Transcriptomic and genomic studies of childhood T cell leukaemia

Lim Song Jie, Bram

Wellcome Sanger Institute
University of Cambridge

This dissertation was submitted for the degree of
Doctor of Philosophy

Peterhouse

September 2025

Declaration

This thesis is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the preface and specified in the text. It is not substantially the same as any work that has already been submitted, or is being concurrently submitted, for any degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the preface and specified in the text. It does not exceed the prescribed word limit for the relevant Degree Committee.

Lim Song Jie, Bram
September 2025

Abstract

Acute lymphoblastic leukaemia (ALL) is the most common cancer in children. It can be of either B cell (B-ALL) or T cell (T-ALL) lineage. T-ALL carries a poorer clinical prognosis and worse overall survival than B-ALL. Despite decades of research including large scale sequencing studies, there is currently no biomarker that is used clinically to predict refractory T-ALL at diagnosis, unlike in B-ALL. To address this, I analysed T-ALL using single cell transcriptome sequencing, which can define refractory leukaemia cell states with higher precision than bulk sequencing methods.

From a combined cohort of 58 individuals with T-ALL, I generated a large, well-annotated cellular atlas, comprising almost 550,000 cells, which includes around 380,000 leukaemia blasts and 170,000 normal cells. T-ALL blasts from different individuals exhibited diverse cell states, resembling various stages of T cell differentiation, as well as unconventional T cells and other innate-like lymphocytes. Within the normal cell compartment, I found that the presence of significant residual blasts following initial treatment resulted in depletion of memory B cells in the bone marrow.

By analysing single leukaemia blast transcriptomes, I discovered that refractory T-ALL is driven by the presence of blasts which express *ZBTB16*, a gene that encodes for the lineage-defining transcription factor of unconventional T cells and other innate-like lymphocytes. I validated this finding in bulk transcriptomes of independent cohorts, including a contemporary trial cohort with over 1,300 individuals with T-ALL, and demonstrated that *ZBTB16* expression is associated with poorer survival. Furthermore, I identified cell surface markers on *ZBTB16*+ blasts, which can be used as potential targets for immunotherapy against these chemotherapy resistant blasts.

Finally, I investigated the origin of refractory *ZBTB16*+ blasts in T-ALL by using somatic mutations derived from whole genome sequencing. I showed that these blasts were not restricted to any particular T-ALL genomic subtype or rearrangement state of their T cell receptor. By projecting somatic mutations, including copy number alterations and single nucleotide variants, onto single leukaemia blast transcriptomes, I demonstrated that in one individual, refractory *ZBTB16*+ blasts could be traced back to a 1% subclone at diagnosis.

Overall, my thesis presents refractory T-ALL as a distinct disease that is intrinsically resistant and exhibits a previously unknown *ZBTB16*+ phenotype. These findings raise the possibility

of clinical risk stratification of T-ALL and may provide new targets for drug development. Moreover, the analytical framework from my PhD may be applied to future sequencing studies of other types of leukaemia.

Acknowledgements

My PhD has been an incredible journey which went by far too quickly. I had the opportunity to explore really cutting-edge science and be surrounded by some of the most brilliant yet humble minds. The work presented in this thesis would not have been possible without the support of many people.

First and foremost, I must thank my primary supervisor, Sam Behjati, for putting your faith in me and challenging me to achieve all that I have. I enjoyed our scientific discussions, punctuated by your entertaining quotes, interesting analogies and opinions about the fundamental facts of life. I would also like to thank my co-supervisor, Muzlifah Haniffa, for your thoughtful insights into developmental immunology and single cell genomics, which form the underlying narrative of my dissertation. Special thanks to Holly Whitfield, who was my approachable day-to-day supervisor for the majority of my PhD, and to Nathan Anderson, who helped me settle in during my first year. My analysis of T-ALL would not have been possible without David O'Connor, who first conceived the project and painstakingly assembled the clinical cohorts. To the patients and families who provided these precious leukaemia samples, often during the darkest times of your lives, we are extremely grateful and hope that this piece of work may benefit future patients.

I am very fortunate to have met amazing friends and colleagues along the way, including Mi Trinh, Manas Dave, Anna Wenger, Taryn Treger, Sid Lawrence, Angus Hodder, Sarah Leiter, Henry Lee-Six, Lorcan Pigott-Dix, Jonathan Kennedy, Barbara Walkowiak, Conor Parks, Toochi Ogbonnah, Tim Coorens, Matthew Young, Tzen Szen Toh, Chenqu Suo, Kelvin Tuong, Dan Leongamornlet, Vijay Baskar and Laura Jardine. I am grateful for your generosity with your time and knowledge. More importantly, I will dearly miss our countless lunchtime conversations and numerous outings.

I am grateful to have received immense support from the laboratory, informatics and administrative teams of the Cellular Genomics programme, the Cancer, Ageing and Somatic Mutation programme, and the wider Wellcome Sanger Institute. Special mention goes to Martin Prete for efficiently sorting out my many data transfer requests. The haematology diagnostic laboratory of Great Ormond Street Hospital also played a crucial role in generating data for this project. I am grateful to the MB/PhD programme at the University of Cambridge for the opportunity to explore science alongside my medical studies and to A*STAR Singapore for funding me throughout these years.

I am forever indebted to my parents, Papa and Mama, for all the sacrifices you have made to support me every step along the way. You taught me to live a life of passion, integrity and optimism. And not forgetting my two younger brothers, Greg and Iain, you were a great source of fun and love growing up, and you still are today. Finally, to my partner, Krysia, you bore the greatest emotional brunt from my PhD, patiently listening to my frustrations while managing your own research struggles. Thank you for sticking with me through the ups and downs. I look forward to walking the next phase of our journey together.

Table of contents

List of figures

List of tables

Abbreviations

ALL	Acute lymphoblastic leukaemia
AML	Acute myeloid leukaemia
BAF	B allele frequency
BAM	Binary alignment map
bHLH	Basic helix-loop-helix
BMP	Bone marrow progenitor
Bulk RNA-seq	Bulk RNA sequencing
B-ALL	B cell acute lymphoblastic leukaemia
CAR	Chimeric antigen receptor
CCF	Cancer cell fraction
ChIP-seq	Chromatin immunoprecipitation and sequencing
CI	Confidence interval
CMP	Common myeloid progenitor
COG	Children's Oncology Group
DC	Dendritic cell
DN	Double-negative
DNA	Deoxyribonucleic acid
DNTT	DNA nucleotidylexotransferase (also see TdT)
DP	Double-positive
EFS	Event free survival
EGIL	European Group for the Immunological Characterization of Leukemias
ETP	Early T cell precursor
FISH	Fluorescence in situ hybridisation
GMP	Granulocyte-monocyte progenitor
GO	Gene Ontology
GOSH	Great Ormond Street Hospital
HD domain	Heterodimerisation domain
HR	Hazard ratio
HSC	Haematopoietic stem cell
HSPC	Haematopoietic stem and progenitor cell
IF	Induction failure
ILC	Innate lymphoid cell

Indel	Insertion/deletion
KLRB1	Killer cell lectin like receptor B1
LMPP	Lymphoid-primed multipotent progenitor
logCPM	Log2 counts per million
logFPKM	Log2 fragments per kilobase of transcript per million mapped reads
LOH	Loss of heterozygosity
LTi cell	Lymphoid tissue inducer cell
MAIT cell	Mucosal associated invariant T cell
MEMP	Mast cell-megakaryocyte-erythroid progenitor
MEP	Megakaryocyte-erythroid
MHC	Major histocompatibility complex
MLP	Multi-lymphoid progenitor
MPP	Multipotent progenitor
MRD	Minimal residual disease
NK cell	Natural killer cell
NKT cell	Natural killer T cell
nTh17 cell	Natural T helper 17 cell
NTS	Non-templated sequences
OS	Overall survival
PCA	Principal component analysis
pDC	Plasmacytoid dendritic cell
PLZF	Promyelocytic leukaemia zinc finger (also see ZBTB16)
PMC	Princess Maxima Center
RNA	Ribonucleic acid
RSS	Recombination signal sequences
SCP	Schwann cell precursor
scRNA-seq	Single cell RNA sequencing
scTCR-seq	Single cell T cell receptor sequencing
SNP	Single nucleotide polymorphism
SNV	Single nucleotide variant
SP	Single-positive
SV	Structural variant
TCR	T cell receptor
TdT	Terminal deoxynucleotidyl transferase (also see DNTT)

TRA/TRB/ TRG/TRD	T cell receptor alpha/beta/gamma/delta
T-ALL	T cell acute lymphoblastic leukaemia
UMAP	Uniform manifold approximation and projection
VAF	Variant allele frequency
VDJ	Variable-diversity-joining
WES	Whole exome sequencing
WGS	Whole genome sequencing
WHO	World Health Organization
ZBTB16	Zinc finger and BTB domain containing 16 (also see PLZF)

Chapter 1: Introduction

1.1 Relationship between T cell development and T cell leukaemia

The history of T cell development and T cell leukaemia goes back to the late 1950s, even before the formal discovery of the T cell itself. In 1958, Jacques Miller, having recently completed his medical degree in Australia, arrived in the United Kingdom to pursue a PhD at the Institute of Cancer Research in London. He investigated a form of lymphocytic leukaemia in mice which characteristically involves the thymus. It had been shown that inoculation of mice with leukaemia extracts could induce leukaemia, but only if this was performed during the neonatal period (Gross, 1951). Miller became interested in the role of the thymus in leukaemogenesis. He performed thymectomy (surgical removal of the thymus) on adult mice which he had previously injected leukaemia extracts at birth. None of these mice developed leukaemia (Miller, 1959a). However, implantation of thymus from mice of the same strain into thymectomised inoculated mice restored their potential for leukaemia development (Miller, 1959b). These results demonstrate the role of the thymus in leukaemogenesis.

Miller then wondered if neonatal thymectomy could similarly inhibit leukaemogenesis like adult thymectomy. The experiment, however, took an unexpected turn. The neonatal thymectomised mice became very unhealthy soon after weaning and died prematurely, regardless of whether they had been inoculated with leukaemia extracts. These observations led Miller to make his renowned claim that “the thymus at birth may be essential to life” (Miller, 1961a). Histological examination of mice which were thymectomised at birth revealed marked deficiency of lymphocytes in blood, lymph nodes and spleen. Notably, these mice failed to reject skin which were grafted from mice of a different strain, thereby suggesting an immunological role of the thymus (Miller, 1961b).

At this point, it was known that lymphocytes could produce antibodies and reject skin grafts. Additional data from Miller’s neonatal thymectomised mice showed severe reduction in both of these functions (Miller and Haddow, 1962). In contrast, the situation was different in birds. Removal of an organ in chickens called the bursa of Fabricius led to defects in antibody generation, whereas thymectomy resulted in failure to reject skin grafts (Warner et al., 1962). Miller had to resolve this conundrum. In 1966, he moved back to Australia and set up his lab at the Walter and Eliza Hall Institute of Medical Research in Melbourne. Together with his first PhD student, Graham Mitchell, they set out to perform a series of complex experiments involving mice whose immune system was completely eradicated by thymectomy and

irradiation (Miller and Mitchell, 1968; Mitchell and Miller, 1968a, 1968b). By injecting thymus cells, bone marrow cells, or both at the same time, they established that the bone marrow cells (now referred to as B cells) produced antibodies, while the thymus cells (now referred to as T cells) provided B cells with the necessary “help” to generate antibodies. These results for the first time defined T cells and B cells, as well as their functional relationship.

Miller’s work on the thymus illustrates the intimate relationship between T cell development and T cell leukaemia. Throughout my doctoral research, I examined childhood T cell leukaemia using single cell RNA sequencing (scRNA-seq) and whole genome sequencing (WGS), through the lens of T cell development. For the introduction of my dissertation, I will cover the advances in unravelling T cell development and the diversity of T cell subsets, followed by an overview of the current state of research in T cell leukaemia. I will then consider how measuring quantitative molecular readouts at single cell resolution has opened up many opportunities in the study of leukaemias.

1.2 T cell development and diverse T cell subtypes

T cells are key members of the adaptive immune system. Mature T cells are defined by surface expression of the T cell receptor (TCR), a heterodimeric molecule that binds to peptide antigens presented on major histocompatibility complex (MHC) molecules. Traditionally, T cells are divided into CD8+ T cells, which kill virus-infected host cells or cancer cells through their cytotoxic effector functions, and CD4+ T cells which coordinate the adaptive immune response through cytokine secretion.

T cell development mainly occurs in the thymus, but involves progenitors from the bone marrow. This highly ordered stepwise process includes commitment to the T cell lineage through a network of transcription factors, somatic recombination of genes to generate TCR, and positive and negative selection in the thymus to ensure the production of functional and non-autoreactive T cells. Aberrations in T cell development due to mutations in key transcription factors and signalling pathways may result in T cell acute lymphoblastic leukaemia (T-ALL) (Aifantis et al., 2008; Gianni et al., 2020).

Conventional CD8+ and CD4+ T cells express αβ TCR, which consists of an α chain and a β chain. However, 5-10% of T cells utilise γδ TCR instead of αβ TCR (Chien et al., 2014). Furthermore, some αβ T cells express αβ TCR with limited diversity, unlike the high TCR sequence diversity seen in conventional αβ T cells (Legoux et al., 2017). These γδ T cells and

αβ T cells with limited TCR diversity represent examples of the burgeoning group of unconventional T cells, which recognise distinct antigens compared to conventional αβ T cells and perform different immunological functions. In addition, natural killer (NK) cells and innate lymphoid cells (ILCs) resemble CD8+ and CD4+ T cells respectively, but lack TCR expression (Vivier et al., 2018).

In this section, I will first give an overview of the development of conventional T cells and highlight key aspects of this complex process, namely the commitment to T cell lineage and somatic recombination of TCR genes. Next, I will describe the unconventional T cells as well as the NK cells and ILCs, covering their developmental origin and their relationship to conventional T cells. Finally, I will highlight how single cell transcriptomic studies have transformed our understanding of T cell development and diversity.

1.2.1 Development of conventional T cells

The development of conventional αβ T cells begins with migration of progenitors from the bone marrow to the thymus (**Figure 1.1**) (Hosokawa and Rothenberg, 2021). These cells lack expression of both CD8 and CD4 surface markers and are hence called double negative (DN) thymocytes. DN thymocytes undergo rearrangement at the TCR-β gene locus. The resulting TCR-β chain pairs up with a surrogate pre-T cell receptor α chain (pTα) to form the pre-T cell receptor (pre-TCR). Signalling by the pre-TCR indicates successful TCR-β chain rearrangement and is required to permit further development. This critical gateway is known as β-selection. In the next stage, cells express both CD8 and CD4, thus becoming double positive (DP) thymocytes, and they undergo rearrangement at the TCR-α gene locus. Successful rearrangement of the TCR-α chain allows it to pair up with the TCR-β chain to form a complete TCR.

DP thymocytes then undergo positive selection, where their TCR has to recognise self-antigen presented by MHC molecules on thymic epithelial cells to avoid cell death. Upon successful positive selection, the cells become single-positive (SP) cells, expressing either CD8 or CD4, depending on whether they recognise MHC class I or class II molecules, respectively. Following this, SP cells undergo negative selection, where cells which interact too strongly with self-antigens displayed on thymic epithelial cells are eliminated by apoptosis. SP cells which pass negative selection exit the thymus as mature naive CD8+ or CD4+ T cells and migrate to the periphery.

Figure 1.1

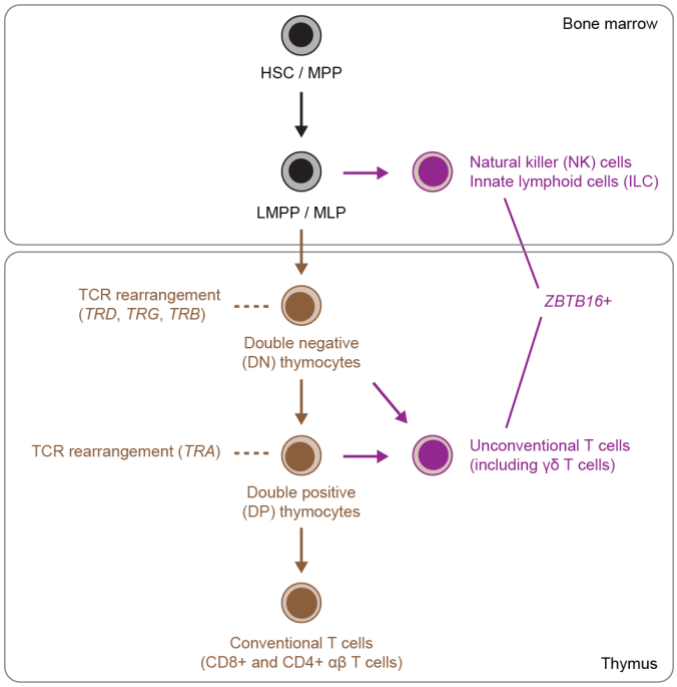


Figure 1.1: Schematic of T cell development.

Progenitors from the bone marrow migrate to the thymus and become double negative (DN) thymocytes. Here, they undergo TCR rearrangement at the *TRD*, *TRG* and/or *TRB* loci. Next, they transition to double positive (DP) thymocytes, during which rearrangement at the *TRA* locus occurs. Finally, the cells mature into conventional CD8+ and CD4+ αβ T cells, which will exit the thymus into the periphery. In the bone marrow, the same progenitors can develop into natural killer (NK) cells or innate lymphoid cells (ILC), instead for travelling to the thymus to become a T cell. Alternatively, DN and DP thymocytes in the thymus may mature into unconventional T cells, which include γδ T cells.

1.2.2 Commitment to T cell lineage

T cell lineage commitment is driven by a network of lineage-specific transcription factors. Notch signalling in the thymus is necessary for commitment to T cell lineage (Hosokawa and Rothenberg, 2021; Miyazaki et al., 2014). Deletion of *Notch1* in bone marrow progenitors or deletion of its ligand Delta-like-4 (*Dll4*) in the thymic epithelium results in the complete absence of T cells (Hozumi et al., 2008; Radtke et al., 1999). Provision of the Notch ligand Delta-like-1 in vitro stimulated the differentiation of haematopoietic progenitors into T cells (Schmitt and Zúñiga-Pflücker, 2002). An isoform of the E2A transcription factor, E47, was found to promote upregulation of *Notch1* and subsequently act in concert to upregulate T cell specific genes such as *Cd7* and *Ptcra* (pTα) (Ikawa et al., 2006). Additionally, E47 also increases chromatin accessibility of the TCR-β locus in preparation for somatic recombination (Agata et al., 2007).

Notch signalling in DN cells activates the expression of TCF1 (encoded by *Tcf7*), which in turn upregulates GATA3 and BCL11B (Germar et al., 2011; Weber et al., 2011). GATA3 blocks B cell lineage differentiation (García-Ojeda et al., 2013) while BCL11B inhibits NK and myeloid cell fates (Ikawa et al., 2010; L. Li et al., 2010; P. Li et al., 2010). BCL11B was found to suppress *Zbtb16*, which is crucial for the development of innate T cells and ILCs, and also suppress *Id2*, which is required for NK cells and ILCs (Hosokawa et al., 2018). TCF1 also activates the transcription of genes required for TCR rearrangement and TCR signalling, such as *Cd3g*, *Lck* and *Rag2* (Weber et al., 2011). Altogether, a network of transcription factors consisting of Notch, E47, TCF1, GATA3 and BCL11B promote the expression of T cell development genes and inhibit alternative cell fates.

1.2.3 Somatic recombination of T cell receptor genes

There are four different TCR chains encoded in the human genome: TCR-α (*TRA* on chromosome 14q11), TCR-β (*TRB* on chromosome 7q34), TCR-γ (*TRG* on chromosome 7p14) and TCR-δ (*TRD* on chromosome 14q11) (Lefranc et al., 1999). The *TRB* and *TRD* loci consist of the variable (V), diversity (D), joining (J) and constant (C) gene segments, while the *TRA* and *TRG* loci contain only V, J and C segments. Remarkably, the *TRD* locus is located within the *TRA* locus. Rearrangement of the *TRA* locus results in deletion of the *TRD*. The germline configurations of the TCR loci are illustrated in detail in **Figure 1.2**.

Somatic recombination or rearrangement assembles the separate germline-encoded VDJ gene segments, through genomic deletions or inversions of the intervening gene segments

(Krangel, 2009). On the *TRB* and *TRD* loci, the D-to-J joining occurs first, followed by the V segment to the DJ segment. These steps occur at the genomic level. Next, the *TRB* gene is expressed and RNA splicing joins the C segment to the VDJ segment at the transcript level. A productive TCR chain is formed when successful translation occurs with no premature stop codon. Somatic recombination of the *TRA* and *TRG* loci follow a similar process, except that V-to-J joining occurs directly owing to the absence of D segments. Rearrangement of the *TRB*, *TRG* and *TRD* loci occur earlier at the DN cell stage while *TRA* rearrangement occurs later at the DP cell stage (Hayday and Pennington, 2007).

VDJ recombination is mediated by the RAG endonucleases, RAG1 and RAG2, which are guided by recombination signal sequences (RSS) (Fugmann et al., 2000). These are conserved noncoding DNA sequences that flank the individual V, D and J gene segment coding sequences. An RSS consists of a highly conserved heptamer block (CACAGTG), which is always contiguous with the coding sequence, followed by a 12-bp or 23-bp long spacer sequence, followed by a less conserved nonamer block (ACAAAAACC).

RAG endonucleases bind VDJ gene segments at RSS sites and cleave DNA at the boundary between the RSS and the flanking coding sequence (Bassing et al., 2002). This generates a hairpin end next to the VDJ coding sequencing and a blunt end adjacent to the RSS motif. Hairpin ends from two VDJ gene segments are brought in close proximity by the RAG complex. Opening of the hairpin ends at a site other than the original breakpoint results in addition of P-nucleotides, named as such due to the resulting palindromic sequence. The lymphocyte-specific terminal deoxynucleotidyl transferase (TdT) enzyme adds non-templated sequences (NTS), also known as N-nucleotides, to the opened hairpins, resulting in further diversification of the antigen receptor (Komori et al., 1993). Nucleotide deletions may also occur at these VDJ junctions, mediated by a yet to be discovered exonuclease.

RAG-mediated recombination is not only confined to the genomic loci of B cell and T cell antigen receptors. By systematically interrogating the whole genomes of *ETV6-RUNX1* B-ALL, Papaemmanuil et al. (2014) found many recurrent structural variants of B-ALL driver genes to be mediated by RAG recombination, characterised by the presence of RSS motifs near the breakpoints and the incorporation of NTS at junctions. More recently, Machado et al. (2022) demonstrated through WGS of single lymphocyte derived colonies that a substantial fraction of deletions found in normal lymphocytes are attributable to off-target RAG activity.

Figure 1.2

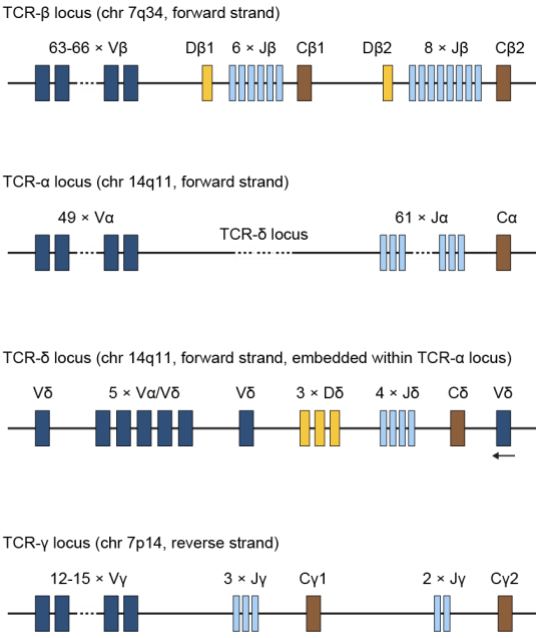


Figure 1.2: Germline configuration of TCR genes in the human genome.

The human TCR-β locus (*TRB* on chromosome 7q34) has a cluster of 63-66 Vβ gene segments located some distance from two separate clusters, that each contain one Dβ gene segment together with 6 or 8 Jβ gene segments and a single C gene segment. The human TCRα locus (*TRA* on chromosome 14q11) consists of 49 Vα gene segments, located considerably upstream from 61 Jα gene segments, followed by a single Cα gene. The TCRα locus is interrupted between the Vα and Jα gene segments by the TCR-δ locus (*TRD*), which contains 8 Vδ gene segments, 3 Dδ gene segments, 4 Jδ gene segments and a single Cδ gene segment. Five of the so-called Vδ gene segments can be used interchangeably to make either TCR-α or TCR-δ chain. The last Vδ gene segment, located at the 3' end of the TCR-δ locus, is in the reverse orientation. Rearrangement at the TCR-α locus will delete the intervening DNA including the TCR-δ locus. Finally, the human TCR-γ locus (*TRG* on chromosome 7p14) consists of 12-15 Vγ gene segments, followed by two separate clusters, each containing 3 or 2 Jγ gene segments and one Cγ gene segment.

1.2.4 Unconventional T cells

Unconventional T cells refer to T cell subsets other than those derived from canonical CD8+ and CD4+ αβ T cells (Godfrey et al., 2015; Pellicci et al., 2020). This burgeoning collection includes γδ T cells, natural killer T (NKT) cells, mucosal associated invariant T (MAIT) cells, CD8αα T cells and natural T helper 17 (nTh17) cells, among others. Unconventional T cell subsets are generally marked by expression of *ZBTB16* (zinc finger and BTB domain containing 16) (Alonzo and Sant'Angelo, 2011) and *KLRB1* (killer cell lectin like receptor B1) (Fergusson et al., 2011), genes which are of great relevance to this thesis.

- **γδ T cells** express TCR-γ chain in combination with TCR-δ chain and represent 5-10% of T cells in the human (Chien et al., 2014). The two main subsets of γδ T cells in humans are Vδ2-Vγ9 γδ T cells (defined by their specific TCR-δ and TCR-γ gene usage) which are found predominantly in the bloodstream, and Vδ1 γδ T cells which are more common in the skin and large intestine. Vδ2-Vγ9 γδ T cells respond to bacterial phosphoantigens presented on butyrophilin 3A1 molecules (Vavassori et al., 2013), while Vδ1 γδ T cells recognise diverse antigens including lipids presented on MHC class-I-like protein CD1d (Uldrich et al., 2013). The developmental origin of γδ T cells in humans remains ambiguous. It was initially claimed that human γδ T cells develop first in the fetal liver and then in the fetal thymus, with little contribution by the postnatal thymus (McVay et al., 1998; McVay and Carding, 1996). However, more recent studies have suggested that most γδ T cells in adult human derive from the postnatal rather than fetal thymus (Papadopoulou et al., 2019; Perriman et al., 2023).

