor repertoire advanced data analysis. *BMC Bioinformatics, 16*(1), 175. doi:10.1186/s12859-015-0613-1

Ndifon, W., Gal, H., Shifrut, E., Aharoni, R., Yissachar, N., Waysbort, N., . . . Friedman, N. (2012). Chromatin conformation governs T-cell receptor Jbeta gene segment usage. *Proc Natl Acad Sci U S A, 109*(39), 15865-15870. doi:10.1073/pnas.1203916109

Newell, E. W., & Becht, E. (2018). High-Dimensional Profiling of Tumor-Specific Immune Responses: Asking T Cells about What They "See" in Cancer. *Cancer Immunol Res, 6*(1), 2-9. doi:10.1158/2326-6066.Cir-17-0519

Nikolic-Zugic, J., & Bevan, M. J. (1990). Role of self-peptides in positively selecting the T-cell repertoire. *Nature, 344*(6261), 65-67. doi:10.1038/344065a0

O'Connor, D., Enshaei, A., Bartram, J., Hancock, J., Harrison, C. J., Hough, R., . . . Goulden, N. (2018). Genotype-Specific Minimal Residual Disease Interpretation Improves Stratification in Pediatric Acute Lymphoblastic Leukemia. *J Clin Oncol, 36*(1), 34-43. doi:10.1200/JCO.2017.74.0449

Oakes, T., Heather, J. M., Best, K., Byng-Maddick, R., Husovsky, C., Ismail, M., . . . Chain, B. (2017). Quantitative Characterization of the T Cell Receptor Repertoire of Naïve and Memory Subsets Using an Integrated Experimental

and Computational Pipeline Which Is Robust, Economical, and Versatile. *Frontiers in Immunology, 8*(1267). doi:10.3389/fimmu.2017.01267

Oh, D. Y., Cham, J., Zhang, L., Fong, G., Kwek, S. S., Klinger, M., . . . Fong, L. (2017). Immune Toxicities Elicted by CTLA-4 Blockade in Cancer Patients Are Associated with Early Diversification of the T-cell Repertoire. *Cancer Res, 77*(6), 1322-1330. doi:10.1158/0008-5472.CAN-16-2324

Ohnmacht, G. A., & Marincola, F. M. (2000). Heterogeneity in expression of human leukocyte antigens and melanoma-associated antigens in advanced melanoma. *J Cell Physiol, 182*(3), 332-338. doi:10.1002/(SICI)1097-4652(200003)182:3<332::AID-JCP3>3.0.CO;2-Z

Pasetto, A., Gros, A., Robbins, P. F., Deniger, D. C., Prickett, T. D., Matus-Nicodemos, R., . . . Rosenberg, S. A. (2016). Tumor- and Neoantigen-Reactive T-cell Receptors Can Be Identified Based on Their Frequency in Fresh Tumor. *Cancer Immunol Res, 4*(9), 734-743. doi:10.1158/2326-6066.Cir-16-0001

Peacock, T., Heather, J. M., Ronel, T., & Chain, B. (2020). Decombinator V4 - an improved AIRR-compliant software package for T cell receptor sequence annotation. *Bioinformatics*. doi:10.1093/bioinformatics/btaa758

Pogorelyy, M. V., Minervina, A. A., Shugay, M., Chudakov, D. M., Lebedev, Y. B., Mora, T., & Walczak, A. M. (2019). Detecting T cell receptors involved in immune responses from single repertoire snapshots. *PLoS Biol, 17*(6), e3000314. doi:10.1371/journal.pbio.3000314

Qi, Q., Cavanagh, M. M., Le Saux, S., NamKoong, H., Kim, C., Turgano, E., . . . Goronzy, J. J. (2016). Diversification of the antigen-specific T cell receptor repertoire after varicella zoster vaccination. *Sci Transl Med, 8*(332), 332ra346. doi:10.1126/scitranslmed.aaf1725

Qi, Q., Liu, Y., Cheng, Y., Glanville, J., Zhang, D., Lee, J. Y., . . . Goronzy, J. J. (2014). Diversity and clonal selection in the human T-cell repertoire. *Proc Natl Acad Sci U S A, 111*(36), 13139-13144. doi:10.1073/pnas.1409155111

Reuben, A., Spencer, C. N., Prieto, P. A., Gopalakrishnan, V., Reddy, S. M., Miller, J. P., . . . Wargo, J. A. (2017). Genomic and immune heterogeneity are associated with differential responses to therapy in melanoma. *NPJ Genom Med, 2*. doi:10.1038/s41525-017-0013-8

Ricketts, C. J., De Cubas, A. A., Fan, H., Smith, C. C., Lang, M., Reznik, E., . . . Linehan, W. M. (2018). The Cancer Genome Atlas Comprehensive Molecular Characterization of Renal Cell Carcinoma. *Cell Rep, 23*(1), 313-326 e315. doi:10.1016/j.celrep.2018.03.075

Rizvi, N. A., Hellmann, M. D., Snyder, A., Kvistborg, P., Makarov, V., Havel, J. J., . . . Chan, T. A. (2015). Cancer immunology. Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer. *Science, 348*(6230), 124-128. doi:10.1126/science.aaa1348

Robert, L., Tsoi, J., Wang, X., Emerson, R., Homet, B., Chodon, T., . . . Ribas, A. (2014). CTLA4 blockade broadens the peripheral T-cell receptor repertoire. *Clin Cancer Res, 20*(9), 2424-2432. doi:10.1158/1078-0432.CCR-13-2648

Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics, 26*(1), 139-140.

Roh, W., Chen, P. L., Reuben, A., Spencer, C. N., Prieto, P. A., Miller, J. P., . . . Futreal, P. A. (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*(379). doi:10.1126/scitranslmed.aah3560

Roithmaier, S., Haydon, A. M., Loi, S., Esmore, D., Griffiths, A., Bergin, P., . . . Schwarz, M. A. (2007). Incidence of malignancies in heart and/or lung

transplant recipients: a single-institution experience. *J Heart Lung Transplant, 26*(8), 845-849. doi:10.1016/j.healun.2007.05.019

Rooney, M. S., Shukla, S. A., Wu, C. J., Getz, G., & Hacohen, N. (2015). Molecular and genetic properties of tumors associated with local immune cytolytic activity. *Cell, 160*(1-2), 48-61. doi:10.1016/j.cell.2014.12.033

Rosati, E., Dowds, C. M., Liaskou, E., Henriksen, E. K. K., Karlsen, T. H., & Franke, A. (2017). Overview of methodologies for T-cell receptor repertoire analysis. *BMC Biotechnology, 17*(1), 61. doi:10.1186/s12896-017-0379-9

Rosenthal, R., Cadieux, E. L., Salgado, R., Bakir, M. A., Moore, D. A., Hiley, C. T., . . . consortium, T. R. (2019). Neoantigen-directed immune escape in lung cancer evolution. *Nature, 567*(7749), 479-485. doi:10.1038/s41586-019-1032-7

Rudd, B. D., Venturi, V., Davenport, M. P., & Nikolich-Zugich, J. (2011). Evolution of the antigen-specific CD8+ TCR repertoire across the life span: evidence for clonal homogenization of the old TCR repertoire. *J Immunol, 186*(4), 2056-2064. doi:10.4049/jimmunol.1003013

Satija, R., Farrell, J. A., Gennert, D., Schier, A. F., & Regev, A. (2015). Spatial reconstruction of single-cell gene expression data. *Nat Biotechnol, 33*(5), 495-502. doi:10.1038/nbt.3192

Schalper, K. A., Rodriguez-Ruiz, M. E., Diez-Valle, R., Lopez-Janeiro, A., Porciuncula, A., Idoate, M. A., . . . Melero, I. (2019). Neoadjuvant nivolumab modifies the tumor immune microenvironment in resectable glioblastoma. *Nat Med, 25*(3), 470-476. doi:10.1038/s41591-018-0339-5

Scheper, W., Kelderman, S., Fanchi, L. F., Linnemann, C., Bendle, G., de Rooij, M. A. J., . . . Schumacher, T. N. (2019). Low and variable tumor reactivity of the intratumoral TCR repertoire in human cancers. *Nat Med, 25*(1), 89-94. doi:10.1038/s41591-018-0266-5

Schreiber, R. D., Old, L. J., & Smyth, M. J. (2011). Cancer immunoediting: integrating immunity's roles in cancer suppression and promotion. *Science, 331*(6024), 1565-1570. doi:10.1126/science.1203486

Seliger, B., Maeurer, M. J., & Ferrone, S. (2000). Antigen-processing machinery breakdown and tumor growth. *Immunol Today, 21*(9), 455-464. doi:10.1016/s0167-5699(00)01692-3

Shankaran, V., Ikeda, H., Bruce, A. T., White, J. M., Swanson, P. E., Old, L. J., & Schreiber, R. D. (2001). IFNgamma and lymphocytes prevent primary tumour development and shape tumour immunogenicity. *Nature, 410*(6832), 1107-1111. doi:10.1038/35074122

Shugay, M., Bagaev, D. V., Turchaninova, M. A., Bolotin, D. A., Britanova, O. V., Putintseva, E. V., . . . Chudakov, D. M. (2015). VDJtools: Unifying Post-analysis of T Cell Receptor Repertoires. *PLOS Computational Biology, 11*(11), e1004503. doi:10.1371/journal.pcbi.1004503