- **NKT cells** were originally defined as T cells which expressed the NK cell protein NK1.1 (CD161 or KLRB1) (Makino et al., 1995). However, they have since been defined by their specificity for lipid antigens presented by CD1d molecules (Godfrey et al., 2004), in contrast to conventional T cells which recognise peptide antigens presented in MHC molecules. There exist two broad groups of NKT cells. Type 1 NKT cells, sometimes referred to as invariant NKT cells, express an invariant TCR-α chain (Vα10 and Jα18) and a limited set of TCR-β chain in humans. They make up around 0.1% of T cells in human blood (Chan et al., 2013). Type 2 NKT cells in humans express diverse TCRs that recognise a broader range of lipid antigens. Type 2 NKT cells are more abundant in humans, but their lack of any defining marker makes them challenging to study (De Lalla et al., 2011). The development of NKT cells occurs in the postnatal thymus (Hammond et al., 1998). NKT cell development branches off the CD8+ CD4+ DP cell stage of conventional T cell development and is dependent on interactions with CD1d+ DP cells

(Coles and Raulet, 2000; Gapin et al., 2001). NKT cell development also requires the transcription factor ZBTB16 (Kovalovsky et al., 2008; Savage et al., 2008).

- **MAIT cells** are a subset of T cells which respond to microbial vitamin B metabolites presented on the MHC class-I-like protein MR1 (Kjer-Nielsen et al., 2012; Tilloy et al., 1999; Treiner et al., 2003). Human MAIT cells typically express an invariant TCR-α (Vα7.2 and Jα33) and a constrained TCR-β chain (Vβ6 or Vβ20) (Ussher et al., 2014). MAIT cells make up around 5% of T cells in human blood (Le Bourhis et al., 2010) and up to 45% of T cells in the human liver (Dusseaux et al., 2011). The development of MAIT cells exhibit many similarities compared to NKT cells. Notably, MAIT cells develop in the postnatal thymus (Martin et al., 2009) and arise from interactions between MR1-expressing DP cells (Seach et al., 2013). ZBTB16 is also required for MAIT cell development (Koay et al., 2016).

- **CD8αα T cells** express the CD8αα homodimer instead of the CD8αβ found on conventional CD8+ T cells and exhibit signs of previous antigen exposure (Konno et al., 2002). They include a subset of intraepithelial lymphocytes (IELs) which play important roles in mucosal immunity (Cheroutre et al., 2011). CD8αα IELs arise directly from thymic precursors instead of deriving from conventional naive T cells (Ruscher et al., 2017; Verstichel et al., 2017).

- **nTh17 cells** closely resemble canonical Th17 cells, as both secrete IL-17 cytokines in response to extracellular bacteria (Marks et al., 2009). Whereas canonical Th17 cells are induced from CD4+ T cells by exposure to IL-6 and TGF-β (Veldhoen et al., 2006), nTh17 cells directly originate from thymic progenitors (Marks et al., 2009).

A common theme among unconventional T cells is their direct derivation from thymic precursors by agonist selection (Stritesky et al., 2012). In the development of conventional CD8+ or CD4+ T cells, thymocytes which react too strongly to self-antigen are eliminated by apoptosis (Ashby and Hogquist, 2024). Similarly, thymocytes which fail to react at all do not receive sufficient stimulation to survive. However, there are a subset of thymocytes which react more strongly than others, but below the threshold for negative selection; these thymocytes become natural regulatory T cells, in contrast to induced regulatory T cells which arise from CD4+ T cells in the periphery (Jordan et al., 2001). Agonist selection has since also been demonstrated in the development of various subtypes of unconventional T cells (Stritesky et al., 2012). Unconventional T cells bearing αβ TCR may derive from the DP cell

stage in the thymus, whereas γδ T cells arise earlier from the DN cell stage (**Figure 1.1**) (Hayday and Pennington, 2007).

1.2.5 Innate lymphoid cells

Innate lymphoid cells (ILCs) are lymphocytes which lack antigen receptors generated by RAG recombination (i.e. TCR and BCR) (Artis and Spits, 2015; Eberl et al., 2015; Vivier et al., 2018). This definition includes NK cells, which were discovered in 1975 (Kiessling et al., 1975). NK cells mirror CD8+ cytotoxic T cells by killing virus-infected cells and tumour cells (Wolf et al., 2023). The remaining ILCs are non-cytotoxic and comprise group 1 ILC (ILC1), group 2 ILC (ILC2), group 3 ILC (ILC3) and lymphoid tissue inducer (LTi) cells. The LTi cells, which promote the formation of lymph nodes during fetal development, were discovered in 1997 (Mebius et al., 1997). However, the concept of a family of ILCs only emerged within the last 15 years, following the discovery of the remaining three ILC classes (Spits et al., 2013).

ILC1, ILC2 and ILC3 closely resemble their corresponding CD4+ helper T cell counterpart: Th1, Th2 and Th17 cells. ILC1 secrete IFN-γ and offer early protection against viral infections (Bernink et al., 2013; Fuchs et al., 2013; Weizman et al., 2017). ILC2 respond to parasitic helminths by producing type 2 cytokines such as IL-5 and IL-13 (Moro et al., 2010; Neill et al., 2010; Price et al., 2010). ILC3 secrete IL-22 in response to extracellular bacteria and play a role in the containment of intestinal commensal microbes (Cella et al., 2009; Luci et al., 2009; Satoh-Takayama et al., 2008).

Studies in mice have uncovered insights into the stages and transcription factor associated with ILC development. The transcription factor ID2 is crucial for the development of all ILCs including NK cell, but not B cells and T cells (Boos et al., 2007; Moro et al., 2010; Yokota et al., 1999), whereas GATA3 is required for 'helper' ILCs and LTi cells, not NK cells (Yagi et al., 2014). Finally, ZBTB16 is needed for the development of helper ILCs, but not LTi cells (Constantinides et al., 2014; Klose et al., 2014). Together, the transcription factors ID2, GATA3 and ZBTB16 delineate at least three progenitor stages which give rise to NK cells, LTi cells and the helper ILCs.

Individual ILC subsets are further driven by specific transcription factors, which parallel those of their corresponding helper T cell subsets. In addition to the transcription factors described earlier, NK cells require both Eomes and T-bet (Gordon et al., 2012), whereas ILC1s are dependent on the master transcription factor of Th1 cells, T-bet (Klose et al., 2014; Szabo et

al., 2000). On the other hand, the Th2 master regulator, GATA3, is necessary for ILC2 development and maintenance (Hoyler et al., 2012; Klein Wolterink et al., 2013; Mjösberg et al., 2012; Zheng and Flavell, 1997). ILC2s also require the activity of RORα although this is not the case for Th2 cells (Halim et al., 2012; Wong et al., 2012). A recent study in mice has shown that RORα expression in thymocytes can switch their developmental pathway from conventional T cells to ILC2s (Ferreira et al., 2021). Finally, both ILC3s and LTi cells are dependent on the Th17 transcription factor, RORγt, in mice (Eberl et al., 2004; Ivanov et al., 2006; Satoh-Takayama et al., 2008; Sun et al., 2000). Interestingly, humans deficient in *RORC* (which encodes for RORγt), still produce the IL-22+ subset of ILC3s, but not IL-17+ ILC3s (Lim et al., 2017), highlighting subtle differences between ILC development in human and mice.

1.2.6 Contribution of single cell sequencing to T cell biology

Many discoveries about T cell development and the diversity of T cell subtypes were derived from mouse models through genetic knockouts of crucial transcription factors. Immune cell populations were profiled by flow cytometry using a limited set of markers. The emergence of single cell transcriptomics and related technologies have contributed to the study of immunology, by delivering quantitative molecular readouts on a cell by cell basis (Ginhoux et al., 2022). Furthermore, transcriptomic profiling of single immune cells in the human body allows us to study human immunology in humans themselves, rather than extrapolating insights from mice (Behjati et al., 2018; Haniffa et al., 2021; Webb and Haniffa, 2023).

Single cell transcriptomics has uncovered heterogeneity in seemingly homogeneous populations of immune cells as well as overlapping features across seemingly distinct cell types. For instance, Rebuffet et al. (2024) showed that human NK cells could be delineated into three major groups and provided detailed descriptions about their gene expression profiles: NK1 cells showing the classic cytotoxic phenotype, NK2 cells exhibit higher protein synthesis and proliferative capacity, and NK3 cells possess features of "adaptive immunity". On the other hand, Jaeger et al. (2024) found that human ILC1s and NK cells lie on a transcriptional spectrum, with certain subsets bearing features of both cell types, unlike the clear distinctions observed in mice. Both of these studies highlight the heterogeneity and overlap among the NK cell and ILC1 populations in humans. In addition, the resolution of single cell methods, allows the identification of rare or unexpected cell populations for which markers are currently not known. Thomson et al. (2023) discovered a novel CD8αα T cell

subset in children that is poised for rapid immune response and is lost with age, providing some insights into the differences between children and adult immunity.

Single cell sequencing studies, now in the order of millions of cells, provide detailed descriptions of T cell development in humans. The landmark cellular atlas of the human thymus from fetal to adulthood by (Park et al., 2020) provided unbiased, quantitative transcriptomic definitions of the developmental stages of conventional T cells and also defined the transcriptomic features of unconventional T cells. Additionally, the authors identified two subsets of unconventional T cells which were fetal specific, namely NKT-like cells and Th17-like cells. A subsequent study characterising the single cell immune landscape across multiple organs in fetal development suggested that these cell types, which the authors renamed as type 1 and type 3 innate T cells, may originate from thymocyte-thymocyte interactions at the DP cell stage, rather than interactions between thymocytes and thymic epithelial cells (Suo et al., 2022).

Overall, single cell sequencing of human T cells and innate lymphocytes have provided a detailed descriptions of their developmental processes and cellular heterogeneity. They will serve as a valuable reference map for understanding T cell leukaemia in the context of T cell development.

1.3 T cell acute lymphoblastic leukaemia

Acute lymphoblastic leukaemia (ALL) is the most common childhood cancer (Hunger and Mullighan, 2015). ALL is a malignancy of developing lymphocytes and predominantly presents with symptoms of bone marrow failure, including the triad of anaemia, neutropenia and thrombocytopenia. Leukaemia cells, also known as lymphoblasts or simply blasts, can additionally infiltrate many other organs, including the liver, spleen, lymph nodes, mediastinum, testes and central nervous system, and perturb their normal function.

ALL is usually of B cell lineage (B-ALL), but can be of T cell lineage (T-ALL) in up to 15% of cases. T-ALL has an overall worse clinical prognosis than B-ALL, with higher rates of refractory disease, relapse and death (Goldberg et al., 2003). Moreover, there is currently no biological or genomic feature that is used to clinically risk stratify T-ALL at diagnosis, unlike B-ALL where clinical outcome correlates strongly with clinical features such as age and white cell count at presentation, as well as genomic mutations (Brady et al., 2022; Iacobucci et al., 2021; Roberts and Mullighan, 2020; Ueno et al., 2020).

ALL treatment comprises four phases: induction (4 weeks), consolidation (12 weeks), intensification (8 weeks) and long-term maintenance (2 years) (Malard and Mohty, 2020). Induction aims to achieve remission, defined as the eradication of leukaemia blasts and restoration of normal haematopoiesis. Induction utilises either three or four chemotherapy drugs: dexamethasone, PEGylated L-asparaginase, vincristine and additionally daunorubicin (for T-ALL and high-risk B-ALL) (Teachey and O'Connor, 2020). Induction failure, defined as the presence of ≥5% blasts in the bone marrow at the end of induction, is a predictor for poor clinical outcome and around 50% of such patients will eventually die from their leukaemia (O'Connor et al., 2017).

In this section, I will focus on T-ALL, for which two of the results chapters of my thesis are based upon. I begin with a review of earlier studies into the biology T-ALL. In the past 15 years, next-generation sequencing of T-ALL gave rise to detailed genomic classifications. Despite this progress, genomic studies of T-ALL have not been successfully translated into meaningful clinical risk stratification, unlike in B-ALL. In parallel, the early T cell precursor (ETP) immunophenotype was proposed as a distinct, high-risk T-ALL subtype. However, more recent studies have highlighted heterogeneity within this subtype and its unreliability as a clinical predictor. Finally, single cell sequencing has emerged as a powerful tool to investigate the cellular states of cancers and their microenvironment. I will consider how it can be applied to study leukaemias.

1.3.1 Early studies into the biology of T-ALL

Early work deciphering the biology and genetic drivers of T-ALL came from cytogenetics as well as mutational analysis of suspected cancer genes. About 50% of T-ALL cases contain structural chromosomal aberrations that are detectable by conventional karyotyping (Harrison and Foroni, 2002). These chromosomal rearrangements may place oncogenes under the influence of an active promoter/enhancer resulting in aberrant expression, or they may generate a fusion protein with constitutive activity. Many translocations recurrently involve the TCR loci on chromosome 7q34 (*TRB*), 14q11 (*TRA*/*TRD*) and 7p14 (*TRG*), thereby hijacking active enhancers of the TCR loci. Additionally, deletions have been found by fluorescence in situ hybridisation (FISH) in cases with seemingly normal karyotypes, usually leading to the loss of tumour suppressor genes. Sequencing of suspected driver genes in T-ALL with normal karyotype has uncovered point mutations and small insertions/deletions as additional sources of leukaemia drivers.

One of the first drivers discovered in T-ALL was *NOTCH1*, which was identified as a fusion partner to *TRB* in the rare t(7;9)(q34;q34.3) translocation that truncated the N-terminus leading to a constitutively active form of Notch1 in patients with T-ALL (Ellisen et al., 1991). Notch1 is both a cell surface membrane receptor and a ligand-activated transcription factor (Grabher et al., 2006). Ligand binding triggers a series of proteolytic cleavages, which releases intracellular Notch1 that translocates to the nucleus to regulate gene expression.

However, the t(7;9)(q34;q34.3) translocation involving *NOTCH1* is rare and only accounts for a minority of T-ALL cases. It was later found that in more than 50% of T-ALL, *NOTCH1* was activated by sequence mutations in the extracellular heterodimerisation (HD) domain or the C-terminal PEST domain (Weng et al., 2004). Mutations in the HD domain renders Notch1 more prone to ligand-independent activation whereas mutations in the PEST domain makes it more resistant to degradation. In addition, up to 30% of T-ALL also harbour mutations in *FBXW7*, which encodes a protein that mediates the proteasomal degradation of Notch1 (O'Neil et al., 2007; Thompson et al., 2007).

Although Notch1 signalling is essential for T cell development, its mechanistic role in T-ALL is less clear. The leukaemogenic effects of Notch1 may be due to the upregulation of the *MYC* oncogene, a direct transcriptional target of Notch1 (Palomero et al., 2006; Weng et al., 2006). Notably, Notch1 binds to a long-range distal enhancer of *MYC* which targeted by recurrent chromosomal duplications in 5% of T-ALL patients (Herranz et al., 2014). Given its ubiquity amongst T-ALL, it was previously thought that *NOTCH1* mutation could be a primary driver of leukaemogenesis. However, evidence suggests that *NOTCH1* alterations are subclonal and hence more likely to be secondary events (Mansour et al., 2007).

Deletions on chromosome 9p21 are one of the most common chromosomal anomalies found in T-ALL, with 65% and 15% of cases having homozygous and hemizygous deletions respectively (Cayuela et al., 1996; Hebert et al., 1994). These target the cell cycle regulators *CDKN2A* (p16/p14) and *CDKN2B* (p15), which inhibit the complex of cyclin D and CDK4/6 that keep the cell in a quiescent state. In fact, functional inactivation of the *CDKN2A* and *CDKN2B* loci is a nearly universal event in childhood T-ALL; this may be mediated by sequence mutation or promoter hypermethylation instead of gene deletions (Omura-Minamisawa et al., 2000).

The class II basic helix-loop-helix (bHLH) family of transcription factors are frequently mutated in T-ALL. These include *TAL1*, *TAL2*, *LYL1* and *OLIG2* (previously *BHLHB1*). Activation of

TAL1 may occur as a consequence of the t(1;14)(p32;q11) translocation to the *TRA*/*TRD* locus (Begley et al., 1989; Carroll et al., 1990), but more often results from a submicroscopic interstitial deletion at the chromosome 1p32 locus which generates the *STIL-TAL1* fusion gene (Janssen et al., 1993). More recently, sequence mutations at a precise site 1 kb upstream from the *TAL1* locus were found to create a super-enhancer site for binding of the MYB transcription factor, thereby driving aberrant *TAL1* expression (Mansour et al., 2014). *TAL2*, *LYL1* and *OLIG2* are also over-expressed in T-ALL, although much less common that *TAL1*, and these similarly involve translocation to place them adjacent to TCR enhancers (Mellentin et al., 1989; Wang et al., 2000; Xia et al., 1991).

The LIM domain only gene, *LMO1* and *LMO2*, which are located on chromosome 11p15 and 11p13, respectively, may be aberrantly expressed by translocation to the *TRA*/*TRD* locus on 14q11 (McGuire et al., 1989; Royer-Pokora et al., 1991). Additionally, *LMO2* may be activated by a cryptic deletion upstream of the gene that removes negative regulatory elements (Van Vlierberghe et al., 2006). The LMO proteins themselves do not bind to DNA directly, but instead form heterodimers with class II bHLH proteins such as TAL1 to promote leukaemogenesis (Aplan et al., 1997; Larson et al., 1996).

The homeobox (HOX) genes are key transcriptional regulators of embryonic development, responsible for processes as body patterning along the anterior-posterior axis (Shah and Sukumar, 2010). They are found in four HOX clusters (A-D) on different chromosomes in the human genome. Of these, the *HOXA* genes (A9, A10, A11, A13) on chromosome 7p15 are expressed in a coordinated fashion throughout T cell development (Taghon et al., 2003). Translocations between the *HOXA9* and *TRB* loci, resulting in aberrant expression of *HOXA9* and *HOXA10*, have been identified in T-ALL (Soulier et al., 2005; Speleman et al., 2005).

Whereas the classical HOX genes are found in gene clusters, orphan HOX genes are scattered throughout the genome. Two of these orphan HOX genes, *TLX1* (*HOX11*) and *TLX3* (*HOX11L2*), are relevant in T-ALL. The *TLX1* gene is activated in T-ALL by the t(10;14)(q24;q11) translocation which places it under control of the *TRA*/*TRD* locus (Hatano et al., 1991), whereas *TLX3* is over-expressed by placing it adjacent to the *BCL11B* enhancer in a t(5;14)(q35;q32) rearrangement (Bernard et al., 2001). Interestingly, TLX1 and TLX3 appear to reduce accessibility to the *TRA* locus and block TCR rearrangement there, resulting in a block in T cell maturation (Dadi et al., 2012).

1.3.2 Next generation sequencing of T-ALL

The arrival of massively parallel nucleotide sequencing, also known as next generation sequencing, revolutionised the study of T-ALL. Whole exome sequencing (WES) and transcriptome sequencing accelerated the discovery of many more cancer genes in T-ALL. Whole genome sequencing (WGS) additionally permits the discovery of driver mutations in the noncoding regions of the genome and is recently becoming more common than WES with the decreasing costs of sequencing. Finally, both WES and WGS allow the tracking of leukaemia clones across time using mutations as phylogenetic markers, although only WGS can deliver definitive phylogenies.

The first exome sequencing of T-ALL was in fact performed on just the X chromosome (Van Vlierberghe et al., 2010). Here, the authors identified *PHF6* as a recurrently mutated tumour suppressor gene, either from inactivating sequence mutations or focal deletions. Notably, *PHF6* mutations were found almost exclusively in males with T-ALL – present in 32% of males but only 2% of females – and was associated with mutations in *TLX1* and *TLX3*.

Systematic interrogation of mutations across all genes has led to the discovery of driver mutations in genes previously not considered to be cancer genes. For instance, mutations were found in the ribosomal protein genes, *RPL5* and *RPL10*, in 5-10% of T-ALL cases (De Keersmaecker et al., 2013). The authors demonstrated that recurrent R98S mutation on *RPL10* reduced ribosomal biogenesis and increased proliferation. One possible mechanism is that leukaemia cells with this mutation have greater translation of the anti-apoptotic BCL-2 protein via the IRES pathway (internal ribosome entry site) (Kampen et al., 2019), an alternative protein translation pathway that occurs during cellular stress. Interestingly, another group found that recurrent chromosome 6q deletion, whose role in T-ALL had been elusive for more than 40 years, may be ascribed to a reduction in ribosomal activity due to the loss of *SYNCRIP* and *SNHG5* (Gachet et al., 2018).

Driver mutations can also be identified from transcriptome sequencing, including structural rearrangements which generated a gene fusion that is transcriptionally expressed. Seki et al. (2017) identified novel recurrent gene fusions involving *SPI1* (*STMN1-SPI* and *TCF7-SPI1*) from transcriptome sequencing of children with T-ALL, making up around 4% of cases. These *SPI1* fusion positive blasts exhibited an immunophenotype and transcriptomic profile that was distinct from other known T-ALL subtypes. Critically, the *SPI1* fusion is the first genetic lesion that was found to be associated with very poor prognosis in T-ALL, with 6 out of 7 patients in their cohort dying from relapse within 3 years of diagnosis.

Whereas WES only covers coding regions which make up 1-2% of the human genome, WGS is able to capture mutations in noncoding regions, including long-range enhancer sites which alter gene expression. For instance, Montefiori et al. (2021) discovered a subtype of T-ALL with dysregulated expression of *BCL11B* due to translocation of the gene near super-enhancer sites which are active in immature haematopoietic stem and progenitor cells. Alternatively, it could also be driven by focal amplification of 2.5 kb region which contains a weak *BCL11B* enhancer.

The patterns of clonal evolution and driver of relapse may be deduced from WES and WGS. Sequencing of paired relapse and diagnostic samples reveal diverse patterns of clonal evolution pathways, indicating that relapse can derive from the major clone at diagnosis, the minor clone at diagnosis, be polyclonal in nature, or even a completely separate clone from diagnosis with mutually exclusive mutations (Oshima et al., 2016; Richter-Pechańska et al., 2022; Waanders et al., 2020).

T-ALL relapse can occur early or late. Early relapse is associated with the acquisition of mutations in chemotherapy resistance genes, such as *NT5C2*, which encodes an enzyme that inactivates the cytotoxic metabolites of the nucleoside-analog drug 6-mercaptopurine used for maintenance therapy (Tzoneva et al., 2013). In contrast, recent work on WGS of late relapse in T-ALL suggests that late relapse derives from a pre-leukaemic clone and lack chemotherapy resistance mutations (O'Connor et al., 2024). Most studies of clonal evolution in T-ALL have focused on comparing relapse to diagnosis. Little is known about the phylogenetic relationship of refractory T-ALL between post-treatment and diagnostic timepoints.

1.3.3 Systematic classification of T-ALL subtypes

T-ALL may be classified into subtypes which are parallel to stages of T cell development, suggesting the point at which a differentiation block occurred and the leukaemia diverged from normal development. These were originally based on a limited set of markers detected by flow cytometry ('immunophenotype'). Later approaches were based on gene expression profiles, initially on gene expression microarrays and subsequently by whole transcriptome sequencing.

The first classification scheme of T-ALL was proposed by the European Group for the Immunological Characterization of Leukemias (EGIL). According to EGIL, cytoplasmic or

surface membrane expression of CD3 antigen defines T-ALL, which can then be split into four subtypes: (TI) the most immature subtype or pro-T-ALL is defined by the expression of only CD7; (TII) pre-T-ALL additionally expresses CD2 and/or CD5 and/or CD8; (TIII) cortical T-ALL expresses CD1a; (TIV) mature T-ALL is positive for surface CD3 and negative for CD1a (Bene et al., 1995). Almost thirty years later, this set of markers is still routinely used in clinical flow cytometry. Further refinement of this classification scheme by considering TCR rearrangement status has been proposed although it is not used in clinical practice (Asnafi et al., 2003).

Gene expression profiling by microarray provided a framework to organise the burgeoning collection of translocations in T-ALL and generated several important insights (Ferrando et al., 2002). The TAL1 expressing leukaemias were associated with the expression of LMO1/2 and exhibited a late cortical DP cell phenotype ($CD3^+$, $CD8^+$, $CD4^+$). The TXL1 and TXL3 subtypes were highly similar, with both having an early cortical DP cell phenotype ($CD1a^+$, $CD3^-$, $CD8^+$, $CD4^+$). The LYL1 subtype co-expresses LMO2 and is associated with an immature phenotype ($CD34^+$, $CD3^-$, $CD8^-$, $CD4^-$). Finally, the HOXA expressing leukaemia comprises those with the *HOXA9-TRB* rearrangement, as well as the *KMT2A-MLLT1* (*MLL-ENL*) rearrangement (Chervinsky et al., 1995; Rubnitz et al., 1996), and was described to resemble γδ T cells. The classification by Ferrando et al. (2002) continues to define the major T-ALL subtypes today, linking genomic lesion to gene expression profile.

Modern sequencing studies of T-ALL cohorts in the past 10 years combine transcriptome sequencing with WES/WGS to define the landscape of T-ALL. The first of these by Liu et al. (2017) studied 264 children and young adults with T-ALL enrolled on the Children's Oncology Group (COG) AALL0434 trial between 2007 and 2011, for whom tumour and remission samples were available. They performed either WES or microarray genotyping on extracted DNA from tumour and remission samples, as well as transcriptome sequencing on tumour samples. From these data, the authors defined 8 T-ALL subtypes based on the dysregulated expression of key T-ALL initiating drivers: TAL1, TAL2, TLX1, TLX3, HOXA, LMO1/LMO2, LMO2/LYL1 and NKX2-1. Despite being the first to define the genomic landscape of T-ALL in such a large cohort, this study critically did not include patients who were refractory. To address this gap, O'Connor et al. (2023) performed WGS and transcriptome sequencing on 48 T-ALL patients who were refractory to initial treatment (i.e. induction failure). While certain drivers such as the *TAL1* noncoding mutation were relatively more common at induction failure, the authors did not find a unifying driver of refractory disease.

In a landmark study which was published in 2024, Pölönen et al. (2024) examined the largest T-ALL cohort of over 1,300 patients recruited on the COG AALL0434 trial, including those with

refractory disease. By combining transcriptome sequencing with WES and WGS, the authors delineated 15 T-ALL subtypes with their associated driver lesions: BCL11B, ETP-like, KMT2A, MLLT10, HOXA9 TCR, TLX3, TLX3, NKX2-1, TAL1 DP-like, TAL1 αβ-like, STAG2/LMO2, LMO2 γδ-like, NKX2-5, SPI1 and TME-enriched. Furthermore, they uncovered many new noncoding mutations and suggest noncoding mutations may provide the missing puzzle piece for the T-other subtype of T-ALL.