Simoni, Y., Becht, E., Fehlings, M., Loh, C. Y., Koo, S. L., Teng, K. W. W., . . . Newell, E. W. (2018). Bystander CD8(+) T cells are abundant and phenotypically distinct in human tumour infiltrates. *Nature, 557*(7706), 575-579. doi:10.1038/s41586-018-0130-2

Six, A., Mariotti-Ferrandiz, M. E., Chaara, W., Magadan, S., Pham, H. P., Lefranc, M. P., . . . Boudinot, P. (2013). The past, present, and future of immune repertoire biology - the rise of next-generation repertoire analysis. *Front Immunol, 4*, 413. doi:10.3389/fimmu.2013.00413

Smyth, M. J., Crowe, N. Y., & Godfrey, D. I. (2001). NK cells and NKT cells collaborate in host protection from methylcholanthrene-induced fibrosarcoma. *Int Immunol, 13*(4), 459-463. doi:10.1093/intimm/13.4.459

Smyth, M. J., Thia, K. Y., Street, S. E., Cretney, E., Trapani, J. A., Taniguchi, M., . . . Godfrey, D. I. (2000). Differential tumor surveillance by natural killer (NK) and NKT cells. *J Exp Med, 191*(4), 661-668. doi:10.1084/jem.191.4.661

Springer, I., Besser, H., Tickotsky-Moskovitz, N., Dvorkin, S., & Louzoun, Y. (2020). Prediction of Specific TCR-Peptide Binding From Large Dictionaries of TCR-Peptide Pairs. *Frontiers in Immunology, 11*(1803). doi:10.3389/fimmu.2020.01803

Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W. M., 3rd, . . . Satija, R. (2019). Comprehensive Integration of Single-Cell Data. *Cell, 177*(7), 1888-1902 e1821. doi:10.1016/j.cell.2019.05.031

Sun, J., Sun, B., Gao, Y., He, F., Yang, L., Wang, M., & Zhou, W. (2017). Composition and Variation Analysis of the T Cell Receptor beta-Chain Complementarity Determining Region 3 Repertoire in Neonatal Sepsis. *Scand J Immunol, 86*(5), 418-423. doi:10.1111/sji.12614

Thomas, L. (1959). *Cellular and humoral aspects of the hypersensitive states : a symposium held at the New York academy of medicine*.

Thomas, N., Best, K., Cinelli, M., Reich-Zeliger, S., Gal, H., Shifrut, E., . . . Chain, B. (2014). Tracking global changes induced in the CD4 T-cell receptor repertoire by immunization with a complex antigen using short stretches of CDR3 protein sequence. *Bioinformatics, 30*(22), 3181-3188. doi:10.1093/bioinformatics/btu523

Thomas, N., Heather, J., Ndifon, W., Shawe-Taylor, J., & Chain, B. (2013). Decombinator: a tool for fast, efficient gene assignment in T-cell receptor sequences using a finite state machine. *Bioinformatics, 29*(5), 542-550. doi:10.1093/bioinformatics/btt004

Tickotsky, N., Sagiv, T., Prilusky, J., Shifrut, E., & Friedman, N. (2017). McPAS-TCR: a manually curated catalogue of pathology-associated T cell receptor sequences. *Bioinformatics, 33*(18), 2924-2929. doi:10.1093/bioinformatics/btx286

Tong, Y., Wang, J., Zheng, T., Zhang, X., Xiao, X., Zhu, X., . . . Liu, X. (2020). SETE: Sequence-based Ensemble learning approach for TCR Epitope binding prediction. *Computational Biology and Chemistry, 87*, 107281. doi:https://doi.org/10.1016/j.compbiolchem.2020.107281

Tumeh, P. C., Harview, C. L., Yearley, J. H., Shintaku, I. P., Taylor, E. J., Robert, L., . . . Ribas, A. (2014). PD-1 blockade induces responses by inhibiting adaptive immune resistance. *Nature, 515*(7528), 568-571. doi:10.1038/nature13954

Turajlic, S., Litchfield, K., Xu, H., Rosenthal, R., McGranahan, N., Reading, J. L., . . . Swanton, C. (2017). Insertion-and-deletion-derived tumour-specific neoantigens and the immunogenic phenotype: a pan-cancer analysis. *Lancet Oncol, 18*(8), 1009-1021. doi:10.1016/S1470-2045(17)30516-8

van Dongen, J. J., Langerak, A. W., Bruggemann, M., Evans, P. A., Hummel, M., Lavender, F. L., . . . Macintyre, E. A. (2003). Design and standardization of PCR primers and protocols for detection of clonal immunoglobulin and T-cell receptor gene recombinations in suspect lymphoproliferations: report of the BIOMED-2 Concerted Action BMH4-CT98-3936. *Leukemia, 17*(12), 2257-2317. doi:10.1038/sj.leu.2403202