1.3.4 Early T cell precursor as a distinct T-ALL subtype

The only subtype of T-ALL to be recognised as a distinct entity since the 2017 World Health Organization (WHO) classification of leukaemias is the early T cell precursor (ETP) subtype. It is defined on flow cytometry as expressing markers of the earliest stages of T cell development (cytoplasmic $CD3^{+}$, TdT^{+}, $CD4^{-}$, $CD8^{-}$, $CD1a^{-}$, $CD5^{-/dim}$, MPO^{-}, $CD19^{-}$), and positive for either any myeloid or stem cell marker (CD11b, CD13, CD33, CD117, CD34, HLA-DR) (Arber et al., 2016). The discovery of the ETP subtype was made by Coustan-Smith et al. (2009), on the basis of unsupervised clustering of microarray expression profiles of T cell development genes from 55 T-ALL patients. From this, they identified a set of flow cytometry markers to define the ETP immunophenotype and demonstrated that patients with the ETP profile had a higher risk of treatment failure or relapse compared to non-ETP T-ALL.

Further evidence that ETP T-ALL may represent a distinct entity derives from its driver mutation landscape. ETP T-ALL was found to be enriched for activating mutations in genes regulating cytokine receptor and RAS signalling (*NRAS*, *KRAS*, *FLT3*, *IL7R*, *JAK3*, *JAK1*, *SH2B3*, *BRAF*), inactivating lesions disrupting haematopoietic development (*GATA3*, *ETV6*, *RUNX1*, *IKZF1*, *EP300*) and histone-modifying genes (*EZH2*, *EED*, *SUZ12*, *SETD2*, *EP300*), compared to non-ETP T-ALL (Zhang et al., 2012). Many of these genomic lesions were also later found to be enriched in mixed phenotype acute leukaemia with features of both T and myeloid lineages (Alexander et al., 2018).

Although ETP T-ALL was initially found to be associated with higher risk of treatment failure or relapse (Coustan-Smith et al., 2009), subsequent studies demonstrate this not to be the case (Patrick et al., 2014; Wood et al., 2023). The recent work by Pölönen et al. (2024) found the immunophenotypic definition of ETP to be suboptimal and instead proposed a new definition, referred to as ETP-like, from the clustering of more than 1,300 bulk transcriptomes of T-ALL. It was noted that only one-third of cases within the ETP-like category fulfilled the

flow cytometry definition of ETP. Whether ETP or ETP-like T-ALL represent a biologically meaningful and clinically significant leukaemia entity remains to be seen.

1.4 Single cell sequencing of leukaemias

Single cell sequencing allows us to generate high resolution quantitative molecular readouts of cancer cells and normal cells in leukaemia, which was previously not possible with bulk sequencing methods. These permits the precise elucidation of leukaemia cell states, reconstruction of leukaemia phylogenies and characterisation of the leukaemia microenvironment.

By interrogated individual leukaemia blasts, one can uncover distinct leukaemia cell states, which may be associated with treatment response. Using logistic regression to compare leukaemia blasts against various B cell developmental stages in the fetal bone marrow, Khabirova et al. (2022) showed that *KMT2A*-rearranged infant ALL, which is a highly aggressive subtype, resembles early lymphoid progenitors, whereas *NUTM1*-rearranged infant ALL, which has favourable outcome, resembles later stages of B cell development. The distinct transcriptional cell state of *KMT2A*-rearranged ALL has also been demonstrated by Chen et al. (2022). In addition, the analysis of transcriptional networks from single cell multi-omic sequencing of B-ALL may identify suitable pharmacological targets (Huang et al., 2024).

Bulk WES and WGS will identify the major clone and a few minor clones which are large enough. In contrast, single cell sequencing of leukaemia DNA can identify much smaller clones and reveal the true nature of the leukaemia phylogeny. For instance, single cell targeted DNA sequencing of T-ALL revealed several minor clones with distinct *NOTCH1* mutations in the same sample, indicating strong selective pressures at this gene (Albertí-Servera et al., 2021). The tracing of clones throughout treatment may identify the source of relapse such as the presence of pre-existing CD19-negative blasts which gave rise to relapse following anti-CD19 treatment in B-ALL (Rabilloud et al., 2021).

The success of cancer immunotherapy has generated immense interest in studying the tumour microenvironment. It was initially thought that only solid tumours with high mutational burdens would generate sufficient neoantigens to respond to immunotherapy. However, Zamora et al. (2019) show that even in paediatric ALL, which has one of the lowest mutation burden, there are abundant neoantigens which are capable of activating CD8 T cells. These results were further supported by Wang et al. (2021) found clusters of exhausted T cells from single cell

transcriptome profiling of three B-ALL samples at diagnosis. Components of the immune system other than T cells make also play role. For instance, the enrichment of CD16+ non-classical monocytes in the bone marrow was associated with poorer survival in both paediatric and adult B-ALL (Witkowski et al., 2020).

1.5 Outline of this dissertation

The central focus of my PhD is investigating the transcriptomic features and origin of refractory T-ALL. Through this introduction, I illustrated how our understanding of T cell development has evolved radically especially in the past two decades. I highlighted the growth of T-ALL research from the early days of cytogenetics to contemporary sequencing at base-pair resolution and whole genome coverage. Despite being a well-studied cancer, no biological or genomic marker has successfully made it to the clinic for risk stratification. During my PhD I applied single cell transcriptome sequencing to define the transcriptomic features of refractory T-ALL and combined it with WGS to examine the origin of refractory T-ALL. In addition, I characterised the bone marrow microenvironment of T-ALL at a single cell level.

The remainder of this dissertation is structured as follows:

Chapter 2: Materials and methods. This chapter provides details on how I analysed data from single cell RNA sequencing (scRNA-seq), single cell TCR sequencing (scTCR-seq), bulk RNA sequencing (bulk RNA-seq) and bulk whole genome sequencing (WGS) of T-ALL.

Chapter 3: A single cell transcriptome atlas of T-ALL. This chapter presents a detailed characterisation of leukaemia blasts and normal cells in bone marrow aspirates (and other tissue samples) from patients with T-ALL. I compared leukaemia transcriptomes to various stages of T cell development. I annotated the normal cell compartment to a high resolution and identified various cell types. Finally, I examined for differences in bone marrow cell composition, between patients with induction failure and those who were responsive.

Chapter 4: Transcriptomic features of refractory T-ALL. In this chapter, I discovered that refractory T-ALL is driven by the presence of *ZBTB16*+ blasts. I validated the association between *ZBTB16* expression and treatment responsive across bulk transcriptome sequencing datasets. Finally, I showed that *ZBTB16*+ blasts are a distinct leukaemia cell state with features of innate-like lymphocytes as opposed to stages of conventional T cell development, and express unique cell surface targets for cellular therapy.

Chapter 5: Origin of refractory blasts in T-ALL. This chapter examines the origin of the refractory *ZBTB16*+ blasts identified in the previous chapter. I considered if refractory *ZBTB16*+ blasts were associated with particular genomic subtypes of T-ALL or with specific TCR rearrangement states. I investigated if *ZBTB16* expression might be driven by noncoding mutations and result in allele specific expression. Finally, I projected WGS-derived somatic mutations onto scRNA-seq for phylogenetic construction of leukaemia across treatment timepoints to identify the origin of refractory clones.

Chapter 6: Conclusion. This chapter concludes the thesis with overall insights gained from analyses undertaken in this work.

Chapter 2: Materials and methods

2.1 Sample acquisition and ethics statement

The tissues used in my doctoral studies were accessed with approval by UK NHS research ethics committees. The tissue sources were: VIVO Biobank (National Research Ethics Service reference 16SW0219; VIVO project number 23-VIVO-17), Great Ormond Street Haematology Cell Bank (National Research Ethics Service reference 16/LO/0960), and samples from the diagnostic archives of Great Ormond Street Hospital (GOSH) (National Research Ethics Service reference 16/EE/0394). Patients or guardians provided informed written consent for participation in this study as stipulated by the study protocols.

Two cohorts of patients with T-ALL were used for my dissertation work. The first cohort ("discovery cohort") consists of 21 children, assembled by Dr David O'Connor from GOSH. A total of 29 bone marrow aspirates were obtained, comprising samples at diagnosis ("day 0") for all 21 children as well as samples at the end of induction treatment ("day 28"). The second cohort ("extension cohort") consists of 37 children, assembled by Dr David O'Connor, Dr Jack Bartram and Dr Angus Hodder. A total of 55 samples were available for analysis in this cohort, including at diagnosis, day 7 and day 28 of induction, and at relapse. These samples were mainly bone marrow aspirates, but also include peripheral blood and pleural fluid.

2.2 Single cell RNA sequencing and single cell TCR sequencing

Cells were retrieved from either fresh cell suspensions or viably frozen cell suspensions, and enriched for viable cells using the Dead Cell Removal Kit (Miltenyi) as per manufacturer's instructions. Single cell droplet suspensions were prepared using the 10x Chromium Controller (10x Genomics), followed by reverse transcription and cDNA amplification using either the Single Cell 5' PE kit (10x Genomics) for the discovery cohort or the Single Cell 5' R2-only kit (10x Genomics) for the extension cohort to generate single cell RNA sequencing (scRNA-seq) libraries. T cell receptor (TCR) sequences for single cell TCR sequencing (scTCR-seq) were amplified using the Chromium Single Cell Human TCR Amplification kit (10x Genomics). Libraries were sequenced on the NovaSeq 6000 platform. Sample processing, library preparation and sequencing were performed at University College London for the discovery cohort, and at the Wellcome Sanger Institute for the extension cohort.

2.3 Alignment, quantification and quality control of single cell RNA sequencing

Sequencing reads were processed using Cell Ranger (v7.0.0) (Zheng et al., 2017), where reads were aligned to the GRCh38 human reference genome to generate a matrix of gene expression counts per cell. Ambient mRNA contamination was removed with SoupX (v1.6.2) (Young and Behjati, 2020), which estimates ambient mRNA counts from empty droplets and subtracts this background noise from the expression matrix. Cells were filtered to retain only high quality cells, defined as containing >300 expressed genes, >1000 total read count and <15% reads mapping to mitochondrial genes. Finally, droplets containing more than one cell (doublets) were identified and removed using Scrublet (v0.2.3) (Wolock et al., 2019)

2.4 Normalisation, clustering and annotation of single cell RNA sequencing

Single cell gene expression count matrices were processed using the standard processing pipeline in Scanpy (v1.10.0) (Wolf et al., 2018). Gene expression counts were normalised to 10,000 total counts within each cell and log-transformed. This was followed by selection of highly variable genes, principal component analysis (PCA), uniform manifold approximation and projection (UMAP) and clustering using the Leiden algorithm. Batch correction was not performed to ensure that biological variation between individual leukaemias was preserved. Cell clusters were annotated based on their expression of marker genes of haematopoietic cell types and T-ALL lymphoblasts. Cell cycle phase (G1, S, G2/M) of cells were determined by scoring previously defined S phase and G2/M phase gene sets (Tirosh et al., 2016)

2.5 Comparison of leukaemia and normal cell transcriptomes by logistic regression

Leukaemia cell transcriptomes from scRNA-seq of T-ALL were compared to normal cell transcriptomes from a reference cell atlas of the developing fetal immune system (Suo et al., 2022), using a previously described logistic regression analysis framework (Young et al., 2018). First, cells from the reference cell atlas were randomly split, such that 80% of cells of each cell annotation label were used for model training and the remaining 20% of cells were used for evaluating the trained model. To speed up model training, up to 400 cells for each cell annotation label were randomly sampled. Similarity scores between 0 and 1 of how much each leukaemia cell resembles the various cell annotations from the reference cell atlas were calculated using the trained logistic regression model and the reference cell annotation with the highest similarity score was assigned as the ‘best match cell type’ to that leukaemia cell.

A modified version of this logistic regression analysis was used to compare transcriptomes of normal cells from the fetal immune atlas (Suo et al., 2022) to *ZBTB16*+ leukaemia cells, *ZBTB16*− leukaemia cells and Schwann cell precursors (non-haematopoietic cell negative control). In this case, leukaemia cells and Schwann cell precursors are used to train the logistic regression model.

2.6 Detailed annotation of normal cell types in T-ALL

Normal cells (i.e. cells which are not annotated as leukaemia blasts) from the discovery and extension cohorts were combined and then split into: (1) T cells and NK cells, (2) B cells and plasma cells, and (3) myeloid and erythroid cells. Each of these three groups of normal cells were separately annotated to the highest possible resolution (level 3 cell annotation), using known marker genes of haematopoietic cell types. Cells expressing a highly improbable combination of marker genes for distinct cell lineages (e.g. *CD19* for B cells and *HBA1* for erythroid) were annotated as doublets. Cell annotations of lower resolution (level 1 and 2) were generated by combining closely related cell types under the same label.

2.7 Cell proportion analysis of normal cell types in T-ALL

Cell proportions were calculated for each normal cell type as a percentage of the total number of normal cells (excluding leukaemia blasts and doublets). Significance testing for differences in cell proportions was performed using the *speckle::propeller()* function (Phipson et al., 2022) and *P* values were adjusted by Benjamini-Hochberg correction for multiple hypothesis testing. Cell proportion analysis was performed on all three levels of cell annotation.

2.8 Differential expression analysis to identify marker genes of normal cell types

Marker genes of normal cell types were identified using the *scanpy.tl.rank_genes_groups()* function (Wolf et al., 2018), which performs a differential expression analysis between a group of query cells against a group of reference cells, using a Wilcoxon rank-sum test with Benjamini-Hochberg correction for multiple hypothesis testing.

2.9 Alignment and quantification of bulk RNA sequencing

Bulk RNA sequencing was performed on diagnostic samples for 19 out of 21 patients in our discovery cohort (not available for P073 and P069). Sequencing reads were aligned to the GRCh38 reference human genome using STAR (v2.7.10) (Dobin et al., 2013) and quantified with featureCounts (v2.0.2) (Liao et al., 2014) to generate a matrix of gene expression counts by sample.

2.10 Analysis of bulk RNA sequencing gene expression data

Bulk gene expression count matrices from three T-ALL cohorts were analysed:

(1) Our discovery cohort with 19 patients
(2) An unpublished cohort of 52 patients from the Princess Maxima Center (PMC)
(3) A published cohort of 1,335 patients from the Children's Oncology Group (COG) AALL0434 trial (Pölönen et al., 2024)

Bulk gene expression count matrices were processed using the standard pipeline in edgeR (v3.42.4) (Robinson et al., 2010) and gene module scoring of bulk RNA sequencing samples was performed using singscore (v1.20) (Foroutan et al., 2018).

For bulk RNA sequencing of the discovery cohort, processing steps involved removing genes without a corresponding HGNC gene name, filtering out genes which are lowly expressed using the *filterByExpr()* function in edgeR, library-size normalisation and log-transformation to convert gene expression into log2 count per million (logCPM) values.

For the PMC cohort, gene expression counts were processed into logCPM values with edgeR as described above for the discovery cohort, except that genes were retained only if they achieved an average logCPM of 0.5 across at least 10% of samples.

For the COG AALL0434 cohort, the non-batch corrected count matrix was downloaded from the Synapse portal (https://www.synapse.org/Synapse:syn54032669/wiki/627818; accessed September 2024). Only samples with >60% blast content were used for subsequent analysis. The gene expression counts were processed into logCPM values with edgeR as described above for the discovery cohort, except that genes were retained only if they had an average logCPM of 0.5 in at least 10% of samples.

2.11 Survival analysis on the COG AALL0434 cohort

Survival analysis was performed on the COG AALL0434 cohort (Pölönen et al., 2024) using the survival R package (v3.5) (Therneau, 2024). Kaplan-Meier curves were plotted using overall survival and event free survival data from the authors. Patients were grouped into the top and bottom 33%, according to either *ZBTB16* expression or the *ZBTB16*-dervied gene module score. Cox proportional hazard models were used to compare the predictive effects of multiple variables on overall survival and event free survival in this cohort.

2.12 Drug sensitivity analysis in T-ALL

Normalised LC50 values (lethal concentration to kill 50% of leukaemia blasts) were obtained from a published dataset of drug sensitivity in primary leukaemia blasts (Lee et al., 2023). The concomitant bulk RNA-seq data was downloaded as a matrix of FPKM values (fragments per kilobase of transcript per million mapped reads) (http://permalinks.stjude.cloud/permalinks/all-pharmacotype; accessed February 2026). The bulk RNA-seq data was processed using edgeR (v3.42.4) (Robinson et al., 2010), including log-transformation to logFPKM, and the *ZBTB16*-derived gene module was scored using singscore (v1.20) (Foroutan et al., 2018).

2.13 Differential expression analysis to identify cell surface targets of T-ALL

Cell surface targets against T-ALL blasts were obtained by pseudobulk differential expression analysis comparing T-ALL blasts against normal T cells in my scRNA-seq data. Pseudobulk analysis was performed by aggregating all leukaemia blasts and all normal T cells by patient. Differential expression analysis was performed using the edgeR package (Robinson et al., 2010) and differentially expressed genes were identified as those with an adjusted *P* value <0.05 after Benjamini-Hochberg correction for multiple hypothesis testing. Differentially upregulated genes in leukaemia blasts were then subjected to the following filters:

(1) Expressed in >15% of leukaemia blasts
(2) Expressed in <10% of at least 10 subtypes of conventional T cells from a published normal immune cell atlas (Domínguez Conde et al., 2022)
(3) Annotated as cell surface proteins in Gene Ontology (under GO:0009897) (Ashburner et al., 2000; The Gene Ontology Consortium et al., 2023)

Additionally, pseudobulk differential expression analysis was similarly used to identify cell surface targets against refractory *ZBTB16*+ T-ALL blasts by comparing *ZBTB16*+ blasts from induction failure cases against normal T cells. The resulting list of genes were subjected to the following filters:

(1) Expressed in >10% of *ZBTB16*+ leukaemia blasts from induction failure cases
(2) Expressed in <15% of all leukaemia blasts from responsive cases
(3) Expressed in <10% of at least 10 subtypes of conventional T cells from a published normal immune cell atlas (Domínguez Conde et al., 2022)
(4) Annotated as cell surface proteins in Gene Ontology (under GO:0009897) (Ashburner et al., 2000; The Gene Ontology Consortium et al., 2023)

2.14 Whole genome sequencing and variant calling

Short-insert (500-bp) genomic libraries were constructed and 150-bp paired-end sequencing clusters were generated on the NovaSeq 6000 platform using standard PCR library generation protocol. Sequencing reads were aligned to the GRCh38 Ensembl 103 reference genome using the Burrows-Wheeler Alignment tool (v0.7.17) (Li and Durbin, 2009).

All classes of somatic variants were called from whole genome sequencing (WGS) using the extensively validated pipeline of the Wellcome Sanger Institute, built on the following algorithms: CaVEMan (v1.18.2) (Jones et al., 2016) for single nucleotide variants (SNV), Pindel (v3.10.0) (Ye et al., 2009) for insertions/deletions, ASCAT (v4.5.0) (Van Loo et al., 2010) and Battenberg (v3.5.3) (Nik-Zainal et al., 2012) for copy number alterations, and BRASS (v6.3.4) (Nik-Zainal et al., 2016) and GRIDSS2 (v2.13.1) (Cameron et al., 2021) for structural variants. Copy number alterations in samples which did not have a matched germline sample were called using PURPLE (v3.8.4) (Priestley et al., 2019).

2.15 Alignment and annotation of T cell receptor sequences

TCR gene usage was called using two different methods. In the first method, raw reads from scTCR-seq were aligned to the Cell Ranger VDJ reference (v7.0.0), using Cell Ranger (v7.0.0) (Zheng et al., 2017), generating a FASTA file of TCR sequences. The TCR sequences were then annotated by Dandelion (v0.3.6) (Suo et al., 2024) to give VDJ gene usage calls.

Since scTCR-seq only captures TCR αβ regions (*TRA*/*TRB*), TRUST4 (v1.1.0) (Song et al., 2021), was used instead to recover TCR γδ regions (*TRG*/*TRD*), as well as validate the *TRA*/*TRB* calls from Dandelion. TRUST4 extracts candidate reads from the scRNA-seq BAM file, assembles them into TCR sequences and annotates VDJ gene calls.

2.16 Detecting allele specific expression in bulk RNA sequencing

Allele specific expression of genes was detected in bulk RNA-seq using a modified pipeline of alleleIntegrator (v0.9.1) (Trinh et al., 2022). Heterozygous single nucleotide polymorphisms (SNPs) were identified from WGS of the remission sample. A pileup was performed at these loci on bulk RNA-seq, counting the number of reads supporting the reference and alternate allele, and the B allele frequency (BAF) was calculated. SNPs that significantly deviated from BAF = 0.5 were identified using a two-sided exact binomial test with Benjamini-Hochberg multiple hypothesis correction (false discovery rate of 0.01).

2.17 Detecting copy number alterations in single cell RNA sequencing

Detection of copy number alterations in single leukaemia blast transcriptomes was performed using alleleIntegrator (v0.9.1) (Trinh et al., 2022). Heterozygous SNPs were identified from WGS of its remission sample. Phasing of heterozygous SNPs across copy number segments with allelic imbalance was performed, using WGS of the high tumour purity sample at day 0. The number of reads supporting the major/minor allele across each copy number segment was calculated in each single cell transcriptome. Finally, the posterior probability of each cell harbouring the altered or normal copy number state for each copy number segment was calculated. In the case of P030 where no remission sample is available, heterozygous SNPs were called from the day 28 sample which had a lower tumour purity, by setting the B allele frequency (BAF) deviation parameter to 0.3 instead of the 0.1 default value.

2.18 Reconstructing leukaemia phylogeny from single nucleotide variants

SNVs were called using CaVEMan, as described earlier. For day 28 samples with low tumour purity, CaVEMan was unable to call somatic SNVs using the default normal contamination parameter. To overcome this, CaVEMan was run with a new normal contamination parameter, set by estimating the day 28 tumour purity using SNVs called in the day 0 sample (**Table 2.1**). Following the standard post-processing filters of CaVEMan, SNVs were filtered for those with median alignment score of reads supporting a mutation (ASMD) ≥140 and fewer than half of reads supporting the variant were clipped (CLPM = 0).

Table 2.1: CaVEMan analysis type and normal contamination parameter.

Patient ID	Timepoint	WGS sample ID	Analysis type	Normal contamination
P058	Day 0	PD62354a	Matched	0.1 (default for matched)
P058	Day 28	PD61982a	Matched	0.95
P030	Day 0	PD67564a	Unmatched	0.0 (default for unmatched)
P030	Day 28	PD67564c	Unmatched	0.8

In the case of P030 where no remission sample is available, there was an additional step of removing germline SNVs by a one-sided exact binomial test. At each SNV locus, the number of trials was the total read depth, and the number of successes was the variant read count. The null hypothesis is a binomial distribution with $P = 0.5$ (chromosomes 1 to 22) or $P = 0.95$ (chromosomes X and Y as P030 is male), while the one-sided alternative hypothesis is a distribution with $P < 0.5$ or $P < 0.95$, respectively. SNVs with adjusted P value $<10^{-5}$ following Benjamini-Hochberg correction for multiple hypothesis testing were considered somatic, while the rest were considered germline and removed. SNVs found in copy number altered regions were also removed as the exact binomial test is invalid in these cases.

SNVs from day 0 and day 28 samples were merged to form a union catalogue for that patient. For each locus in this catalogue, pileups were generated across all samples from that patient, with read filters of mapping quality (mq = 30) and base quality (bq = 25), to determine the total read depth and number of variant supporting reads. SNVs with too low or too high read depth (minimum and maximum cutoff values for each sample in **Table 2.2**) were removed as these

were likely due to sequencing or mapping artefacts. SNVs found within 10 base pairs from any insertion/deletion called by Pindel (including non-PASS ones) were also removed, as insertion/deletion mutations could be misinterpreted by variant callers as SNVs.

Table 2.2: Depth filter cutoff values.

Patient ID	Timepoint	WGS sample ID	Average coverage	Minimum depth	Maximum depth
P058	Day 0	PD62354a	107x	20	150
P058	Day 28	PD61982a	429x	150	550
P030	Day 0	PD67564a	39x	15	60
P030	Day 28	PD67564c	218x	100	300

Furthermore, site-specific error rates were calculated by interrogating the same sites in 32 unmatched normal blood samples from a previous study (Coorens et al., 2021), and SNVs which were indistinguishable from background noise (false-discovery rate of 0.01 following Benjamini-Hochberg multiple hypothesis correction) were rejected (Gerstung et al., 2014). All SNVs post-filtering were visually inspected on the IGV genome browser (Robinson et al., 2011), yielding a final catalogue of high quality variants.

Each SNV has an associated variant allele frequency (VAF), which is the number of reads supporting the variant divided by the total of reads at that locus, determined from the pileup. For phylogenetic reconstruction, the cancer cell fraction (CCF) of each SNV, that is the fraction of cancer cells with the mutation, is calculated below, by considering the VAF, copy number profile and tumour purity (Tarabichi et al., 2021):

$$\mathrm{CCF} = \mathrm{VAF} \times \frac{pN_{tum,total} + (1-p)N_{norm,total}}{mp}$$

where p is the inferred tumour purity, $N_{\mathrm{tum,total}}$ is the integer tumour total copy number, $N_{\mathrm{norm,total}}$ is the integer normal total copy number and m is the inferred multiplicity of the mutation. Tumour purity is a parameter inferred from the CCF distribution, such that the CCF values of clonal mutations (found in all cancer cells) are centred around 1 (**Table 2.3**). Multiplicity is an integer with a default value of 1 and refers to the number of chromosome copies bearing the

mutation. If an SNV falls within a duplicated region of the genome (either a copy number gain or a copy-neutral loss of heterozygosity), then there is a possibility that the duplication event occurred after the SNV event and there will be two copies of the SNV instead of one. In these cases, multiplicity is inferred from the calculated CCF: if CCF > 1.5, then multiplicity will be set to a value of 2. Phylogenetic reconstruction of subclones (defined as having CCF < 1) was performed by clustering SNVs according to their CCF at day 0 and day 28, using DPClust (Dentro et al., 2017), which models subclones by a Dirichlet Process.

Table 2.3: Tumour purity parameters.

Patient ID	Timepoint	WGS sample ID	Blast percent from flow cytometry	Tumour purity parameter
P058	Day 0	PD62354a	88%	0.84
P058	Day 28	PD61982a	9%	0.05
P030	Day 0	PD67564a	80%	0.92
P030	Day 28	PD67564c	30%	0.22

2.19 Detecting single nucleotide variants in single cell RNA sequencing

SNVs are detected in the scRNA-seq BAM files by running the *mpileup* function in the SAMtools package (Danecek et al., 2021). This outputs the allelic state for each cellular barcode and each locus. Thus, each cell can be annotated whether it contains the variant allele or the reference allele only, or if there is no coverage at all at that locus and for that cell. In fact, for majority of loci and majority of cells, there will be no coverage on scRNA-seq. This is because only expressed regions of the genome are covered, and this is further limited to the 5' end of transcripts due to the sequencing chemistry used in this study. This limitation is partially overcome by aggregating coverage across SNVs assigned to the same cancer subclone.

Chapter 3: A single cell transcriptome atlas of T-ALL

3.1 Introduction

Acute lymphoblastic leukaemia (ALL) occurs when there is rapid uncontrolled proliferation of malignant lymphocyte progenitors (lymphoblasts or simply blasts), and presents as either B cell lineage (B-ALL) or T cell lineage (T-ALL). T-ALL can be further characterised and classified by its protein marker expression on flow cytometry (immunophenotype), by its mutation status of key leukaemia genes (genomic subtype), or by its expression of gene transcripts (transcriptomic subtype).

Flow cytometry profiling is often used to compare T-ALL lymphoblasts to stages of T cell development and has identified the early T cell precursor (ETP) phenotype (Coustan-Smith et al., 2009), which has been recognised as distinct clinical entity since the 2017 World Health Organization (WHO) classification (Arber et al., 2016). Although flow cytometry provides readouts at single cell resolution, it is limited to a set of around 20-30 markers which must be defined beforehand and have available antibodies.

In contrast, transcriptome sequencing allows one to quantify in an unbiased manner the expression levels of >20,000 genes in the human genome. When combined with genome sequencing, this has provided detailed classification landscapes of T-ALL subtypes, relating genetic mutations to gene expression phenotypes (Brady et al., 2022; Liu et al., 2017; O'Connor et al., 2023; Pölönen et al., 2024). However, majority of transcriptome and genome sequencing studies are performed at the bulk level, where signals are aggregated across leukaemia cells and often contaminated by signals from admixed normal cells (typically non-malignant haematopoietic cells in bone marrow or blood sample).

Single cell RNA sequencing (scRNA-seq) combines the single cell readout advantage of flow cytometry with the unbiased whole transcriptome coverage of bulk RNA sequencing. The greatly enriched gene expression signals from leukaemia blasts can be used to define leukaemia cell states and compare in an unbiased, whole transcriptome fashion to cell types in previously published single cell atlases. Through this approach, the clinically aggressive *KMT2A*-rearrange infant ALL was found to resemble early lymphoid progenitors, unlike other less aggressive infant ALL subtypes (Khabirova et al., 2022). Moreover, scRNA-seq revealed heterogenous leukaemia cell populations in *KMT2A*-rearranged ALL, including a highly immature leukaemia cell state associated with poorer prognosis (Chen et al., 2022).

One could also profile the composition of normal cells in the leukaemia microenvironment using scRNA-seq. For instance, scRNA-seq found evidence of exhausted effector T cells in B-ALL, dispelling the myth that ALL is an 'immunological cold' tumour (Wang et al., 2021). In another scRNA-seq study of B-ALL, leukaemia cells were found to remodel the monocyte compartment in the bone marrow (Witkowski et al., 2020). Therefore, scRNA-seq opens up new avenues of analyses with the potential of discovering novel leukaemia biology.

In this chapter, I generated and analysed an scRNA-seq atlas of 84 samples from 55 patients with T-ALL. I performed detailed annotation of the 378,926 leukaemia blasts and 171,068 normal cells, which add up to a total of 549,994 cells. I revealed interesting insights into the leukaemia cell states of T-ALL and the normal cell composition in the leukaemia microenvironment.

3.2 Single cell transcriptome sequencing of an initial cohort of T-ALL

I started by analysing an initial cohort of 21 children with T-ALL ("discovery cohort"), which was assembled by Dr David O'Connor from Great Ormond Street Hospital (GOSH). This cohort consisted of both children who responded to induction treatment (n = 8) as well as children with induction failure (n = 13), and represented the typical spectrum of T-ALL genomic subtypes and ETP status (**Table 3.1**). Diagnostic bone marrow aspirates ("day 0") were collected from all 21 children. Post-induction bone marrow aspirates ("day 28") were obtained from 8 children with induction failure. All samples were subjected to combined single cell RNA sequencing (scRNA-seq) and single cell TCR sequencing (scTCR-seq). Enrichment for blasts was not performed to avoid biases associated with cell selection.

Following standard quality control and processing steps of scRNA-seq data with the Scanpy package (Wolf et al., 2018), as described earlier in my methods chapter, I obtained 216,507 high quality cells in total (**Figure 3.1A-B**). The number of high quality cells per scRNA-seq sample ranged from 459 cells to 16,753 cells, with 27 out of 29 samples having at least 1,000 high quality cells. These cells were then clustered on a uniform manifold and projection (UMAP) and annotated based on the expression of known marker genes (**Figure 3.2A-C**). Importantly, leukaemia blasts were distinguished from normal haematopoietic cells. The percentages of blasts within each scRNA-seq sample were broadly consistent with values obtained by clinical flow cytometry (**Figure 3.2B**).

Table 3.1: Clinical metadata and sample availability of discovery cohort.

ID	Sex	Age at diagnosis	ETP status	Genomic subtype	Induction response	Timepoints with scRNA-seq data
P016	M	11-15 years	nonETP	TAL1	IF	D0, D28
P018	F	1-5 years	nonETP	MYC-TRA	IF	D0, D28
P019	F	11-15 years	ETP	HOXA	IF	D0
P021	F	11-15 years	ETP	T-other	IF	D0, D28
P022	M	11-15 years	nonETP	TAL1	IF	D0
P026	M	6-10 years	ETP	MED12	IF	D0
P028	F	11-15 years	nonETP	KMT2A	IF	D0, D28
P029	M	6-10 years	nonETP	TAL1	IF	D0, D28
P030	M	1-5 years	nonETP	TAL1	IF	D0, D28
P058	M	11-15 years	nonETP	TAL1	IF	D0, D28
P069	M	11-15 years	nonETP	T-other	Responsive	D0
P071	F	6-10 years	nonETP	TLX3	Responsive	D0
P072	M	>15 years	nonETP	TAL1	Responsive	D0
P073	M	6-10 years	nonETP	TAL1	Responsive	D0
P074	M	1-5 years	nonETP	TLX1	Responsive	D0
P076	M	1-5 years	nonETP	NUP98	Responsive	D0
P080	M	11-15 years	nonETP	MED12	IF	D0
P084	M	11-15 years	nonETP	MLLT10	IF	D0, D28
P086	F	1-5 years	ETP	KMT2A	IF	D0
P108	M	1-5 years	nonETP	T-other	Responsive	D0
P112	F	6-10 years	nonETP	T-other	Responsive	D0

ETP, early T cell precursor; IF, induction failure.

Figure 3.1

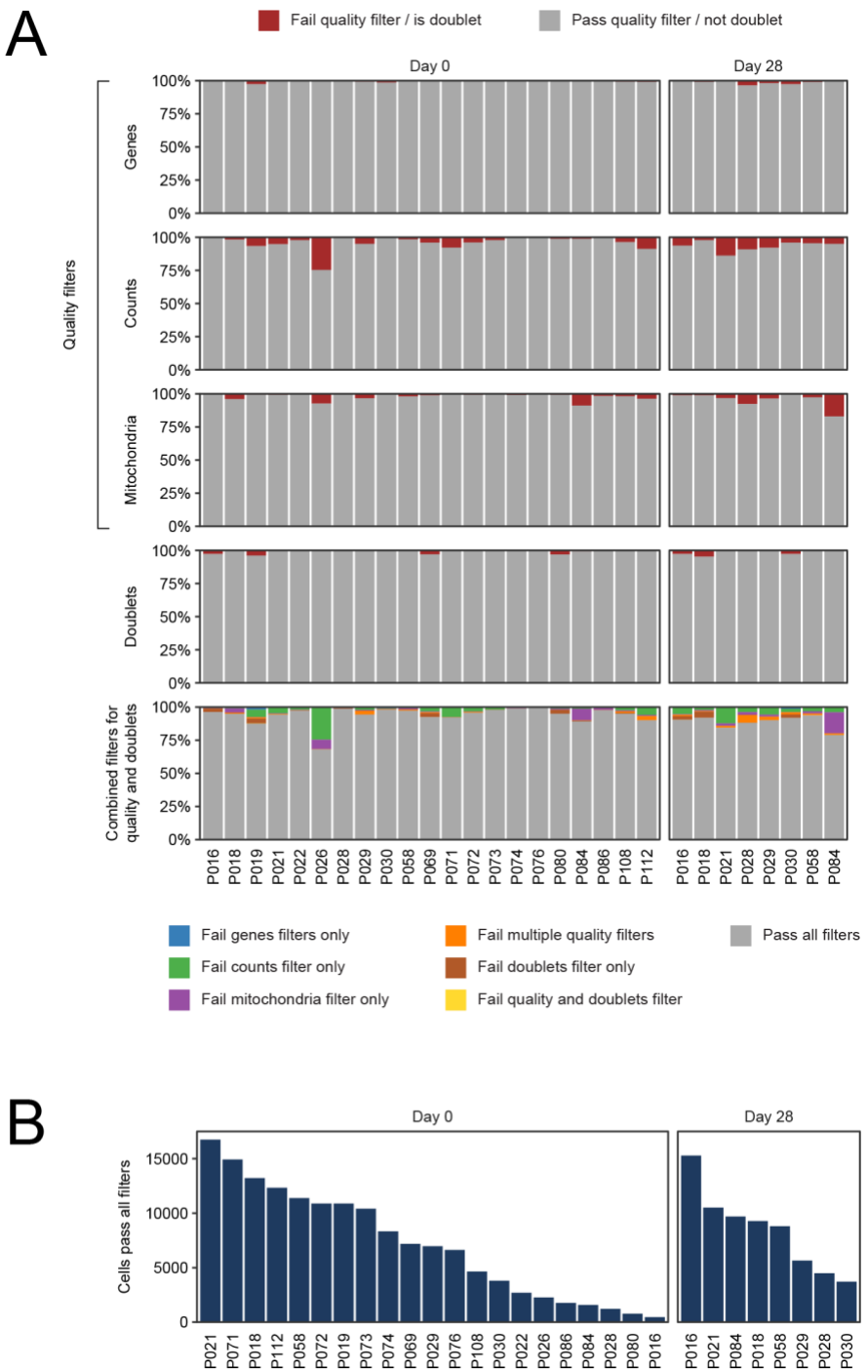

A
Fail quality filter / is doublet
Pass quality filter / not doublet
Day 0
Day 28
Quality filters
Genes
Counts
Mitochondria
Doublets
Combined filters for quality and doublets
100%
75%
50%
25%
0%
P016 P018 P019 P021 P022 P026 P028 P029 P030 P058 P069 P071 P072 P073 P074 P076 P080 P084 P086 P108 P112
P016 P018 P021 P028 P029 P030 P058 P084
Fail genes filters only
Fail counts filter only
Fail mitochondria filter only
Fail multiple quality filters
Fail doublets filter only
Fail quality and doublets filter
Pass all filters
B
Day 0
Day 28
Cells pass all filters
15000
10000
5000
0
P021 P071 P018 P112 P058 P072 P019 P073 P074 P069 P029 P076 P108 P030 P022 P026 P086 P084 P028 P080 P016
P016 P021 P084 P018 P058 P029 P028 P030

Figure 3.1: Quality control of discovery cohort scRNA-seq.

(**A**) Bar plots showing percentage of cells for each scRNA-seq sample that passed each filter. Cell quality filters include: genes filter (top panel, keep cells with >300 expressed genes), counts filter (second panel, keep cells with >1000 total RNA counts) and mitochondria filter (third panel, keep cells with <15% of RNA counts mapping to mitochondrial genes). Additionally, cells called as doublets by Scrublet were discarded (fourth panel). Only cells which passed all four of these filters were kept (bottom panel).

(**B**) Bar plot showing the number of cells for each scRNA-seq sample that passed all filters, faceted by timepoint.

Figure 3.2A-B

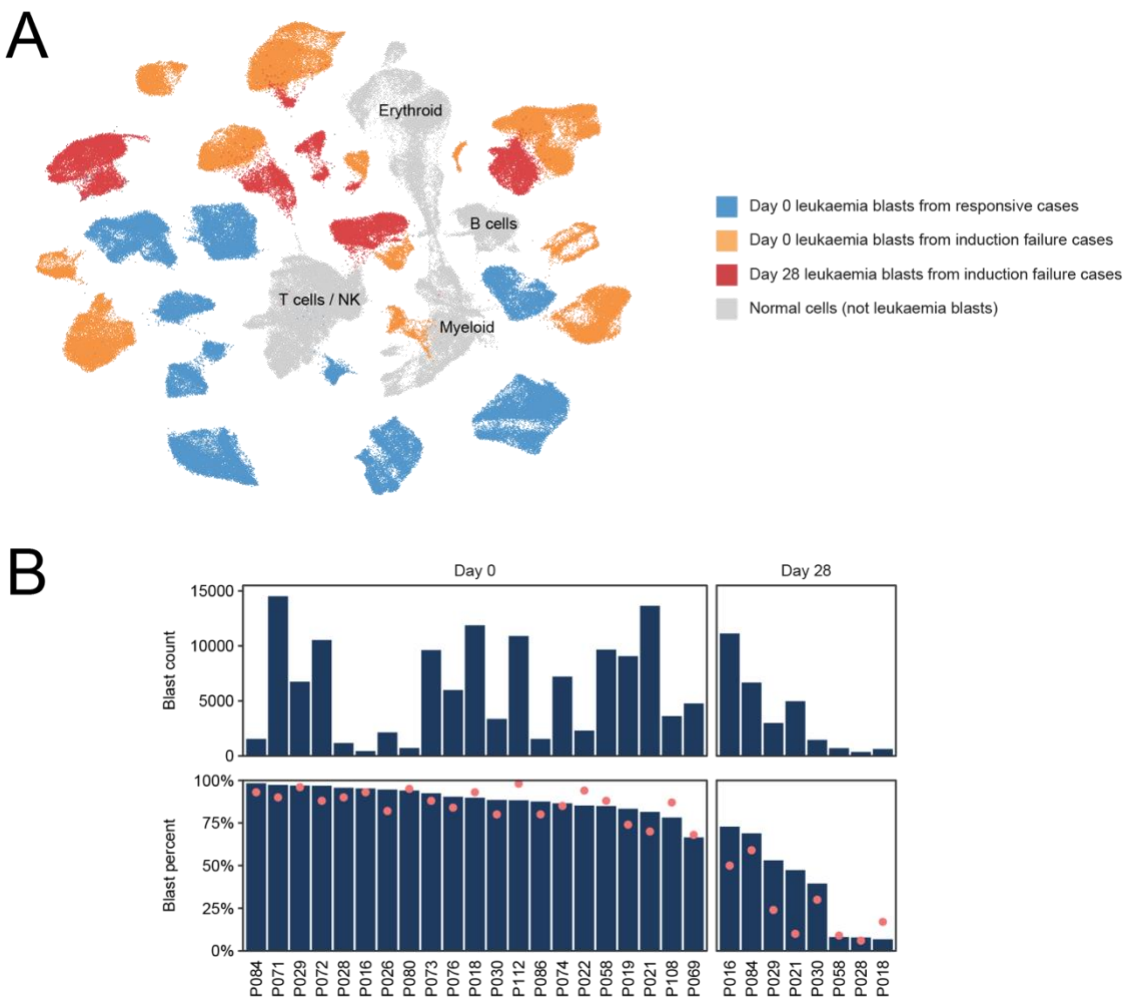

A
Erythroid
B cells
T cells / NK
Myeloid
Day 0 leukaemia blasts from responsive cases
Day 0 leukaemia blasts from induction failure cases
Day 28 leukaemia blasts from induction failure cases
Normal cells (not leukaemia blasts)
B
Day 0
Day 28
Blast count
15000
10000
5000
0
Blast percent
100%
75%
50%
25%
0%
P084
P071
P029
P072
P028
P016
P026
P080
P073
P076
P018
P030
P112
P086
P074
P022
P058
P019
P021
P108
P069
P016
P084
P029
P021
P030
P058
P028
P018

Figure 3.2C

C

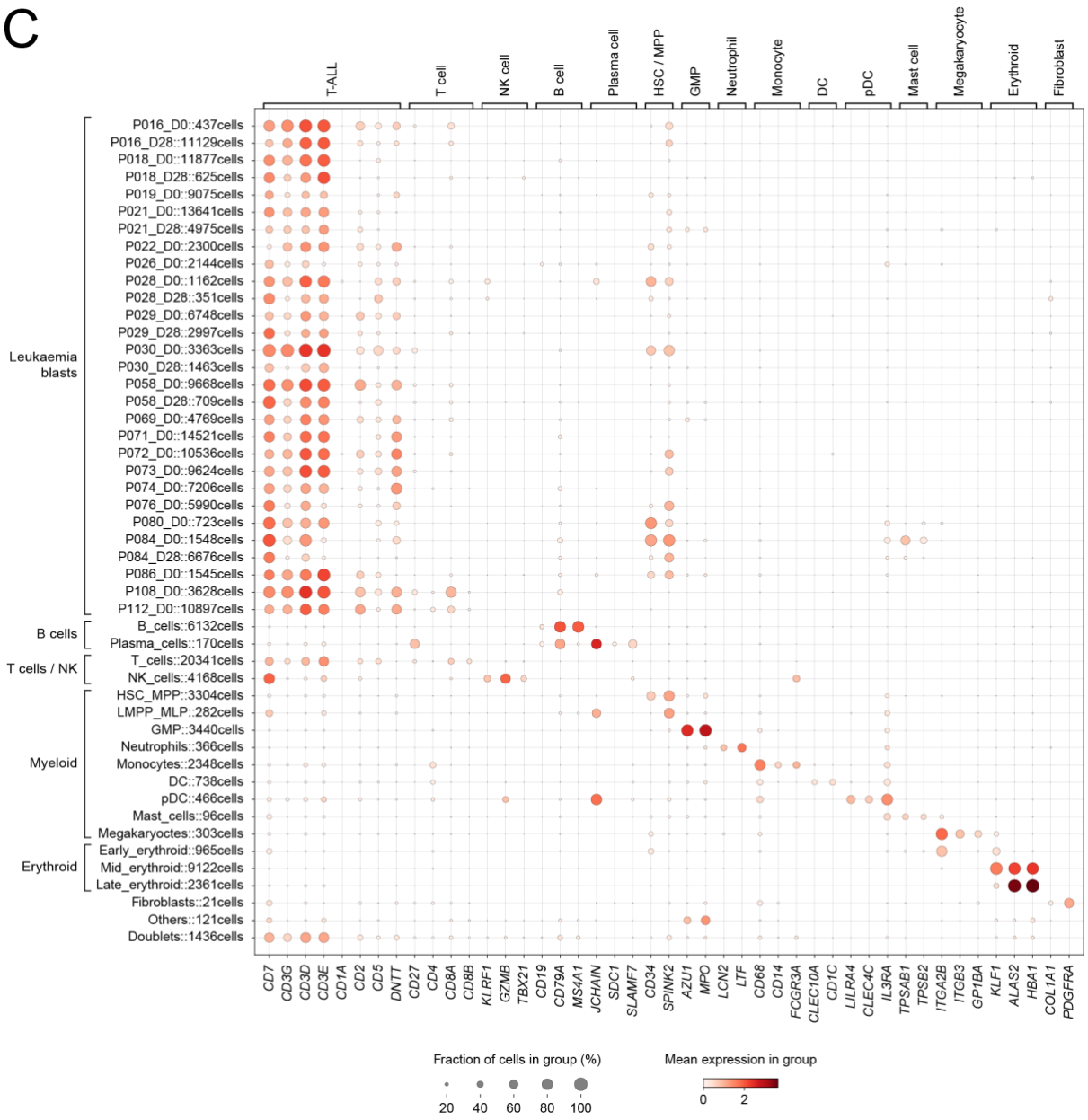

T-ALL
T cell
NK cell
B cell
Plasma cell
HSC / MPP
GMP
Neutrophil
Monocyte
DC
pDC
Mast cell
Megakaryocyte
Erythroid
Fibroblast
Leukaemia blasts
B cells
T cells / NK
Myeloid
Erythroid
P016_D0::437cells
P016_D28::11129cells
P018_D0::11877cells
P018_D28::625cells
P019_D0::9075cells
P021_D0::13641cells
P021_D28::4975cells
P022_D0::2300cells
P026_D0::2144cells
P028_D0::1162cells
P028_D28::351cells
P029_D0::6748cells
P029_D28::2997cells
P030_D0::3363cells
P030_D28::1463cells
P058_D0::9668cells
P058_D28::709cells
P069_D0::4769cells
P071_D0::14521cells
P072_D0::10536cells
P073_D0::9624cells
P074_D0::7206cells
P076_D0::5990cells
P080_D0::723cells
P084_D0::1548cells
P084_D28::6676cells
P086_D0::1545cells
P108_D0::3628cells
P112_D0::10897cells
B_cells::6132cells
Plasma_cells::170cells
T_cells::20341cells
NK_cells::4168cells
HSC_MPP::3304cells
LMPP_MLP::282cells
GMP::3440cells
Neutrophils::366cells
Monocytes::2348cells
DC::738cells
pDC::466cells
Mast_cells::96cells
Megakaryoctes::303cells
Early_erythroid::965cells
Mid_erythroid::9122cells
Late_erythroid::2361cells
Fibroblasts::21cells
Others::121cells
Doublets::1436cells
CD7
CD3G
CD3D
CD3E
CD1A
CD2
CD5
DNTT
CD27
CD4
CD8A
CD8B
KLRF1
GZMB
TBX21
CD19
CD79A
MS4A1
JCHAIN
SDC1
SLAMF7
CD34
SPINK2
AZU1
MPO
LCN2
LTF
CD68
CD14
FCGR3A
CLEC10A
CD1C
LILRA4
CLEC4C
IL3RA
TPSAB1
TPSB2
ITGA2B
ITGB3
GP1BA
KLF1
ALAS2
HBA1
COL1A1
PDGFRA
Fraction of cells in group (%)
20
40
60
80
100
Mean expression in group
0
2

Figure 3.2: Cell annotation of discovery cohort scRNA-seq.

(**A**) UMAP of 216,507 cells from the discovery cohort, including 160,327 leukaemia blasts (coloured) and 56,180 normal cells (grey). Day 0 blasts from individuals who responded to induction (blue); day 0 blasts from individuals with induction failure (orange); day 28 blasts from individuals with induction failure (red).

(**B**) Bar plots showing number of blasts (top panel) and percentage of blasts (navy bars in bottom panel) in each scRNA-seq sample compared to values obtained by clinical flow cytometry (pink dots in bottom panel).

(**C**) Dot plot showing expression of canonical markers to support cell annotation. Dot size corresponds to fraction of cells in each group that expresses each gene and colour intensity indicates mean expression. Leukaemia blasts are split into separate rows for each sample, defined by patient and timepoint. Total cell numbers for each cell annotation are denoted in row labels.

NK, natural killer cell; HSC, haematopoietic stem cell; MPP, multipotent progenitor; LMPP, lymphoid-primed multipotent progenitor; MLP, multi-lymphoid progenitor; GMP, granulocyte-monocyte progenitor; DC, dendritic cell; pDC, plasmacytoid dendritic cell.

3.3 Single cell transcriptome sequencing of an extended cohort of T-ALL

Later in my doctoral studies, I analysed an extended cohort of 37 children with T-ALL ("extension cohort"). These were unselected patients from GOSH, supplied by Dr David O'Connor, Dr Jack Bartram, Dr Angus Hodder. This cohort included seven individuals who were refractory to initial induction and one person who died from treatment-related complications during initial induction (L074) (**Table 3.2**). Diagnostic (day 0) samples were available for 35 out of 37 individuals (missing L074 and T030). Additional timepoints including day 7 and day 28 of induction, as well as relapse samples, were sequenced where possible. Samples were mostly bone marrow aspirate or peripheral blood, except for the diagnostic pleural fluid of L026.

Following quality control and processing of scRNA-seq data, I obtained a total of 333,487 high quality cells from 55 scRNA-seq samples, including two technical replicates which were performed on the diagnostic peripheral blood sample of L086 (**Figure 3.3A-B**). At least 1,000 high quality cells were captured in 54 of the 55 samples, except for the diagnostic sample of T027, which only had 591 cells. While a large fraction of cells from the day 7 sample of T019 was discarded due to low number of expressed genes and high percentage of reads mapping to mitochondrial genes, the remaining cells were of high enough quality for subsequent analyses. The diagnostic peripheral blood sample of L028 contained an exceptionally high number of cells (42,078 cells), but this sample was otherwise deemed acceptable.

As with the discovery cohort, cells from this extension cohort were similarly annotated, distinguishing leukaemia blasts from normal haematopoietic cells (**Figure 3.4A-C**). The percentages of blasts in each scRNA-seq sample were broadly agreeable with values from clinical flow cytometry (**Figure 3.4B**), except for the diagnostic samples T027 and T020, where the percentages of blasts on scRNA-seq were much lower than expected. In three of the day 28 samples (from L028, T007 and T024), no blasts detected in the scRNA-seq data, consistent with their very low percentages of blasts on clinical flow cytometry (<1%).

Table 3.2: Clinical metadata and sample availability of extension cohort.

ID	Sex	Age at diagnosis	ETP status	Genomic subtype	Induction response	Timepoints with scRNA-seq data
L026	F	11-15 years	nonETP	T-other	Responsive	D0 (pleural fluid)
L028	M	6-10 years	nonETP	TLX3	Responsive	D0 (PB), D28
L029	F	1-5 years	nonETP	TAL1	Responsive	D0 (PB)
L074	M	6-10 years	nonETP	T-other	Death	D7
L084	F	6-10 years	ETP	BCL11B	Responsive	D0
L086	M	<1 years	ETP	LMO2-STAG2	IF	D0 (PB)
L088	F	<1 years	nonETP	MYC-TRA	Responsive	D0 (PB)
L095	M	11-15 years	nonETP	T-other	Responsive	D0
T001	M	6-10 years	nonETP	T-other	Responsive	D0
T002	M	6-10 years	nonETP	MYC-TRA	Responsive	D0
T003	M	6-10 years	nonETP	TAL1	Responsive	D0
T004	M	6-10 years	nonETP	TAL1	IF	D0
T005	M	6-10 years	nonETP	TLX3	Responsive	D0
T006	M	6-10 years	nonETP	SET-NUP214	IF	D0
T007	M	1-5 years	nonETP	LMO2-TRA	Responsive	D0, D7, D28
T008	M	11-15 years	nonETP	TAL1	Responsive	D0
T009	M	1-5 years	nonETP	T-other	Responsive	D0
T010	M	6-10 years	nonETP	TAL1	Responsive	D0, D7, D28
T011	M	6-10 years	nonETP	TAL1	Responsive	D0
T012	M	1-5 years	nonETP	MED12	IF	D0
T013	M	11-15 years	nonETP	KMT2A	Responsive	D0, D7
T014	F	11-15 years	ETP	MED12	IF	D0, D7, D28, Relapse
T015	M	6-10 years	nonETP	TLX3	Responsive	D0
T016	M	6-10 years	nonETP	TAL1	Responsive	D0, D7, D28
T017	M	11-15 years	nonETP	NUP98	Responsive	D0
T018	M	6-10 years	nonETP	TLX1	Responsive	D0
T019	M	11-15 years	nonETP	HOXA	IF	D0, D7
T020	M	6-10 years	nonETP	TAL1	Responsive	D0
T021	M	1-5 years	nonETP	TLX3	Responsive	D0
T022	M	1-5 years	ETP	MLLT10	Responsive	D0
T023	M	1-5 years	nonETP	KMT2A	Responsive	D0
T024	M	11-15 years	nonETP	TAL1	Responsive	D0, D7, D28, Relapse
T026	M	1-5 years	nonETP	TLX3	Responsive	D0
T027	M	6-10 years	nonETP	TAL1	Responsive	D0
T028	F	6-10 years	nonETP	KMT2A	Responsive	D0
T029	M	6-10 years	nonETP	TAL1	Responsive	D0, D7, D28, Relapse
T030	F	1-5 years	nonETP	T-other	IF	Relapse

ETP, early T cell precursor; IF, induction failure; PB, peripheral blood.