Wall, R. J., & Shani, M. (2008). Are animal models as good as we think? *Theriogenology, 69*(1), 2-9. doi:10.1016/j.theriogenology.2007.09.030

Wang, T., Wang, C., Wu, J., He, C., Zhang, W., Liu, J., . . . Liu, X. (2017). The Different T-cell Receptor Repertoires in Breast Cancer Tumors, Draining Lymph Nodes, and Adjacent Tissues. *Cancer Immunol Res, 5*(2), 148-156. doi:10.1158/2326-6066.CIR-16-0107

Weaver, C., & Murphy, K. (2016). *Janeway's Immunobiology*(9 ed.).

Weber, G., Gerdemann, U., Caruana, I., Savoldo, B., Hensel, N. F., Rabin, K. R., . . . Bollard, C. M. (2013). Generation of multi-leukemia antigen-specific T cells to enhance the graft-versus-leukemia effect after allogeneic stem cell transplant. *Leukemia, 27*(7), 1538-1547. doi:10.1038/leu.2013.66

Whiteside, T. L. (2008). The tumor microenvironment and its role in promoting tumor growth. *Oncogene, 27*(45), 5904-5912. doi:10.1038/onc.2008.271

Wisskirchen, K., Metzger, K., Schreiber, S., Asen, T., Weigand, L., Dargel, C., . . . Protzer, U. (2017). Isolation and functional characterization of hepatitis B virus-specific T-cell receptors as new tools for experimental and clinical use. *PLoS One, 12*(8), e0182936. doi:10.1371/journal.pone.0182936

Wright, G., Watt, E., Inglott, S., Brooks, T., Bartram, J., & Adams, S. P. (2019). Clinical benefit of a high-throughput sequencing approach for minimal residual disease in acute lymphoblastic leukemia. *Pediatr Blood Cancer, 66*(8), e27787. doi:10.1002/pbc.27787

Wu, D., Sherwood, A., Fromm, J. R., Winter, S. S., Dunsmore, K. P., Loh, M. L., . . . Robins, H. (2012). High-throughput sequencing detects minimal residual disease in acute T lymphoblastic leukemia. *Sci Transl Med, 4*(134), 134ra163. doi:10.1126/scitranslmed.3003656

Wu, T. D., Madireddi, S., de Almeida, P. E., Banchereau, R., Chen, Y. J., Chitre, A. S., . . . Grogan, J. L. (2020). Peripheral T cell expansion predicts tumour infiltration and clinical response. *Nature, 579*(7798), 274-278. doi:10.1038/s41586-020-2056-8

Yanagi, Y., Yoshikai, Y., Leggett, K., Clark, S. P., Aleksander, I., & Mak, T. W. (1984). A human T cell-specific cDNA clone encodes a protein having extensive homology to immunoglobulin chains. *Nature, 308*(5955), 145-149. doi:10.1038/308145a0

Yost, K. E., Satpathy, A. T., Wells, D. K., Qi, Y., Wang, C., Kageyama, R., . . . Chang, H. Y. (2019). Clonal replacement of tumor-specific T cells following PD-1 blockade. *Nat Med, 25*(8), 1251-1259. doi:10.1038/s41591-019-0522-3

Yu, X. G., Lichterfeld, M., Williams, K. L., Martinez-Picado, J., & 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*(22), 12666-12669. doi:10.1128/JVI.01450-07

Zamora, A. E., Crawford, J. C., Allen, E. K., Guo, X. J., Bakke, J., Carter, R. A., . . . Thomas, P. G. (2019). Pediatric patients with acute lymphoblastic leukemia generate abundant and functional neoantigen-specific CD8(+) T cell responses. *Sci Transl Med, 11*(498). doi:10.1126/scitranslmed.aat8549

Zarnitsyna, V. I., Evavold, B. D., Schoettle, L. N., Blattman, J. N., & Antia, R. (2013). Estimating the diversity, completeness, and cross-reactivity of the T cell repertoire. *Front Immunol, 4*, 485. doi:10.3389/fimmu.2013.00485

Zhang, Z., Xiong, D., Wang, X., Liu, H., & Wang, T. (2021). Mapping the functional landscape of T cell receptor repertoires by single-T cell transcriptomics. *Nat Methods, 18*(1), 92-99. doi:10.1038/s41592-020-01020-3

Zhao, J., Chen, A. X., Gartrell, R. D., Silverman, A. M., Aparicio, L., Chu, T., . . . Rabadan, R. (2019). Immune and genomic correlates of response to anti-PD-1 immunotherapy in glioblastoma. *Nat Med, 25*(3), 462-469. doi:10.1038/s41591-019-0349-y