Figure 3.3

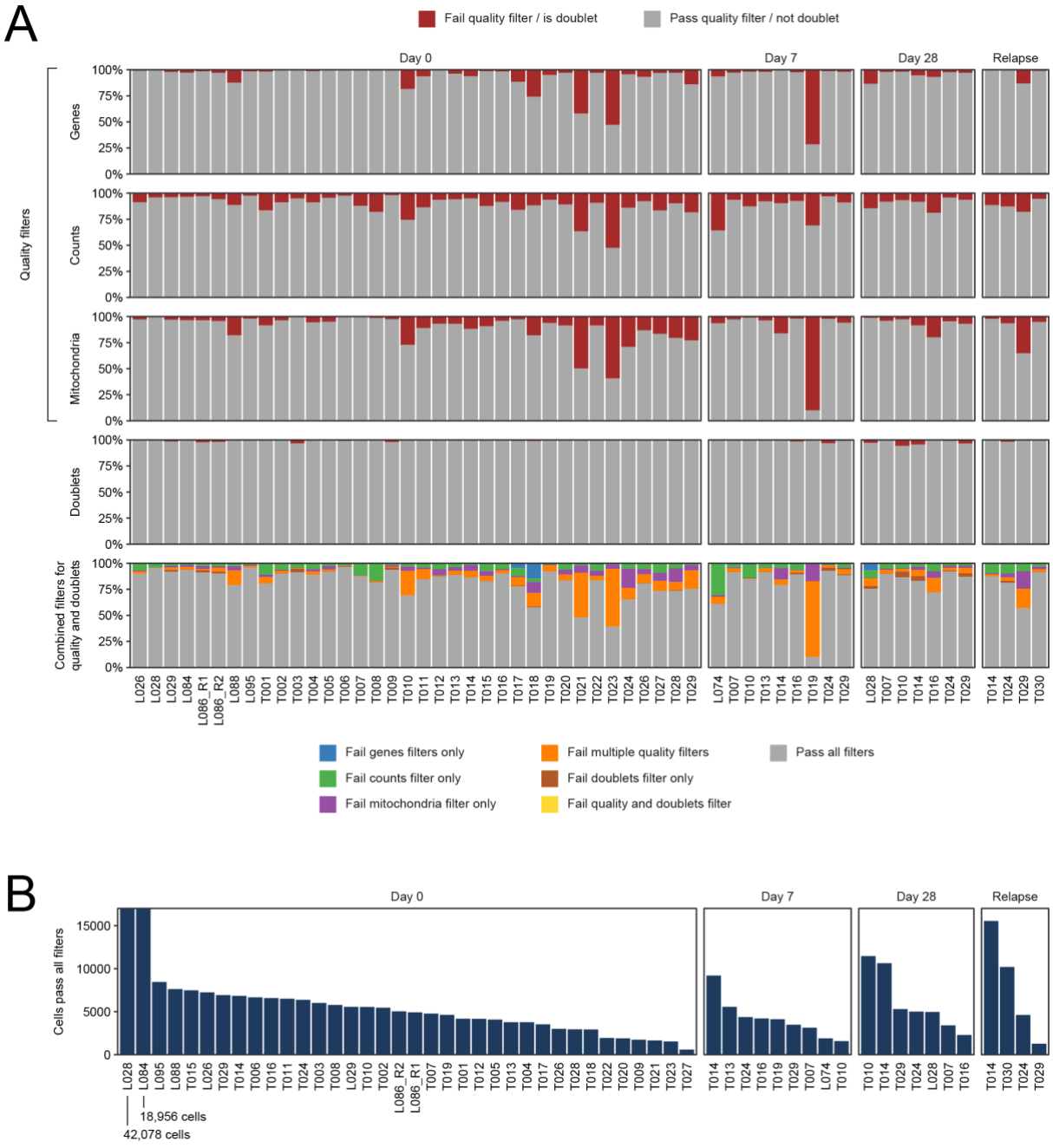

A
Fail quality filter / is doublet
Pass quality filter / not doublet
Day 0
Day 7
Day 28
Relapse
Quality filters
Genes
Counts
Mitochondria
Doublets
Combined filters for quality and doublets
100%
75%
50%
25%
0%
L026 L028 L029 L084 L086_R1 L086_R2 L088 L095 T001 T002 T003 T004 T005 T006 T007 T008 T009 T010 T011 T012 T013 T014 T015 T016 T017 T018 T019 T020 T021 T022 T023 T024 T026 T027 T028 T029
L074 T007 T010 T013 T014 T016 T019 T024 T029
L028 T007 T010 T014 T016 T024 T029
T014 T024 T029 T030
Fail genes filters only
Fail counts filter only
Fail mitochondria filter only
Fail multiple quality filters
Fail doublets filter only
Fail quality and doublets filter
Pass all filters
B
Day 0
Day 7
Day 28
Relapse
Cells pass all filters
15000
10000
5000
0
L028 L084 L095 L088 T015 L026 T029 T014 T006 T016 T011 T024 T003 T008 L029 T010 T002 L086_R2 L086_R1 T007 T019 T001 T012 T005 T013 T004 T017 T026 T028 T018 T022 T020 T009 T021 T023 T027
T014 T013 T024 T016 T019 T029 T007 L074 T010
T010 T014 T029 T024 L028 T007 T016
T014 T030 T024 T029
18,956 cells
42,078 cells

Figure 3.3: Quality control of extension cohort scRNA-seq.

(**A**) Bar plots showing percentage of cells for each scRNA-seq sample that passed each filter. Cell quality filters include: genes filter (top panel, keep cells with >300 expressed genes), counts filter (second panel, keep cells with >1000 total RNA counts) and mitochondria filter (third panel, keep cells with <15% of RNA counts mapping to mitochondrial genes). Additionally, cells called as doublets by Scrublet were discarded (fourth panel). Only cells which passed all four of these filters were kept (bottom panel).

(**B**) Bar plot showing the number of cells for each scRNA-seq sample that passed all filters, faceted by timepoint. Two scRNA-seq technical replicates were performed on the day 0 sample of L086.

Figure 3.4A-B

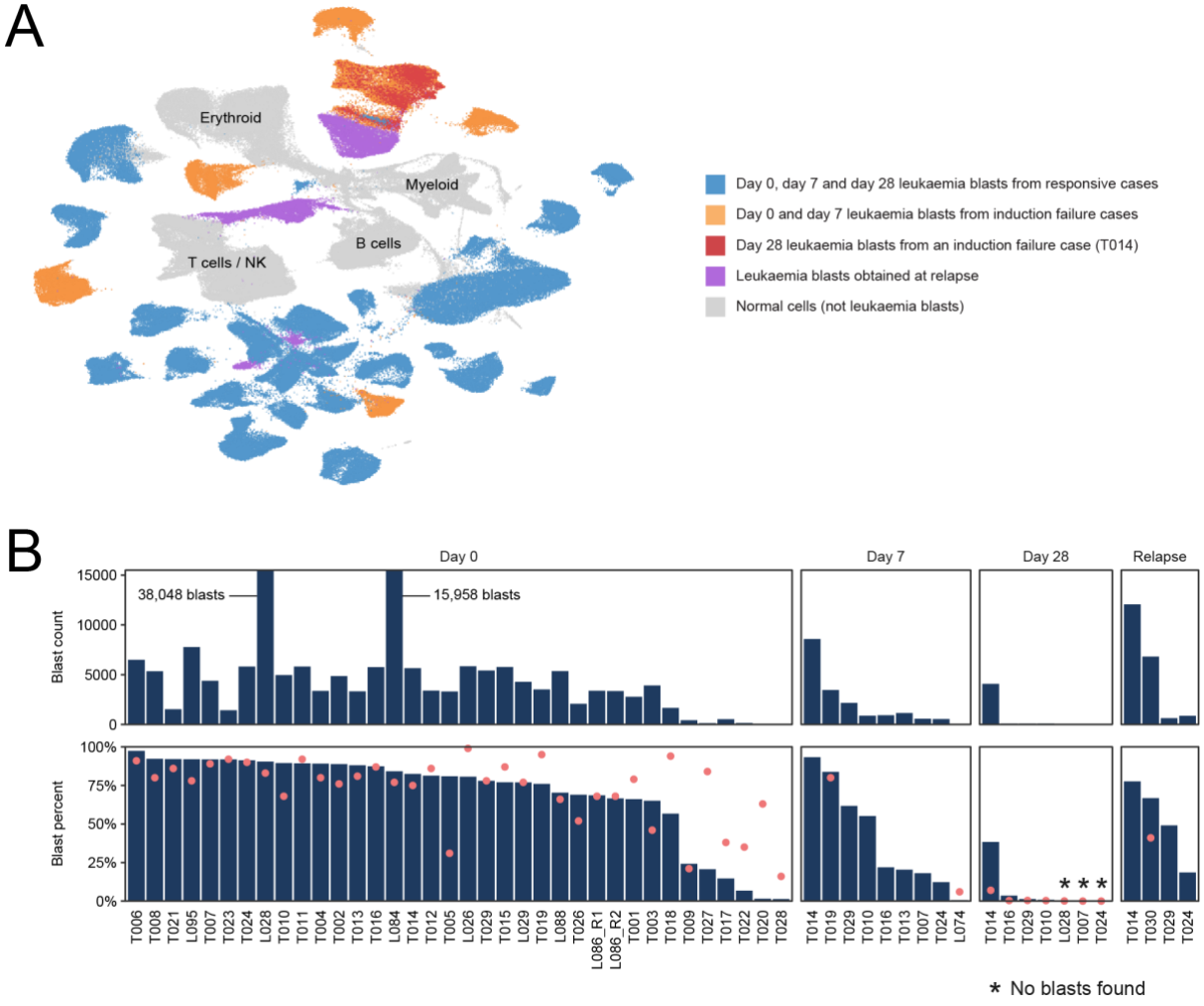

A
Erythroid
Myeloid
B cells
T cells / NK
Day 0, day 7 and day 28 leukaemia blasts from responsive cases
Day 0 and day 7 leukaemia blasts from induction failure cases
Day 28 leukaemia blasts from an induction failure case (T014)
Leukaemia blasts obtained at relapse
Normal cells (not leukaemia blasts)
B
Day 0
Day 7
Day 28
Relapse
Blast count
38,048 blasts
15,958 blasts
Blast percent
T006 T008 T021 L095 T007 T023 T024 L028 T010 T011 T004 T002 T013 T016 L084 T014 T012 T005 L026 T029 T015 L029 T019 L088 T026 L086_R1 L086_R2 T001 T003 T018 T009 T027 T017 T022 T020 T028
T014 T019 T029 T010 T016 T013 T007 T024 L074
T014 T016 T029 T010 L028 T007 T024
T014 T030 T029 T024
* No blasts found

Figure 3.4C


C
T-ALL
T cell
NK cell
B cell
Plasma cell
HSC / MPP
GMP
Neutrophil
Monocyte
Macrophage
DC
pDC
MEMP / MEP
Megakaryote
Erythroid
Fibroblast
Leukaemia blasts
B cells
T cells / NK
Myeloid
Erythroid
L026_D0::5842cells
L028_D0::38048cells
L029_D0::4283cells
L074_D07::13cells
L084_D0::15958cells
L086_D0_R1::3373cells
L086_D0_R2::3355cells
L088_D0::5359cells
L095_D0::7773cells
T001_D0::2768cells
T002_D0::4851cells
T003_D0::3916cells
T004_D0::3363cells
T005_D0::3298cells
T006_D0::6495cells
T007_D0::4385cells
T007_D07::568cells
T008_D0::5339cells
T009_D0::419cells
T010_D0::4961cells
T010_D07::872cells
T010_D28::93cells
T011_D0::5808cells
T012_D0::3388cells
T013_D0::3324cells
T013_D07::1133cells
T014_D0::5647cells
T014_D07::8579cells
T014_D28::4071cells
T014_Relapse::12065cells
T015_D0::5768cells
T016_D0::5746cells
T016_D07::919cells
T016_D28::80cells
T017_D0::517cells
T018_D0::1660cells
T019_D0::3517cells
T019_D07::3451cells
T020_D0::29cells
T021_D0::1525cells
T022_D0::131cells
T023_D0::1413cells
T024_D0::5817cells
T024_D07::535cells
T024_Relapse::855cells
T026_D0::2067cells
T027_D0::122cells
T028_D0::39cells
T029_D0::5406cells
T029_D07::2152cells
T029_D28::69cells
T029_Relapse::626cells
T030_Relapse::6808cells
B_cells::20739cells
Plasma_cells::293cells
T_cells::36279cells
NK_cells::7828cells
HSC_MPP::2180cells
GMP::3310cells
Neutrophils::745cells
Monocytes::10125cells
Macrophages::523cells
DC::780cells
pDC::617cells
MEM_MEP::1671cells
Megakaryotes::866cells
Erythroid::26893cells
Fibroblasts_BM::49cells
Fibroblasts_Pleural::59cells
Doublets_B_T::443cells
Doublets_Leuk_T::657cells
Doublets_T_Ery::635cells
Doublets_T_Mono::196cells
CD7
CD3G
CD3D
CD3E
CD1A
CD2
CD5
DNTT
CD27
CD4
CD8A
CD8B
KLRF1
GZMB
TBX21
CD19
CD79A
MS4A1
JCHAIN
SDC1
SLAMF7
CD34
SPINK2
AZU1
MPO
LCN2
LTF
CD68
CD14
FCGR3A
C1QA
SPP1
CLEC10A
CD1C
LILRA4
CLEC4C
IL3RA
GATA2
GATA1
ITGA2B
ITGB3
GP1BA
KLF1
ALAS2
HBA1
COL1A1
PDGFRA
Fraction of cells in group (%)
20 40 60 80 100
Mean expression in group
0 2 4

Figure 3.4: Cell annotation of extension cohort scRNA-seq.

(**A**) UMAP of 333,487 cells from the extension cohort, including 218,599 leukaemia blasts (coloured) and 114,888 normal cells (grey). Day 0, day 7 and day 28 blasts from individuals who responded to induction (blue); day 0 and day 7 blasts from individuals with induction failure (orange); day 28 blasts from individuals with induction failure (red); blasts found at relapse (purple).

(**B**) Bar plots showing number of blasts (top panel) and percentage of blasts (navy bars in bottom panel) in each scRNA-seq sample compared to values obtained by clinical flow cytometry (pink dots in bottom panel). No blasts were found in three of the day 28 samples (indicated by asterisk). Two scRNA-seq technical replicates were performed on the day 0 sample of L086.

(**C**) Dot plot showing expression of canonical markers to support cell annotation. Dot size corresponds to fraction of cells in each group that expresses each gene and colour intensity indicates mean expression. Leukaemia blasts are split into separate rows for each sample, defined by patient and timepoint. Two scRNA-seq technical replicates were performed on the day 0 sample of L086. Total cell numbers for each cell annotation are denoted in row labels.

NK, natural killer cells; HSC, haematopoietic stem cell; MPP, multipotent progenitor; GMP, granulocyte-monocyte progenitor; DC, dendritic cell; pDC, plasmacytoid dendritic cell.

3.4 Comparing T-ALL blasts to stages of T cell differentiation

T-ALL has traditionally been compared to stages of T cell development based on the expression of specific cell markers (Bene et al., 1995). To obtain an unbiased, quantitative comparison of T-ALL blasts to a reference cell atlas of T cell development, I used a logistic regression framework that was previously developed in the group (Young et al., 2018).

I chose a recently published cell atlas of the developing fetal immune system (Suo et al., 2022) as the reference cell atlas to train a logistic regression model, because of its coverage of the various T cell developmental stages (**Table 3.3**). Glial cells from this cell atlas were included in model training as a non-haematopoietic cell internal control. In addition, Schwann cell precursors (SCP) from a separate fetal adrenal gland cell atlas were included as an external control (Kildisiute et al., 2021). Prior to model training, 20% of cells from each cell type were excluded from training and used for model evaluation (**Figure 3.5A**).

Subsequently, I applied the trained logistic regression model to determine for each single cell leukaemia transcriptome from diagnostic samples, which reference cell type it best matches to (**Figure 3.5B**). T-ALL genomic subtypes which are associated with early differentiation state (BCL11B, MED12, NUP98, SET-NUP214, KMT2A, MLLT10, HOXA) (Pölönen et al., 2024) exhibit stronger cell matching to haematopoietic progenitor cell types such as that of HSC / MPP or LMPP / MLP, compared to genomic subtypes associated with later differentiation state (TLX3, TLX3, TAL1). In general, T-ALL blasts most strongly matched to DN thymocytes (early). Unusually, P018 did not match to DN thymocytes (early) and instead exhibited the strongest match to Type 3 innate T cells (Th17-like), suggesting that it may possess a distinct leukaemia cell state from that of typical T-ALL.

Highly similar cell types may confuse the training of a logistic regression model. To avoid this issue and reduce the number of cell types for model training, I collapsed similar cell types together (**Figure 3.5C**). Applying this new logistic regression model to single cell leukaemia transcriptomes similarly revealed the early differentiation states of some genomic subtypes, and the resemblance of P018 to unconventional T cells and innate lymphoid cells (ILC) (**Figure 3.5D**). Interestingly, L086, which belongs to the LMO2-STAG2 subtype, shows a distinct matching to MEMP / MEP, but not to other early progenitor stages (HSC / MPP or LMPP / MLP), reflecting a bias towards the MEMP / MEP lineage (Pölönen et al., 2024). Overall, I am able to discern broad differences in the transcriptome similarity of T-ALL blasts to various T cell subsets, although it is difficult to pin down specific differentiation states.

Table 3.3: Cell labels from a reference cell atlas of T cell development.

Broad cell categories	Descriptive cell labels	Cell labels in original paper
HSC / MPP	HSC / MPP	HSC_MPP
	MPP (cycling)	CYCLING_MPP
LMPP / MLP	LMPP / MLP	LMPP_MLP
DN thymocytes	DN thymocytes (early)	DN(early)_T
	DN thymocytes (proliferating)	DN(P)_T
	DN thymocytes (quiescent)	DN(Q)_T
DP thymocytes	DP thymocytes (proliferating)	DP(P)_T
	DP thymocytes (quiescent)	DP(Q)_T
Conventional T cells	Immature αβ T cells	ABT(ENTRY)
	CD8+ αβ T cells	CD8+_T
	CD4+ αβ T cells	CD4+_T
	Regulatory T cells	TREG
	T cells (cycling)	CYCLING_T
Unconventional T cells	CD8αα+ T cells	CD8AA
	Type 1 innate T cells (NKT-like)	TYPE_1_INNATE_T
	Type 3 innate T cells (Th17-like)	TYPE_3_INNATE_T
NK cells	NK cells	NK
	NK cells (cycling)	CYCLING_NK
ILC	ILC2	ILC2
	ILC3	ILC3
	ILC (cycling)	CYCLING_ILC
B cell progenitors	Pre-pro-B cells	PRE_PRO_B
	Pro-B cells	PRO_B
Myeloid progenitors	CMP	CMP
	GMP	GMP
pDC	pDC	PDC
	pDC (cycling)	CYCLING_PDC
MEMP / MEP	MEMP	MEMP
	MEP	MEP
Glial cells	Glial cells	GLIAL
SCP	SCP	SCP

HSC, haematopoietic stem cell; MPP, multipotent progenitor; LMPP, lymphoid-primed multipotent progenitor; MLP, multi-lymphoid progenitor; DN, double-negative; DP, double-positive; NK, natural killer; ILC, innate lymphoid cells; CMP, common myeloid progenitor; GMP, granulocyte-monocyte progenitor; pDC, plasmacytoid dendritic cells; MEMP, mast cell-megakaryocyte-erythroid progenitor; MEP, megakaryocyte-erythroid progenitor; SCP, Schwann cell precursor.

Figure 3.5A

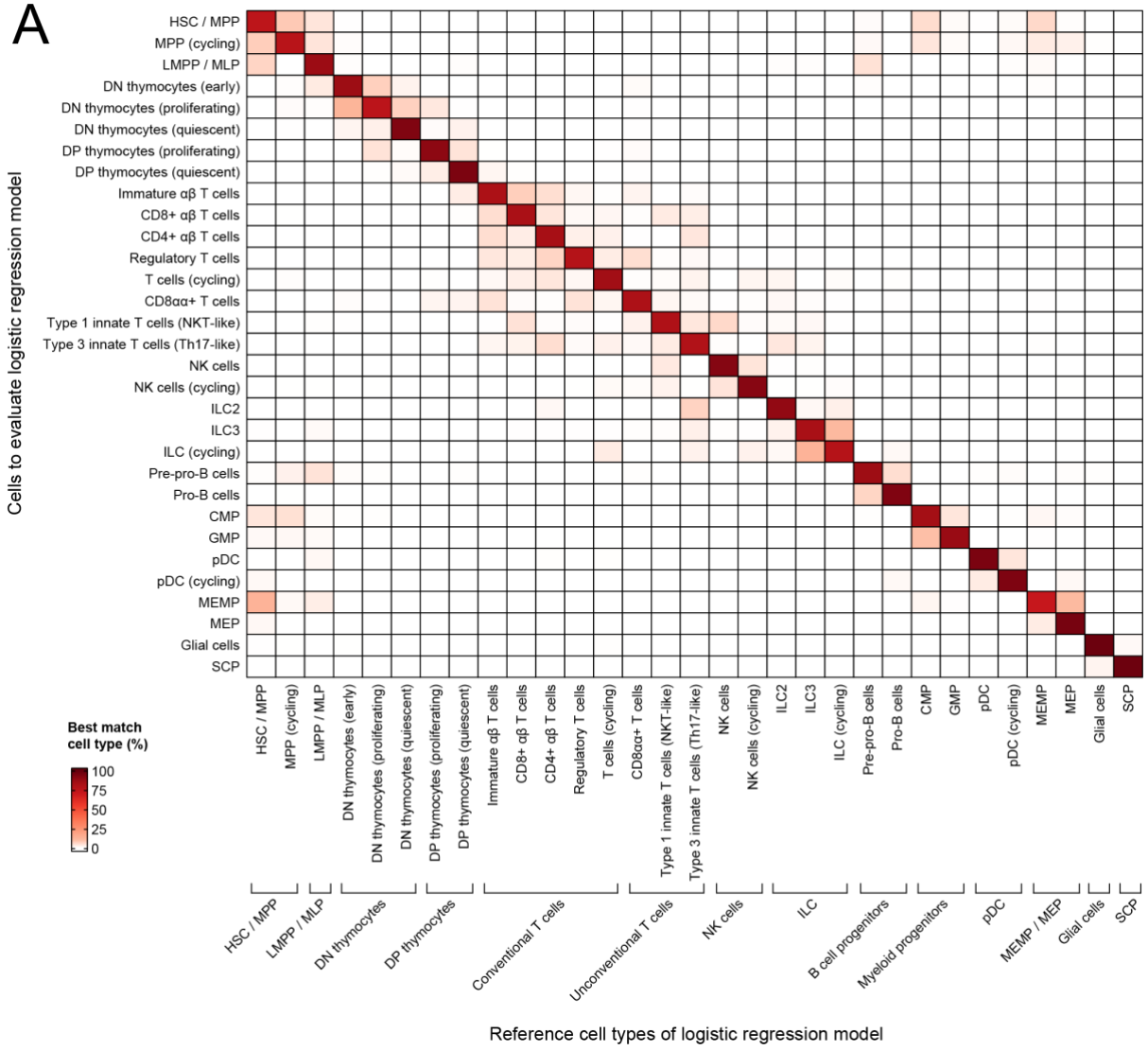

A
Cells to evaluate logistic regression model
HSC / MPP
MPP (cycling)
LMPP / MLP
DN thymocytes (early)
DN thymocytes (proliferating)
DN thymocytes (quiescent)
DP thymocytes (proliferating)
DP thymocytes (quiescent)
Immature αβ T cells
CD8+ αβ T cells
CD4+ αβ T cells
Regulatory T cells
T cells (cycling)
CD8αα+ T cells
Type 1 innate T cells (NKT-like)
Type 3 innate T cells (Th17-like)
NK cells
NK cells (cycling)
ILC2
ILC3
ILC (cycling)
Pre-pro-B cells
Pro-B cells
CMP
GMP
pDC
pDC (cycling)
MEMP
MEP
Glial cells
SCP
Best match cell type (%)
100
75
50
25
0
HSC / MPP
LMPP / MLP
DN thymocytes
DP thymocytes
Conventional T cells
Unconventional T cells
NK cells
ILC
B cell progenitors
Myeloid progenitors
pDC
MEMP / MEP
Glial cells
SCP
Reference cell types of logistic regression model

Figure 3.5B

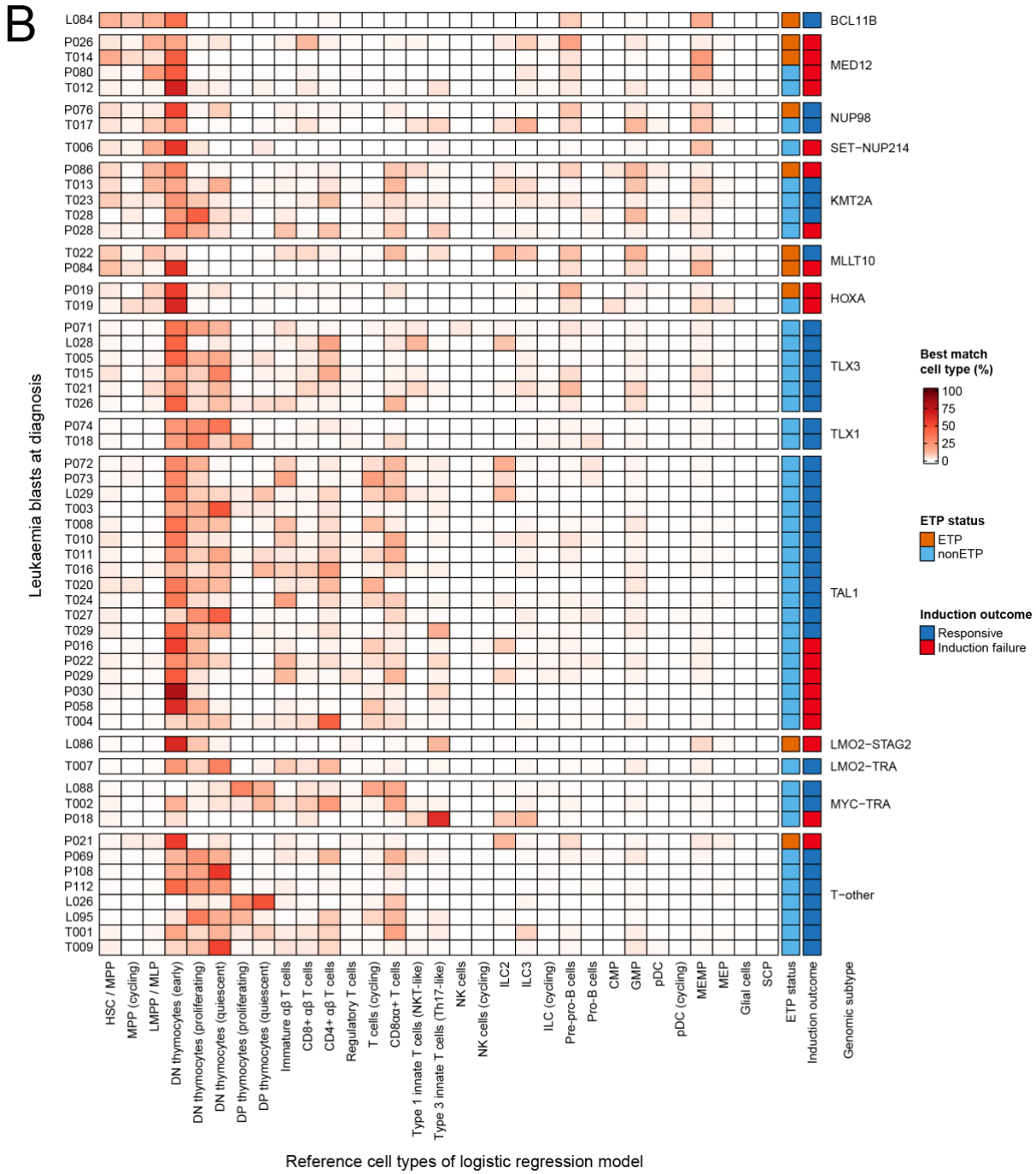

B
Leukaemia blasts at diagnosis
L084
P026
T014
P080
T012
P076
T017
T006
P086
T013
T023
T028
P028
T022
P084
P019
T019
P071
L028
T005
T015
T021
T026
P074
T018
P072
P073
L029
T003
T008
T010
T011
T016
T020
T024
T027
P029
P016
P022
P029
P030
P058
T004
L086
T007
L088
T002
P018
P021
P069
P108
P112
L026
L095
T001
T009
BCL11B
MED12
NUP98
SET−NUP214
KMT2A
MLLT10
HOXA
TLX3
TLX1
TAL1
LMO2−STAG2
LMO2−TRA
MYC−TRA
T−other
Best match cell type (%)
100
75
50
25
0
ETP status
ETP
nonETP
Induction outcome
Responsive
Induction failure
HSC / MPP
MPP (cycling)
LMPP / MLP
DN thymocytes (early)
DN thymocytes (proliferating)
DN thymocytes (quiescent)
DP thymocytes (proliferating)
DP thymocytes (quiescent)
Immature αβ T cells
CD8+ αβ T cells
CD4+ αβ T cells
Regulatory T cells
T cells (cycling)
CD8αα+ T cells
Type 1 innate T cells (NKT-like)
Type 3 innate T cells (Th17-like)
NK cells
NK cells (cycling)
ILC2
ILC3
ILC (cycling)
Pre-pro-B cells
Pro-B cells
CMP
GMP
pDC
pDC (cycling)
MEMP
MEP
Glial cells
SCP
ETP status
Induction outcome
Genomic subtype
Reference cell types of logistic regression model

Figure 3.5C

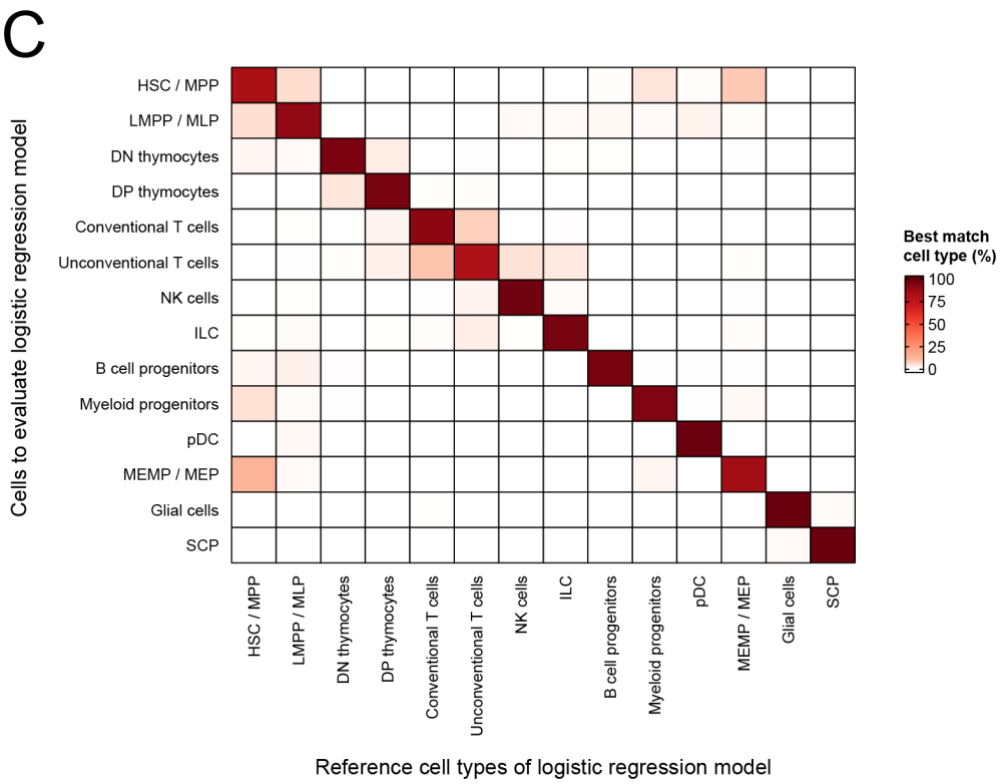

C
Cells to evaluate logistic regression model
HSC / MPP
LMPP / MLP
DN thymocytes
DP thymocytes
Conventional T cells
Unconventional T cells
NK cells
ILC
B cell progenitors
Myeloid progenitors
pDC
MEMP / MEP
Glial cells
SCP
Reference cell types of logistic regression model
Best match cell type (%)
100
75
50
25
0

Figure 3.5D

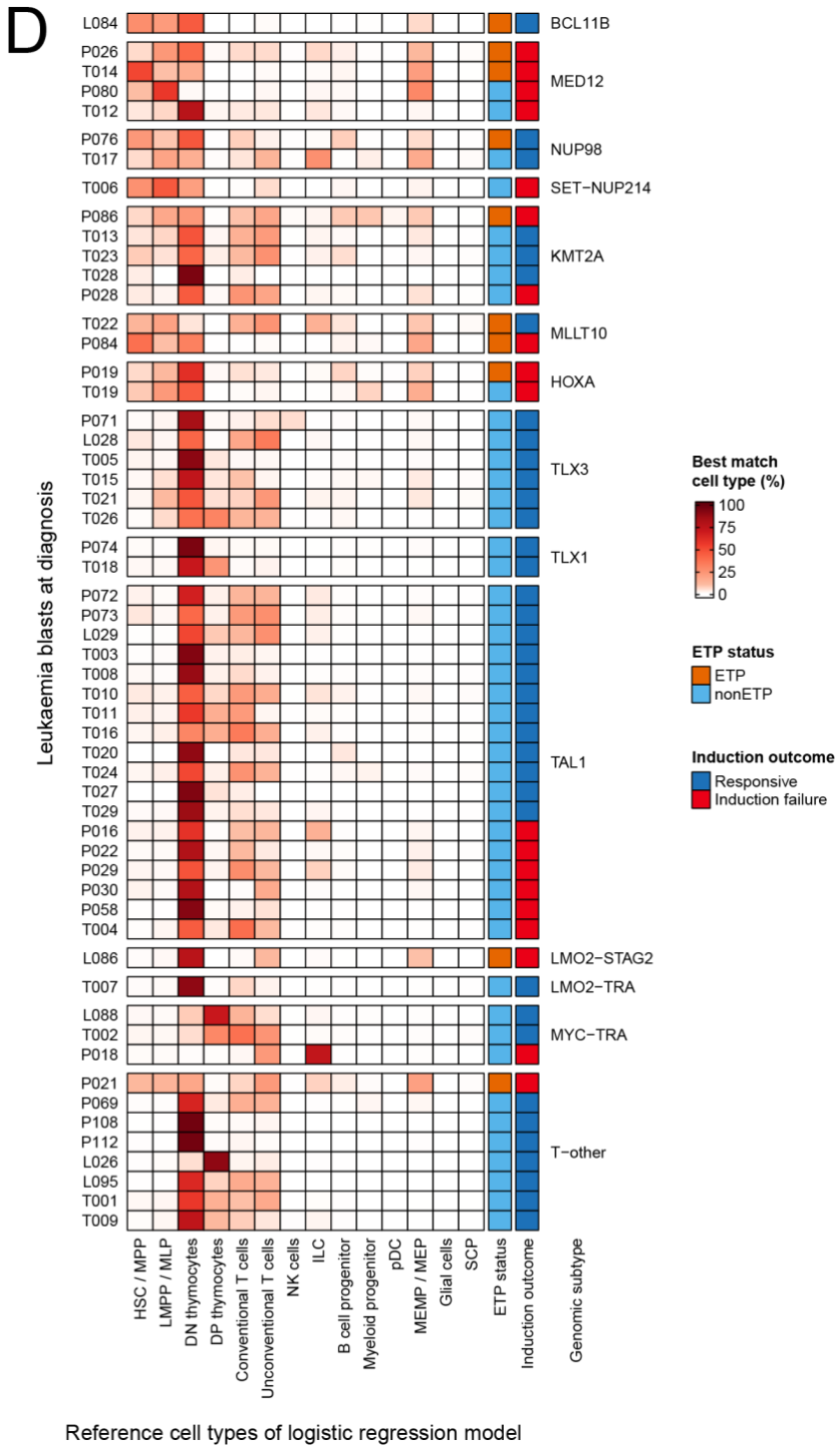

D
Leukaemia blasts at diagnosis
L084
P026
T014
P080
T012
P076
T017
T006
P086
T013
T023
T028
P028
T022
P084
P019
T019
P071
L028
T005
T015
T021
T026
P074
T018
P072
P073
L029
T003
T008
T010
T011
T016
T020
T024
T027
T029
P016
P022
P029
P030
P058
T004
L086
T007
L088
T002
P018
P021
P069
P108
P112
L026
L095
T001
T009
BCL11B
MED12
NUP98
SET−NUP214
KMT2A
MLLT10
HOXA
TLX3
TLX1
TAL1
LMO2−STAG2
LMO2−TRA
MYC−TRA
T−other
HSC / MPP
LMPP / MLP
DN thymocytes
DP thymocytes
Conventional T cells
Unconventional T cells
NK cells
ILC
B cell progenitor
Myeloid progenitor
pDC
MEMP / MEP
Glial cells
SCP
ETP status
Induction outcome
Genomic subtype
Best match cell type (%)
100
75
50
25
0
ETP status
ETP
nonETP
Induction outcome
Responsive
Induction failure
Reference cell types of logistic regression model

Figure 3.5: Projecting leukaemia blasts onto a fetal immune cell atlas by logistic regression.

(**A**) A logistic regression model was trained using cell types from a fetal immune cell atlas reference. Glial cells here represent an internal negative control cell type from the same fetal immune cell atlas. Additionally, Schwann cell precursors (SCP) from an independent fetal adrenal cell atlas were added as an external negative control cell type. Prior to model training, 20% of cells from each cell type were isolated as a testing set from the training set of cells. This testing set of cells was used to evaluate the predictive accuracy of the logistic regression model. Heatmap shows the percentage of cells of each cell type from the testing set of cells (y-axis) which best match to the reference cell types (x-axis) in the logistic regression model.

(**B**) Heatmap shows the percentage of blasts from each diagnostic sample (y-axis) which best match to the reference cell types (x-axis) in the logistic regression model. Genomic subtype, ETP status and induction outcome are indicated on the left margin.

(**C**) Similar to **Figure 3.5A**. However, similar cell types were collapsed (as described in **Figure 3.5A**) to reduce the number of categories to train the logistic regression model.

(**D**) Similar to **Figure 3.5B**, but using the new cell type labels in **Figure 3.5C**.

ETP, early T cell precursor; HSC, haematopoietic stem cell; MPP, multipotent progenitor; LMPP, lymphoid-primed multipotent progenitor; MLP, multi-lymphoid progenitor; DN, double-negative; DP, double-positive; NK, natural killer; ILC, innate lymphoid cells; CMP, common myeloid progenitor; GMP, granulocyte-monocyte progenitor; pDC, plasmacytoid dendritic cells; MEMP, mast cell-megakaryocyte-erythroid progenitor; MEP, megakaryocyte-erythroid progenitor; SCP, Schwann cell precursor.

3.5 Unified annotation of normal cells from both cohorts of T-ALL

I merged normal cells (i.e. cells which are not leukaemia blasts) from both the discovery and extension T-ALL cohorts to generate a unified cell annotation of normal cells in T-ALL (**Table 3.4** and **Figure 3.6A-C**). My annotation of normal cells consists of three hierarchical levels, depicting different levels of granularity. Level 3 cell annotation is the finest resolution level. Level 2 cell annotation combines similar level 3 cell annotations together, but keeps progenitor stages (e.g. B_progenitors), mature stages (e.g. B_naive) and activated stages (e.g. B_memory) separate. Finally, Level 1 cell annotation has the broadest resolution and considers major branches of the haematopoietic tree. Here, monocytes, macrophages and dendritic cells (DC) are grouped together (Mono_Mac_DC), but plasmacytoid dendritic cells (pDC) are kept as a distinct entity (Villani et al., 2017). There was sufficient mixing of normal cells from the discovery and extension cohorts (**Figure 3.6A**). Thus, batch correction in the cell embedding space was not performed to preserve differences in biological signals.

The major haematopoietic lineages could be identified on a UMAP by plotting the expression of marker genes (**Figure 3.6D**): T cells (*CD7*, *CD3D*), NK cells (*CD7*, *KLRF1*, absence of *CD3D*), B cells and plasma cells (*CD19*), haematopoietic stem and progenitor cells (*CD34*), granulocyte-monocyte progenitors (*MPO*), monocytes (*CD68*) and erythroid cells (*HBA1*). Cell types which have a distinct gene expression profile, but are not sufficiently prevalent to be clearly seen on a UMAP, were annotated using a dot plot (**Figure 3.6E**), such as for neutrophils (which express *LCN2* and *LTF*). Cell clusters containing doublets could be identified by their expression of marker genes for two distinct lineages (**Figure 3.6E**), such as *CD19* for B cells and *HBA1* for erythroid cells.

T cells express both *CD7* and the CD3 genes (*CD3G*, *CD3D*, *CD3E*), whereas NK cells express *CD7* and *KLRF1* but not the CD3 genes (**Figure 3.7A-E**). T cells can be broadly divided in those which are CD4+ and those which are CD8+. NK cells are classified into CD16+ (*FCGR3A*) and CD56+ (*NCAM1*). A subset of CD8+ T cells express effector genes such as granzymes (*GZMB*, *GZMK*), perforin (*PRF1*) and interferon gamma (*IFNG*), while other CD8+ T cells exhibit a proliferating phenotype (*MKI67*) or exhausted phenotype (*CTLA4*, *PCDC1*, *LAG3*, *TOX*, *HAVCR2*, *SLAMF6*). Notably, *GZMB* and *GZMK* expression appear mutually exclusive, suggesting that there may be different flavours of cytotoxic CD8+ T cells (Li et al., 2019). A subset of CD4+ T cells are *FOXP3*+ regulatory T cells. Finally, some T cells are do not fall under the categories of conventional CD8+ or CD4+ T cells. Instead, these are unconventional T cells, which express *ZBTB16* and *KLRB1* (Alonzo and Sant'Angelo, 2011; Fergusson et al., 2011). Within the unconventional T cell population are mucosal associated

invariant T (MAIT) cells, which express the marker genes *TRAV1-2* and *SLC4A10* (Domínguez Conde et al., 2022).

B cells express both *CD19* and *MS4A1* (CD20) (**Figure 3.8A-D**). Plasma cells are positive for *CD19*, but do not express *MS4A1*. Instead, plasma cells express a unique set of marker genes (*JCHAIN*, *SLAMF7*, *SDC1*, *TNFRSF17*). Memory B cells express *TNFRSF13B* and *CD80*. Both plasma cells and memory B cells express class-switched immunoglobulins (*IGHG1*, *IGHA1*), in addition to non-class-switched immunoglobulins (*IGHM*, *IGHD*) seen in naive B cells. A subset of memory B cells express *ITGAX*, corresponding to a previously described age-related B cell that may be involved in autoimmune disease (Domínguez Conde et al., 2022). The significance of *ITGAX*+ memory B cells in T-ALL remains unclear. I found stages of B cell development, including pro-B cells (*IGLL1*, *VPREB1*, *MME*), pre-B cells (*SPIB*, *IL4R*, *TCL1A*) and immature B cells (*MS4A1*, *CR2*, *CD22*, *CD40*), consistent with bone marrow being the site of B lymphopoiesis.

Various myeloid cells, erythroid cells and fibroblasts were found in my data (**Figure 3.9A-D**). Haematopoietic stem and progenitor cells (HSPC) are quiescent cells which express *CD34* and *SPINK2*, whereas granulocyte-monocyte progenitors (GMP) are proliferating cells which express *AZU1* and *MPO*. Neutrophils are a distinct population marked by *LCN2* and *LTF*. The *CD68*+ myeloid compartment includes monocytes, macrophages, DCs and pDCs. Monocytes can be subdivided into classical CD14+ monocytes and nonclassical CD16+ monocytes (*FCGR3A*). Macrophages express many overlapping markers with monocytes, but they form a distinct cluster on the UMAP plot (**Figure 3.9A-B**) and express additional macrophage-specific markers (*C1QA*, *CD163*, *SPP1*). DCs mainly express the markers of DC2 (*CLEC10A*, *CD1C*) and weakly express markers of DC1 and other DC subsets (*CLEC9A*, *BATF3*, *ANPEP*). I was unable to further subdivide the DC cell cluster in my data. The pDCs are a distinct population which express *JCHAIN*, *LILRA4* and *CLEC4C*. Megakaryocyte-erythrocyte-mast cell progenitors (MEMP) and megakaryocyte-erythrocyte progenitors (MEP) are progenitor cells marked by the expression of *GATA2* and *GATA1*, and give rise to megakaryocytes (*ITGA2B*, *ITGB3*, *GP1BA*), erythroid lineages (*ALAS2*, *HBA1*) and mast cells (*TPSAB2*, *TPSB2*). The erythroid lineage is further subdivided based on proliferation state and expression of *GATA2*, *GATA1* and *KLF1*. Fibroblasts (*COL1A1*, *PDGFRA*) were detected in small numbers, either from bone marrow samples, or from pleural fluid (patient L026 only).

Table 3.4: Relationship between level 1, 2 and 3 cell annotation labels.

Level 1 cell annotation	Level 2 cell annotation	Level 3 cell annotation
NK_cells	NK_cells	NK_CD16
		NK_CD56
T_cells	T_innate	T_innate
		T_innate_MAIT
	T_CD8_naive	T_CD8
	T_CD8_activated	T_CD8_GZMB
		T_CD8_GZMK
		T_CD8_MKI67
		T_CD8_CTLA4
	T_CD4_naive	T_CD4
	T_CD4_regulatory	T_CD4_FOXP3
B_cells	B_progenitors	Pro_B
		Pre_B
		Immature_B
	B_naive	B_cells
	B_memory	B_mem
		B_mem_ITGAX
Plasma_cells	Plasma_cells	Plasma_cells
HSPC	HSPC	HSPC
GMP	GMP	GMP
Neutrophils	Neutrophils	Neutrophils
Mono_Mac_DC	Monocytes	Monocytes_CD14
		Monocytes_CD16
	Macrophages	Macrophages
	DC	DC
pDC	pDC	pDC
MEMP_MEP	MEMP_MEP	MEMP_MEP
Mast_cells	Mast_cells	Mast_cells
Megakaryocytes	Megakaryocytes	Megakaryocytes
Erythroid	Erythroid_progenitors	Erythroid_Early
		Erythroid_Mid_Cycling
		Erythroid_Late_Quiescent
	Erythrocytes	Erythrocytes
Fibroblasts	Fibroblasts	Fibroblasts_BM
		Fibroblasts_Pleural
Doublets	Doublets_T_B	Doublets_T_B
		Doublets_B_T_Leuk
	Doublets_T_Myeloid	Doublets_Myeloid_T_Leuk
	Doublets_T_Mono	Doublets_T_Mono
		Doublets_Mono_T_Leuk

	Doublets_T_pDC	Doublets_pDC_T_Leuk
	Doublets_T_Megakaryocytes	Doublets_T_Megakaryocytes
	Doublets_T_Erythroid	Doublets_T_Erythroid
		Doublets_Erythroid_T_Leuk
	Doublets_T_Leuk	Doublets_T_Leuk
	Doublets_B_Myeloid	Doublets_B_Myeloid
	Doublets_B_Erthroid	Doublets_B_Erthroid

HSPC, haematopoietic stem and progenitor cell; GMP, granulocyte-monocyte progenitor; DC, dendritic cell; pDC, plasmacytoid dendritic cell; MEMP / MEP, megakaryocyte-erythrocyte-(mast cell) progenitor.

Figure 3.6

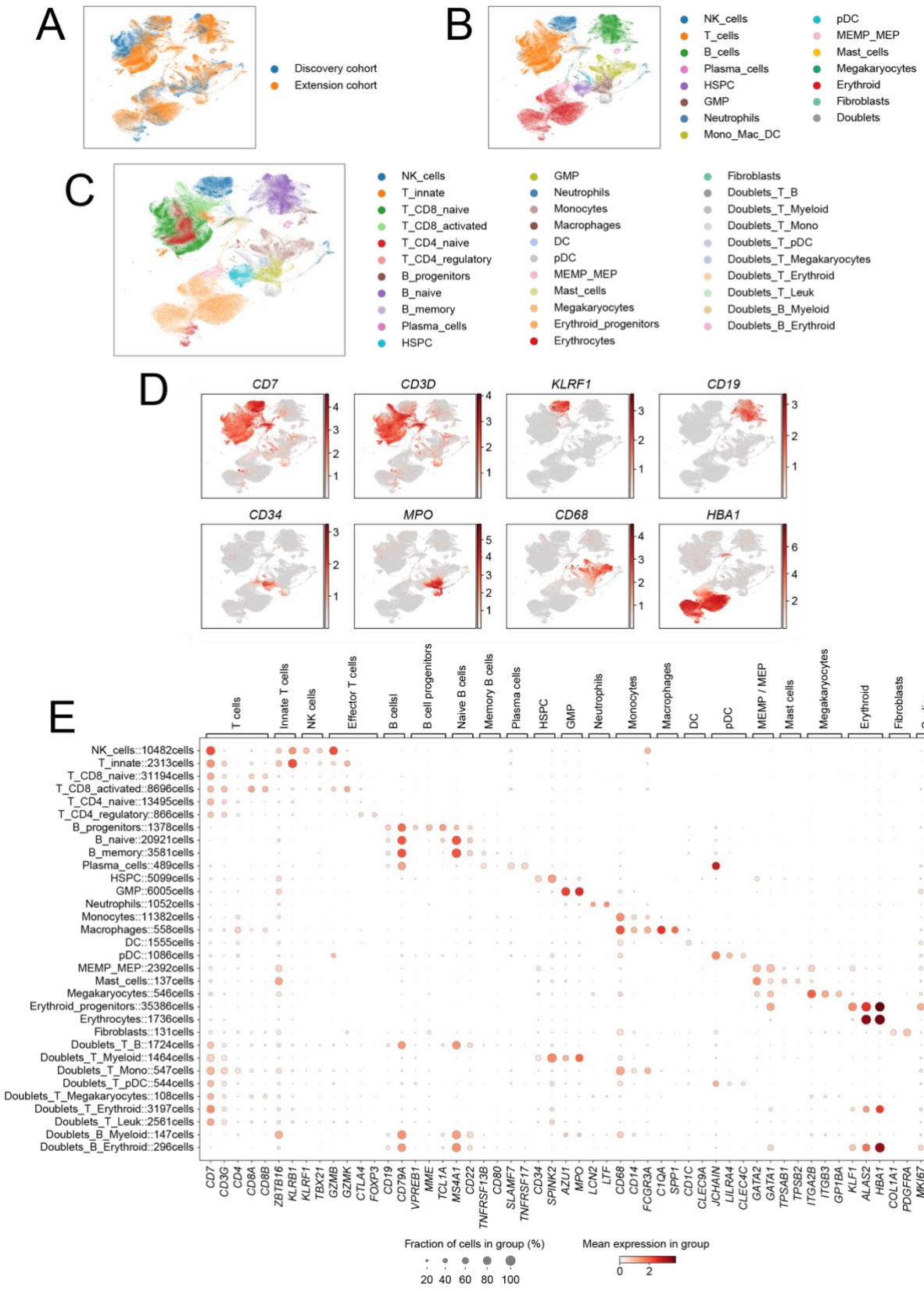

A
Discovery cohort
Extension cohort
B
NK_cells
T_cells
B_cells
Plasma_cells
HSPC
GMP
Neutrophils
Mono_Mac_DC
pDC
MEMP_MEP
Mast_cells
Megakaryocytes
Erythroid
Fibroblasts
Doublets
C
NK_cells
T_innate
T_CD8_naive
T_CD8_activated
T_CD4_naive
T_CD4_regulatory
B_progenitors
B_naive
B_memory
Plasma_cells
HSPC
GMP
Neutrophils
Monocytes
Macrophages
DC
pDC
MEMP_MEP
Mast_cells
Megakaryocytes
Erythroid_progenitors
Erythrocytes
Fibroblasts
Doublets_T_B
Doublets_T_Myeloid
Doublets_T_Mono
Doublets_T_pDC
Doublets_T_Megakaryocytes
Doublets_T_Erythroid
Doublets_T_Leuk
Doublets_B_Myeloid
Doublets_B_Erythroid
D
CD7
CD3D
KLRF1
CD19
CD34
MPO
CD68
HBA1
E
T cells
Innate T cells
NK cells
Effector T cells
B cells!
B cell progenitors
Naive B cells
Memory B cells
Plasma cells
HSPC
GMP
Neutrophils
Monocytes
Macrophages
DC
pDC
MEMP / MEP
Mast cells
Megakaryocytes
Erythroid
Fibroblasts
Cycling
NK_cells::10482cells
T_innate::2313cells
T_CD8_naive::31194cells
T_CD8_activated::8696cells
T_CD4_naive::13495cells
T_CD4_regulatory::866cells
B_progenitors::1378cells
B_naive::20921cells
B_memory::3581cells
Plasma_cells::489cells
HSPC::5099cells
GMP::6005cells
Neutrophils::1052cells
Monocytes::11382cells
Macrophages::558cells
DC::1555cells
pDC::1086cells
MEMP_MEP::2392cells
Mast_cells::137cells
Megakaryocytes::546cells
Erythroid_progenitors::35386cells
Erythrocytes::1736cells
Fibroblasts::131cells
Doublets_T_B::1724cells
Doublets_T_Myeloid::1464cells
Doublets_T_Mono::547cells
Doublets_T_pDC::544cells
Doublets_T_Megakaryocytes::108cells
Doublets_T_Erythroid::3197cells
Doublets_T_Leuk::2561cells
Doublets_B_Myeloid::147cells
Doublets_B_Erythroid::296cells
CD7 CD3G CD4 CD8A CD8B ZBTB16 KLRB1 KLRF1 TBX21 GZMB GZMK CTLA4 FOXP3 CD19 CD79A VPREB1 MME TCL1A MS4A1 CD22 TNFRSF13B CD80 SLAMF7 TNFRSF17 CD34 SPINK2 AZU1 MPO LCN2 LTF CD68 CD14 FCGR3A C1QA SPP1 CD1C CLEC9A JCHAIN LILRA4 CLEC4C GATA2 GATA1 TPSAB1 TPSB2 ITGA2B ITGB3 GP1BA KLF1 ALAS2 HBA1 COL1A1 PDGFRA MKI67 PCNA
Fraction of cells in group (%)
20 40 60 80 100
Mean expression in group
0 2

Figure 3.6: Overview of normal cells from both T-ALL cohorts.

(**A**) UMAP of all normal cells, coloured by whether they are from the discovery or extension cohorts.

(**B**) UMAP of all normal cells, coloured by level 1 cell annotation.

(**C**) UMAP of all normal cells, coloured by level 2 cell annotation.

(**D**) UMAP showing expression of marker genes: *CD7*, *CD3D*, *KLRF1*, *CD19*, *CD34*, *MPO*, *CD68, HBA1*.

(**E**) Dot plot showing expression of markers gene to support level 2 cell annotation. Dot size corresponds to fraction of cells in each group that expresses each gene and colour intensity indicates mean expression. Total cell numbers for each cell annotation are denoted in row labels.

NK cells, natural killer cells; HSPC, haematopoietic stem and progenitor cell; GMP, granulocyte-monocyte progenitor; DC, dendritic cell; pDC, plasmacytoid dendritic cell; MEMP / MEP, megakaryocyte-erythrocyte-(mast cell) progenitor.

Figure 3.7

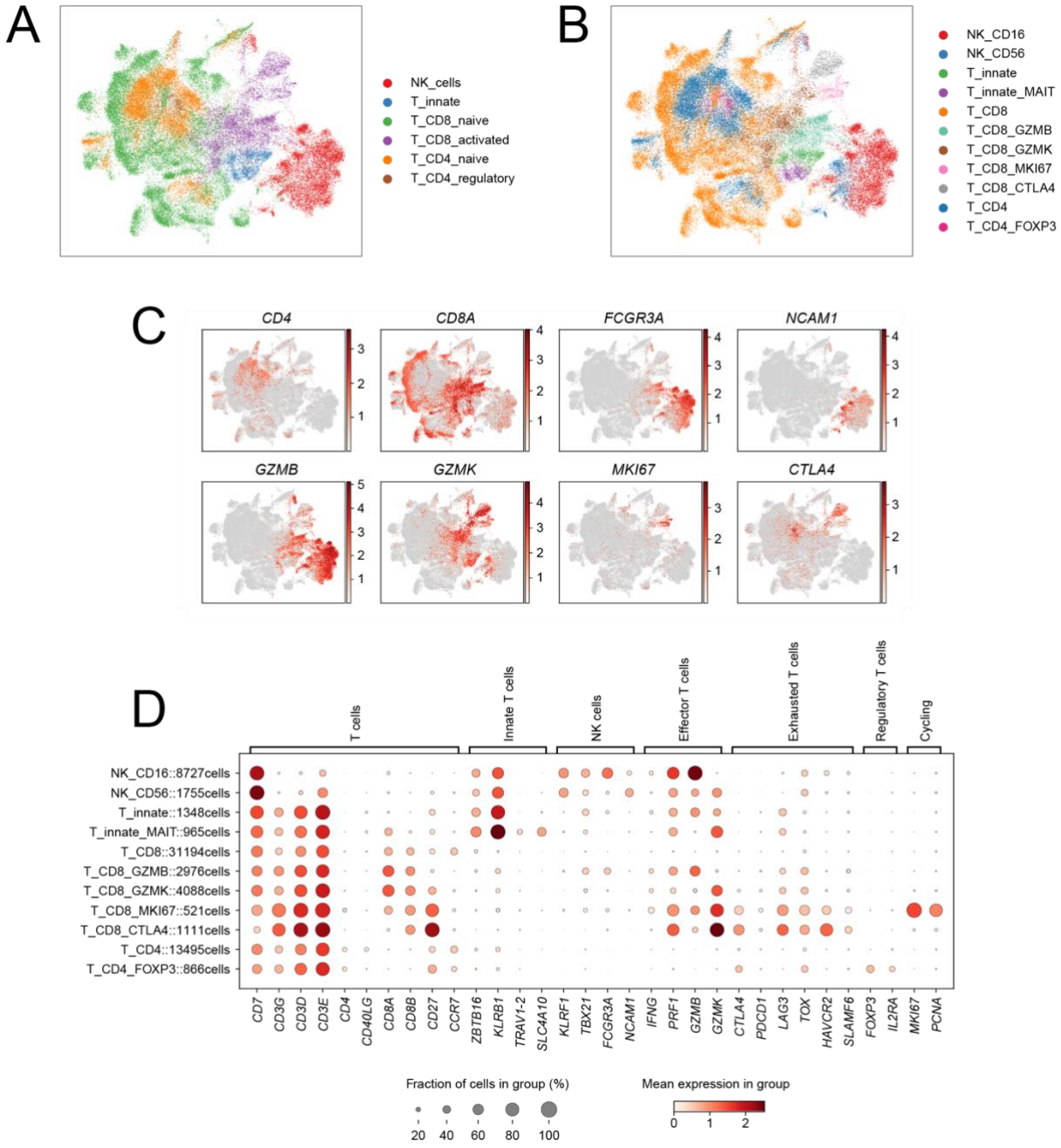

A
NK_cells
T_innate
T_CD8_naive
T_CD8_activated
T_CD4_naive
T_CD4_regulatory
B
NK_CD16
NK_CD56
T_innate
T_innate_MAIT
T_CD8
T_CD8_GZMB
T_CD8_GZMK
T_CD8_MKI67
T_CD8_CTLA4
T_CD4
T_CD4_FOXP3
C
CD4
CD8A
FCGR3A
NCAM1
GZMB
GZMK
MKI67
CTLA4
D
T cells
Innate T cells
NK cells
Effector T cells
Exhausted T cells
Regulatory T cells
Cycling
NK_CD16::8727cells
NK_CD56::1755cells
T_innate::1348cells
T_innate_MAIT::965cells
T_CD8::31194cells
T_CD8_GZMB::2976cells
T_CD8_GZMK::4088cells
T_CD8_MKI67::521cells
T_CD8_CTLA4::1111cells
T_CD4::13495cells
T_CD4_FOXP3::866cells
CD7 CD3G CD3D CD3E CD4 CD40LG CD8A CD8B CD27 CCR7 ZBTB16 KLRB1 TRAV1-2 SLC4A10 KLRF1 TBX21 FCGR3A NCAM1 IFNG PRF1 GZMB GZMK CTLA4 PDCD1 LAG3 TOX HAVCR2 SLAMF6 FOXP3 IL2RA MKI67 PCNA
Fraction of cells in group (%)
20 40 60 80 100
Mean expression in group
0 1 2

Figure 3.7: T cells and NK cells in the T-ALL microenvironment.

(**A**) UMAP of T cells and NK cells, coloured by level 2 cell annotation.

(**B**) UMAP of T cells and NK cells, coloured by level 3 cell annotation.

(**C**) UMAP showing expression of marker genes: *CD4*, *CD8A*, *FCGR3A*, *NCAM1*, *GZMB*, *GZMK*, *MKI67*, *CTLA4*.

(**D**) Dot plot showing expression of markers gene to support level 3 cell annotation. Dot size corresponds to fraction of cells in each group that expresses each gene and colour intensity indicates mean expression. Total cell numbers for each cell annotation are denoted in row labels.

Figure 3.8

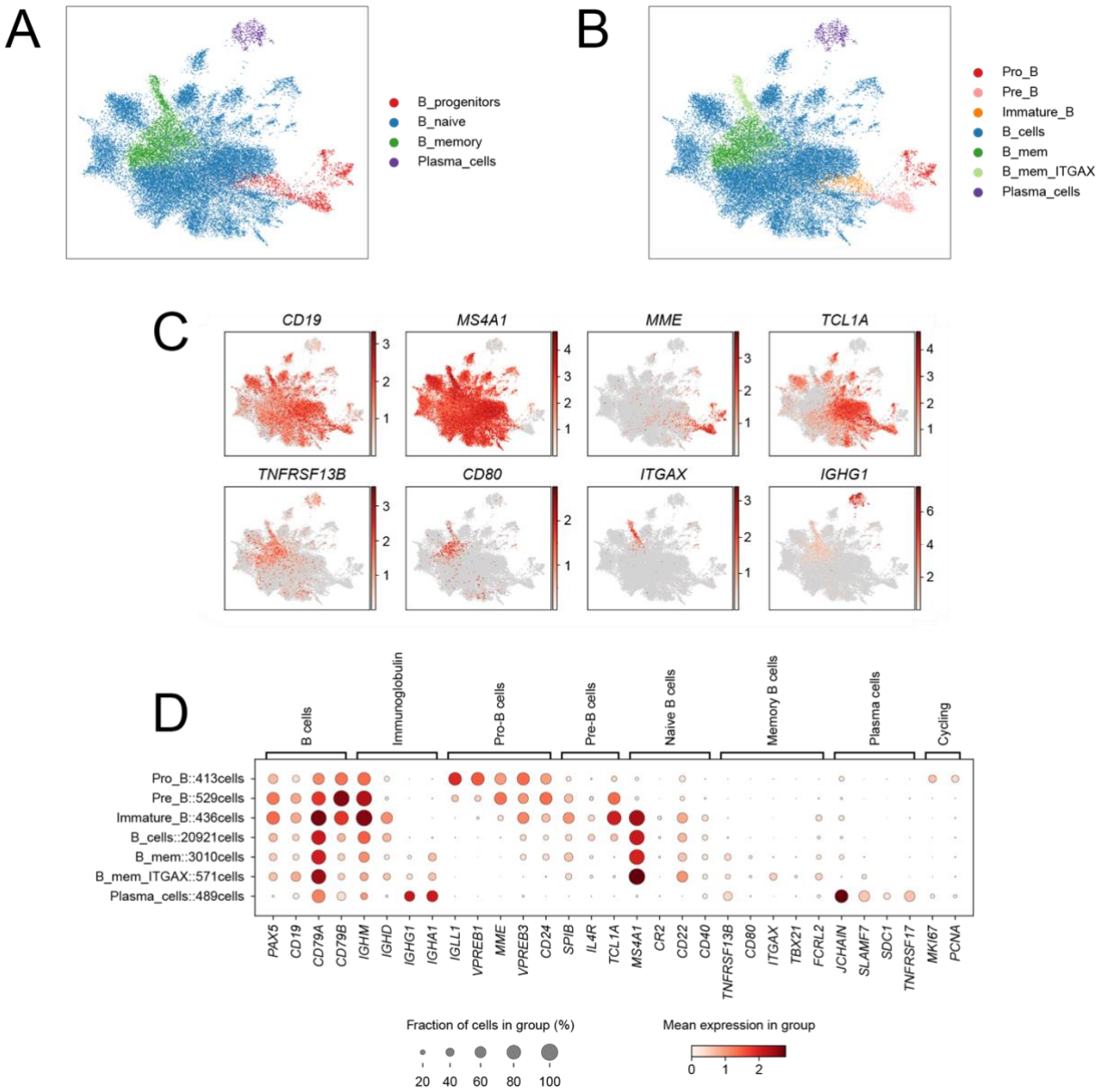

A
B_progenitors
B_naive
B_memory
Plasma_cells
B
Pro_B
Pre_B
Immature_B
B_cells
B_mem
B_mem_ITGAX
Plasma_cells
C
CD19
MS4A1
MME
TCL1A
TNFRSF13B
CD80
ITGAX
IGHG1
D
B cells
Immunoglobulin
Pro-B cells
Pre-B cells
Naive B cells
Memory B cells
Plasma cells
Cycling
Pro_B::413cells
Pre_B::529cells
Immature_B::436cells
B_cells::20921cells
B_mem::3010cells
B_mem_ITGAX::571cells
Plasma_cells::489cells
PAX5
CD19
CD79A
CD79B
IGHM
IGHD
IGHG1
IGHA1
IGLL1
VPREB1
MME
VPREB3
CD24
SPIB
IL4R
TCL1A
MS4A1
CR2
CD22
CD40
TNFRSF13B
CD80
ITGAX
TBX21
FCRL2
JCHAIN
SLAMF7
SDC1
TNFRSF17
MKI67
PCNA
Fraction of cells in group (%)
20 40 60 80 100
Mean expression in group
0 1 2

Figure 3.8: B cells and plasma cells in the T-ALL microenvironment.

(**A**) UMAP of B cells and plasma cells, coloured by level 2 cell annotation.

(**B**) UMAP of B cells and plasma cells, coloured by level 3 cell annotation.

(**C**) UMAP showing expression of marker genes: *CD19*, *MS4A1*, *MME*, *TCL1A*, *CD27*, *TNFRSF13B*, *ITGAX*, *IGHG1*.

(**D**) Dot plot showing expression of markers gene to support level 3 cell annotation. Dot size corresponds to fraction of cells in each group that expresses each gene and colour intensity indicates mean expression. Total cell numbers for each cell annotation are denoted in row labels.

Figure 3.9

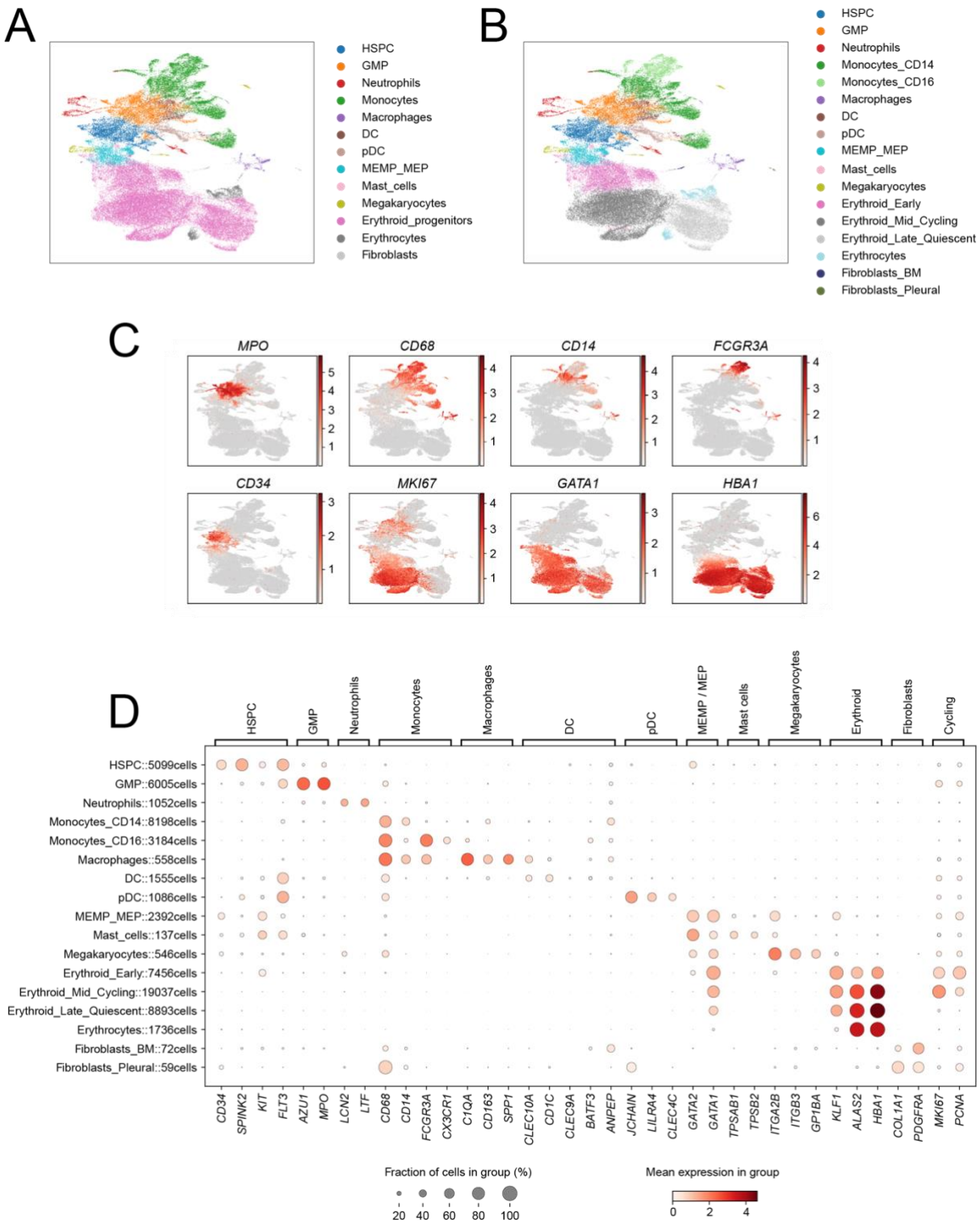

A
HSPC
GMP
Neutrophils
Monocytes
Macrophages
DC
pDC
MEMP_MEP
Mast_cells
Megakaryocytes
Erythroid_progenitors
Erythrocytes
Fibroblasts
B
HSPC
GMP
Neutrophils
Monocytes_CD14
Monocytes_CD16
Macrophages
DC
pDC
MEMP_MEP
Mast_cells
Megakaryocytes
Erythroid_Early
Erythroid_Mid_Cycling
Erythroid_Late_Quiescent
Erythrocytes
Fibroblasts_BM
Fibroblasts_Pleural
C
MPO
CD68
CD14
FCGR3A
CD34
MKI67
GATA1
HBA1
D
HSPC
GMP
Neutrophils
Monocytes
Macrophages
DC
pDC
MEMP / MEP
Mast cells
Megakaryocytes
Erythroid
Fibroblasts
Cycling
HSPC::5099cells
GMP::6005cells
Neutrophils::1052cells
Monocytes_CD14::8198cells
Monocytes_CD16::3184cells
Macrophages::558cells
DC::1555cells
pDC::1086cells
MEMP_MEP::2392cells
Mast_cells::137cells
Megakaryocytes::546cells
Erythroid_Early::7456cells
Erythroid_Mid_Cycling::19037cells
Erythroid_Late_Quiescent::8893cells
Erythrocytes::1736cells
Fibroblasts_BM::72cells
Fibroblasts_Pleural::59cells
CD34 SPINK2 KIT FLT3 AZU1 MPO LCN2 LTF CD68 CD14 FCGR3A CX3CR1 C1QA CD163 SPP1 CLEC10A CD1C CLEC9A BATF3 ANPEP JCHAIN LILRA4 CLEC4C GATA2 GATA1 TPSAB1 TPSB2 ITGA2B ITGB3 GP1BA KLF1 ALAS2 HBA1 COL1A1 PDGFRA MKI67 PCNA
Fraction of cells in group (%)
20 40 60 80 100
Mean expression in group
0 2 4

Figure 3.9: Myeloid cells, erythroid cells and fibroblasts in the T-ALL microenvironment.

(**A**) UMAP of myeloid cells, erythroid cells and fibroblasts, coloured by level 2 cell annotation.

(**B**) UMAP of myeloid cells, erythroid cells and fibroblasts, coloured by level 3 cell annotation.

(**C**) UMAP showing expression of marker genes: *MPO*, *CD68*, *CD14*, *FCGR3A*, *CD34*, *MKI67*, *GATA1*, *HBA1*.

(**D**) Dot plot showing expression of markers gene to support level 3 cell annotation. Dot size corresponds to fraction of cells in each group that expresses each gene and colour intensity indicates mean expression. Total cell numbers for each cell annotation are denoted in row labels.

HSPC, haematopoietic stem and progenitor cell; GMP, granulocyte-monocyte progenitor; DC, dendritic cell; pDC, plasmacytoid dendritic cell; MEMP / MEP, megakaryocyte-erythrocyte-(mast cell) progenitor.

3.6 Bone marrow cell composition at the end of induction treatment

The goal of induction treatment is to eliminate leukaemia blasts from the bone marrow in order to restore normal bone marrow function, in particular haematopoiesis. Hence, it is plausible that failure to eliminate blasts at the end of induction (day 28) may negatively impact normal bone marrow function and manifest in changes to the bone marrow cell composition. To this end, I analysed the cell composition in 15 samples of day 28 T-ALL bone marrow (6 responsive and 9 induction failure cases). Cell composition was calculated for each normal cell type as a percentage of the total number of normal cells (excluding leukaemia blasts and doublets). All samples had at least 2,000 normal cells and there was no difference in the total number of normal cells between induction failure and those who were responsive (**Figure 3.10A**).

Cell proportions of normal cells in the bone marrow at the end of induction treatment were analysed across all three levels of cell annotations. Significance testing for differences in cell proportion was performed using the *speckle::propeller()* function (Phipson et al., 2022). When day 28 samples of induction failure patients were compared to day 28 samples of responsive ones, I found that only memory B cells were significantly depleted and only erythrocytes were significantly enriched. The change in erythrocyte population is unlikely to be biologically meaningful as it is dependent on the red cell depletion step that is performed prior to single cell droplet preparation. In contrast, depletion of memory B cells in induction failure bone marrows may suggest attenuated antibody-mediated immunity in this group of patients. The presence of residual leukaemia blasts may have altered the bone marrow niche for memory B cells, resulting in their depletion.

To rule out the possibility that differences in memory B cell abundance was due to improper annotation of memory B cells in my data, I investigated the differences in marker gene expression of my annotation of memory B cells and naive B cells (**Figure 3.11A-C**). Memory B cells specifically expressed *TNFRSF13B*, *CD27*, *CD80* and *CD86*, whereas naive B cells specifically expressed *TCL1A* and *IL4R*. Memory B cells expressed class-switched immunoglobulin genes (*IGHG1*, *IGHG2*, *IGHA1*, *IGHA2*), which were absent in naive B cells. Memory B cells also had lower expression of non-class-switched immunoglobulin genes (*IGHM*, *IGHD*) compared to naive B cells. These results support the proper annotation of memory B cells and naive B cells in my data.

Figure 3.10A-C

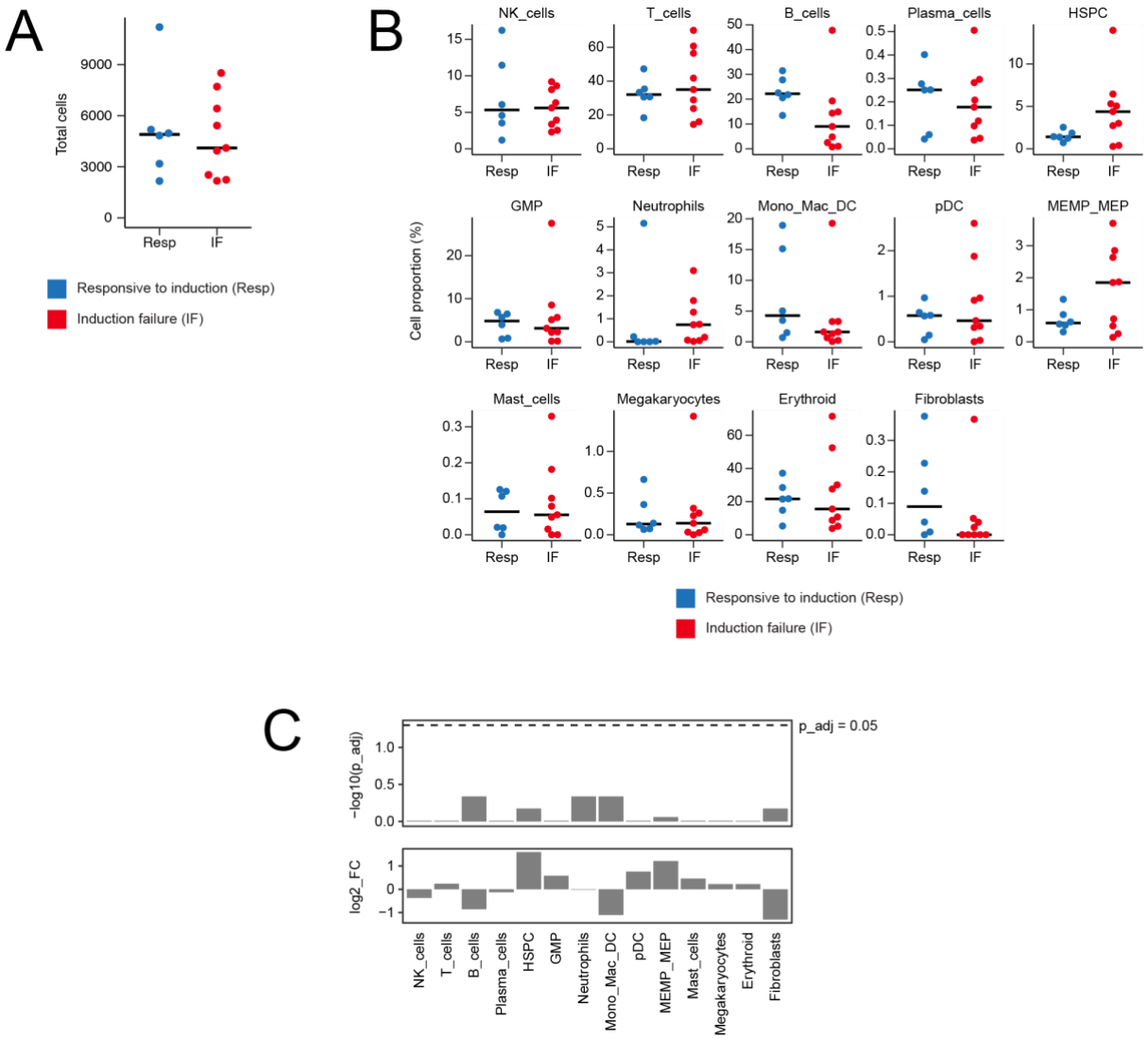


Figure 3.10D-E

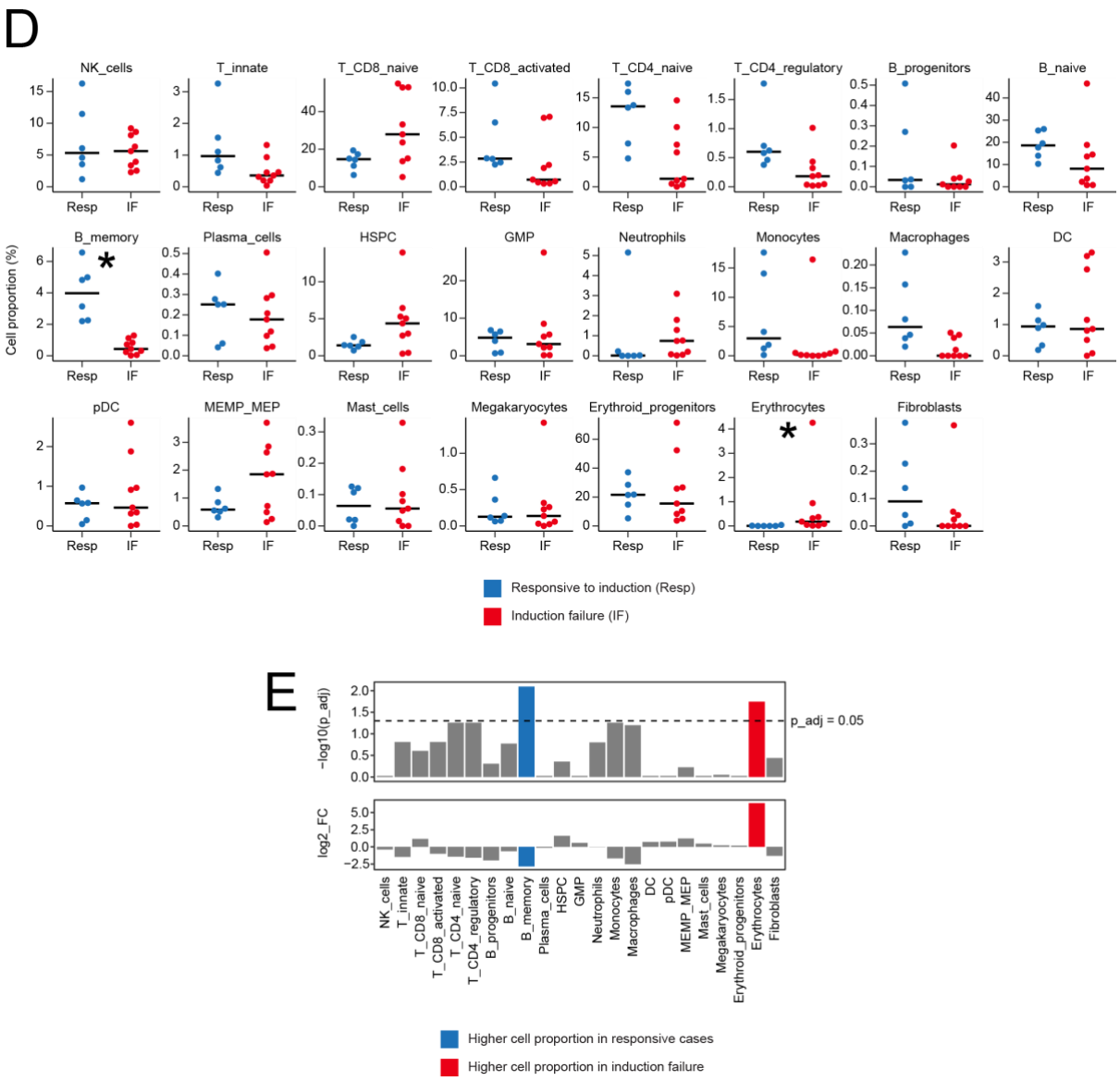


Figure 3.10F-G

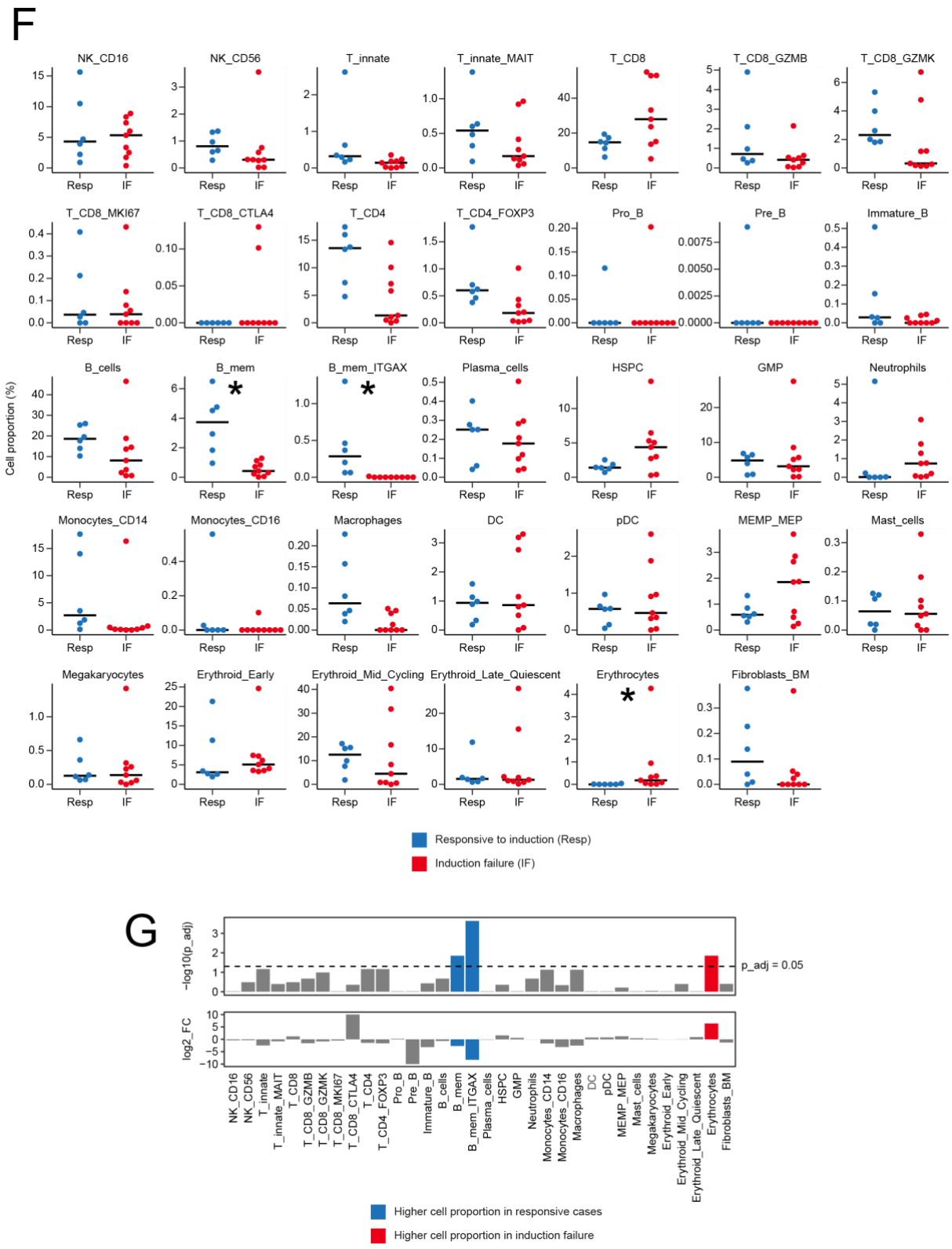


Figure 3.10: Cell proportions of normal cells in the bone marrow after induction treatment.

(**A**) Number of normal cells in scRNA-seq of the bone marrow of each patient after induction treatment. Horizontal black bars indicate median values.

(**B,D,F**) Cell proportions of each normal cell type in the day 28 bone marrow of responsive and induction failure patients. Horizontal black bars indicate median values. Cell types where there are significant differences in cell proportions are marked by asterisks (*). Level 1 (**B**), level 2 (**D**) and level 3 (**F**) cell annotation.

(**C,E,G**) Significant differences in cell proportions were identified with the *speckle::propeller()* function, which outputs fold changes and *P* values adjusted for multiple hypothesis testing by Benjamini-Hochberg correction. Dashed line indicates the 0.05 false discovery rate cutoff. Coloured bars indicate significant changes in cell proportions when induction failure patients are compared to responsive patients: increased (red) and decreased (blue) in induction failure. Level 1 (**C**), level 2 (**E**) and level 3 (**G**) cell annotation.

Figure 3.11

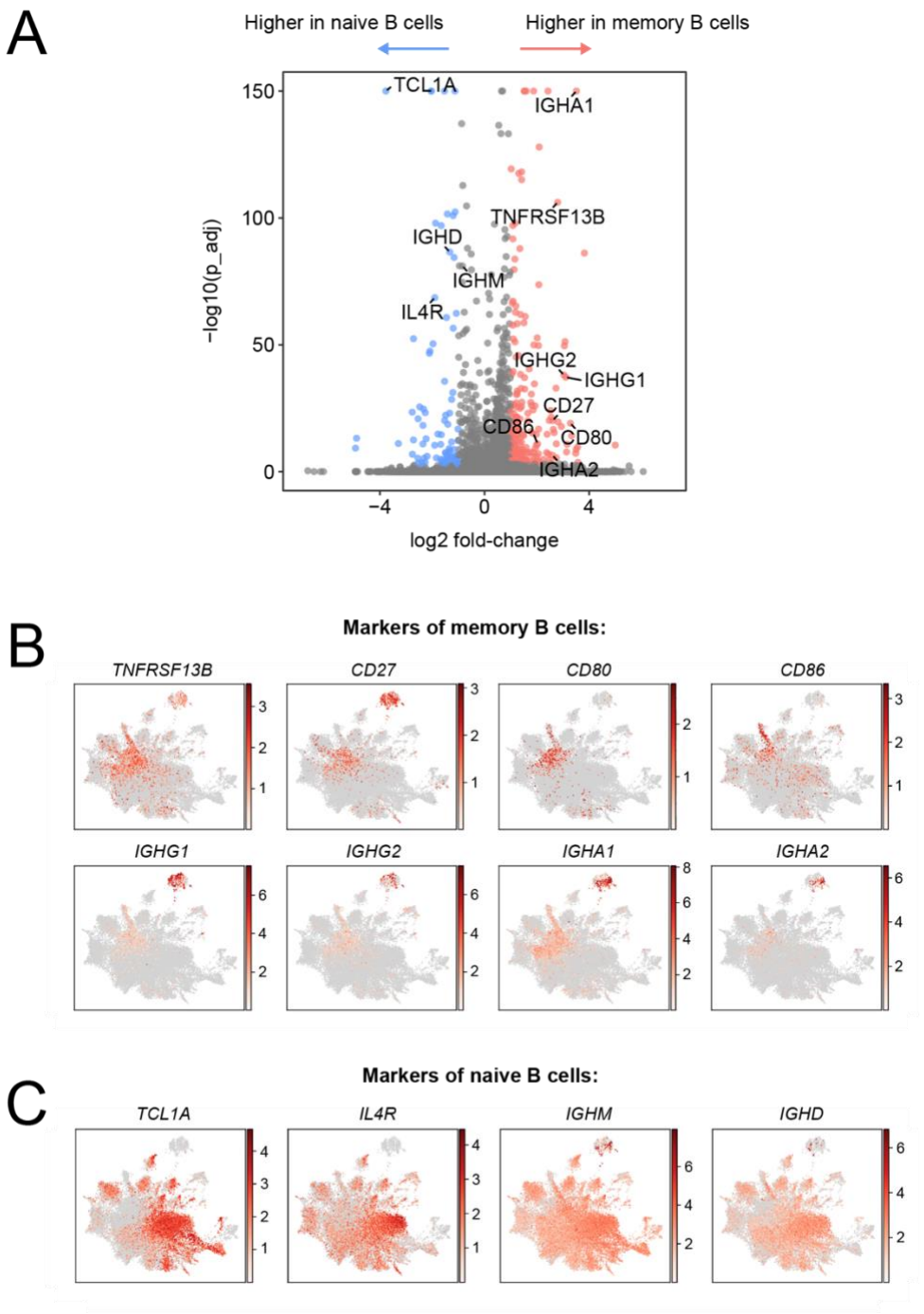

A
Higher in naive B cells
Higher in memory B cells
TCL1A
IGHA1
TNFRSF13B
IGHD
IGHM
IL4R
IGHG2
IGHG1
CD27
CD86
CD80
IGHA2
−log10(p_adj)
log2 fold-change
B
Markers of memory B cells:
TNFRSF13B
CD27
CD80
CD86
IGHG1
IGHG2
IGHA1
IGHA2
C
Markers of naive B cells:
TCL1A
IL4R
IGHM
IGHD

Figure 3.11: Cell annotation of memory B cells.

(**A**) Volcano plot showing marker genes which distinguish memory B cells from naive B cells (level 2 cell annotation), derived with the *scanpy.tl.rank_genes_groups()* function. *P* values were adjusted for multiple hypothesis testing by Benjamini-Hochberg correction and cut off at 10^{-150} lower limit for plotting. Significantly upregulated genes (in red, adjusted *P* value < 10^{-3} and log2 fold change > 1) and downregulated genes (in blue, adjusted *P* value < 10^{-3} and log2 fold change < −1) are highlighted in colour.

(**B**) UMAP showing expression of marker genes of memory B cells: *TNFRSF13B*, *CD27*, *CD80*, *CD86*, *IGHG1*, *IGHG2*, *IGHA1*, *IGHA2*.

(**C**) UMAP showing expression of marker genes of naive B cells: *TCL1A*, *IL4R*, *IGHM*, *IGHD*.

3.7 Discussion

In this chapter, I generated and analysed a single cell transcriptome atlas of T-ALL, comprising 84 samples from 55 individuals, combined from two sequencing cohorts. I performed detailed cell annotation on almost 550,000 high quality cells, which includes around 380,000 leukaemia blasts and 170,000 normal cells. This represents a rich dataset that will further our understanding of T-ALL biology in ways not previously attainable by flow cytometry or bulk sequencing.

By comparing transcriptomes of T-ALL blasts to stages of normal T cell development, I found that T-ALL blasts most strongly resembled DN thymocytes, which are found in the thymus. However, T-ALL blasts with the ETP immunophenotype, or of a genomic subtype that is associated with early differentiation state, exhibit signals of haematopoietic and lymphoid progenitors in the bone marrow, which have not yet migrated to the thymus. It is currently not known whether T-ALL arises from progenitors cells in the thymus or bone marrow. Transcriptomic similarity of T-ALL blasts to thymic and bone marrow progenitors suggests that both could be cellular sources of T cell leukaemia. However, transcriptomic similarity is not definitive evidence of cell of origin, as cancer cells, particularly paediatric cancers, can undergo de-differentiation to an earlier cell state (Kildisiute et al., 2024). Finally, for one individual, P018, their blasts exhibited features of innate lymphoid cells as opposed to thymocyte stages, suggesting a departure from conventional T cell lineages.

Non-malignant cells (i.e. normal cells) make up around one-third of cells in my single cell atlas of T-ALL. I annotated T cells, B cells, myeloid lineages and erythroid lineages to a high degree of granularity. Potentially interesting cells which I identified include: various effector and exhausted T cell states, *ZBTB16*+ unconventional T cells and MAIT cells, memory B cells and *ITGAX*+ memory B cells, and differentiation stages of myeloid and erythroid development. It is possible that T cells and B cells of the adaptive immune system may have responded to antigens from T-ALL blasts. Evidence of clonally expanded T cells and B cells, from sequencing of their antigen receptors, will be consistent with an adaptive immune response, although these clonally expanded T cells and B cells could already have existed prior to leukaemia occurrence.

The presence of significant residual blasts after induction treatment (i.e. induction failure) may have an effect on the composition of normal cells in the bone marrow. I found that memory B cells are relatively depleted in day 28 bone marrow of induction failure patients compared to day 28 bone marrow of responsive patients. It is possible that residual blasts in the bone

marrow may alter the niche for memory B cells, resulting in their relative depletion. Examination of other T-ALL scRNA-seq datasets which contain day 28 samples, or flow cytometry for memory B cells in day 28 bone marrow samples of an independent T-ALL cohort, may validate this finding.

While I focused in this chapter on defining the landscape of T-ALL blasts from single cell transcriptome sequencing, clinically relevant insights are more likely to arise from interrogating residual blasts after treatment, for these represent refractory blasts by definition. Comparing blasts at the end of induction to diagnostic blasts may yield insights into the transcriptomic features of refractory T-ALL, which I present in **Chapter 4**. Finally, refractory blasts may have evolved from diagnostic blasts by acquiring resistance mutations, or they may represent a distinct independent lineage that is intrinsically resistant. I present in **Chapter 5** phylogenetic analysis of T-ALL blasts to reconcile between these two competing hypotheses.

Chapter 4: Transcriptomic features of refractory T-ALL

4.1 Introduction

T-ALL is a more aggressive disease, with higher rates of induction failure, relapse and death compared to B-ALL (Goldberg et al., 2003). Furthermore, unlike in B-ALL, there are currently no clinical, phenotypic or genetic features that can be reliably used to risk stratify children with T-ALL at the point of diagnosis. Although various markers have previously been proposed, including specific genetic mutations or immunophenotype categories, none have proven to be sufficiently robust for use in current clinical practice (Conter et al., 2016; Coustan-Smith et al., 2009; Jenkinson et al., 2013; Petit et al., 2018; Seki et al., 2017; Taj et al., 2022). The earliest reliable indicator of high-risk T-ALL is induction failure, defined as the presence of ≥5% leukaemia cells (blasts) in the bone marrow after four weeks of induction treatment (O'Connor et al., 2017). Induction failure occurs in about 10% of T-ALL cases, but account for around half of mortality cases at 5 years, making refractory disease an important cause of T-ALL death (Raetz et al., 2023; Teachey and O'Connor, 2020).

Large scale genomic studies have provided substantial insights into the genetic and phenotypic basis of T-ALL (Brady et al., 2022; Liu et al., 2017; O'Connor et al., 2023). Most recently, combined genomic and transcriptomic profiling of >1,300 children from the COG AALL0434 cohort, the largest T-ALL sequencing cohort to date, has defined 15 subtypes of T-ALL (Pölönen et al., 2024). However, beyond providing detailed genomic classification, the nature and cause of refractory T-ALL have eluded these studies. In parallel, the early T cell precursor (ETP) immunophenotype emerged as a distinct T-ALL subtype with poor prognosis (Coustan-Smith et al., 2009). However, data from modern trials showed no significant difference in survival outcome for ETP T-ALL (Patrick et al., 2014; Wood et al., 2023). It is possible that refractory T-ALL is driven by a cancer cell state independent of genetic mutation and not defined by previously known immunophenotypic markers. Single cell transcriptome sequencing provides unbiased, quantitative, molecular readouts at the resolution of individual cells, which may allow for the definition of cancer cell states with higher precision than bulk transcriptome sequencing. Accordingly, I set out to apply single cell transcriptome sequencing to investigate if a specific cell state may account for refractory T-ALL.

4.2 Discovery of *ZBTB16* as a marker of refractory T-ALL

I analysed a discovery cohort of 21 T-ALL patients (**Figure 4.1A**), which I described earlier in **Chapter 3**. From single cell RNA sequencing (scRNA-seq) of this cohort, I obtained 216,507 high quality cells in total, comprising 160,327 leukaemia blasts and 56,180 normal cells. Upon close examination of a Uniform Manifold and Projection (UMAP) plot of these cells (**Figure 4.1B**), I observed that although day 0 and day 28 blasts of each patient tended to cluster next to each other, in two patients, P058 and P030, their day 28 blasts clustered away from their respective day 0 blasts and instead co-located next to each other. These observations suggest that for P058 and P030, there may be a major shift in transcriptomic state of their blasts between diagnosis and post-induction timepoints. Moreover, their co-localisation suggests transcriptomic similarity between day 28 blasts from these two patients. The unique clustering of blasts from P058 and P030 prompted further investigation.

The transcriptomic features of day 28 blasts from P058 and P030 were investigated through two approaches. For the first approach, I looked for genes which were upregulated in the day 28 blasts for each of the two patients compared to their respective day 0 blasts, by using the *scanpy.tl.rank_genes_group()* function (Wolf et al., 2018), which performs a Wilcoxon rank-sum test (**Figure 4.2A**). Remarkably, I found that the most upregulated gene in day 28 blasts from P058 was *ZBTB16*, a transcription factor that is expressed in and required for development of unconventional T cells (Alonzo and Sant'Angelo, 2011). *KLRB1*, another marker of unconventional T cells (Fergusson et al., 2011), was also upregulated in day 28 blasts of P058. In contrast, *DNTT*, which is involved in development of conventional T cells, appeared upregulated in day 0 blasts of P058. In the case of P030, *ZBTB16* appears to be only slightly enriched in day 28 blasts compared to day 0 blasts.

Plotting *ZBTB16* expression on a UMAP of leukaemia blasts from P058 (**Figure 4.2B**) revealed a sharp contrast between its high expression in day 28 blasts compared to its near-complete absence in day 0 blasts, except for a small cluster of *ZBTB16*+ cells which constitute 1% of the day 0 blasts. A similar plot for P030 showed most day 0 and day 28 blasts being *ZBTB16*+, with slightly higher expression level at day 28.

For the second approach, genes upregulated in day 28 blasts for each of the two patients compared to all other blasts were obtained using the *scanpy.tl.rank_genes_group()* function (**Figure 4.2C**). Intersecting the top 100 upregulated genes of day 28 blasts from P058 and P030 yielded 8 genes, among which *ZBTB16* showed the highest log-fold increase. Overall, both analyses suggest that *ZBTB16* is associated with the day 28 blasts of P058 and P030.

4.3 *ZBTB16* expression in blasts is associated with refractory T-ALL

Next, I interrogated *ZBTB16* expression across leukaemia blasts in scRNA-seq samples from the discovery cohort of 21 T-ALL patients (**Figure 4.2D**). Expression was considered by measuring *ZBTB16* mRNA transcript level in individual blasts, or the fraction of blasts which express *ZBTB16* mRNA transcripts. Furthermore, to probe the possibility of heterogeneous clusters of cancer cells, Leiden clustering was performed on blasts within each scRNA-seq sample and the median mRNA transcript level in each cluster of blasts was calculated. *ZBTB16* expression was largely absent in blasts derived from day 0 samples of patients who responded to induction chemotherapy. In contrast, *ZBTB16*+ blasts and clusters of blasts were found in more than half of day 0 and day 28 samples of patients with induction failure. Notably, a cluster of blasts which was positive for *ZBTB16* was found in the day 0 sample of P058, that was otherwise mostly negative.

Day 28 blasts in patients with induction failure represent refractory blasts, as they were not killed by induction chemotherapy. Four patients with induction failure had a high proportion of *ZBTB16*+ blasts at both day 0 and day 28 (P030, P018, P021 and P029) (**Figure 4.2E-F**), consistent with *ZBTB16* expression being associated with a refractory phenotype. The percentage of *ZBTB16*+ blasts increased dramatically from 0.67% of blasts at day 0 to 97.6% at day 28 in P058, suggesting a possible “clonal sweep” by *ZBTB16*+ blasts, which I will elaborate further in **Chapter 5**. Two patients (P016 and P028) had low fractions of *ZBTB16*+ blasts at both timepoints, and in one patient (P084) it decreased following induction. When examining *ZBTB16* mRNA transcript level in only the *ZBTB16*+ blasts from these eight induction failure patients across both timepoints (**Figure 4.2F**), *ZBTB16* levels were found to consistently increase from day 0 to day 28.

These results suggest that *ZBTB16* expression in leukaemia blasts is associated with refractory T-ALL and that measuring *ZBTB16* in blasts could potentially be a strategy for risk stratification in the clinic. A pilot study by Dr Rebecca Thomas and Dr David O’Connor at Great Ormond Street Hospital (GOSH) demonstrated the feasibility of flow cytometry with intracellular staining for measuring ZBTB16 protein levels in blasts (**Figure 4.2G**).

Figure 4.1

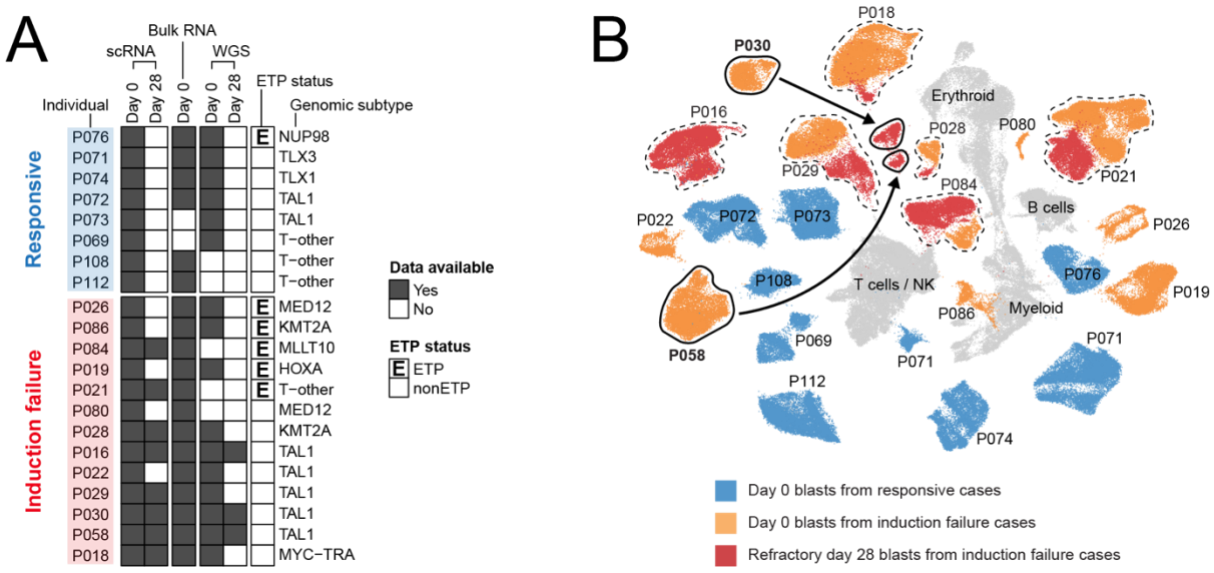


Figure 4.1: Overview of discovery cohort scRNA-seq.

(**A**) Heatmap indicating sequencing data generated for each patient from the discovery cohort and their clinical information (induction outcome, ETP status, genomic subtype).

(**B**) UMAP of 216,507 cells from the discovery cohort, including 160,327 leukaemia blasts (coloured) and 56,180 normal cells (grey). Day 0 blasts from patients who responded to induction treatment (blue); day 0 blasts from patients with induction failure (orange); day 28 refractory blasts from patients with induction failure (red). Circles and arrows highlight that day 28 blasts from patients P030 and P058 clustered separately from their respective day 0 blasts on the UMAP.

scRNA, single cell RNA sequencing; WGS, whole genome sequencing; ETP, early T cell precursor.

Figure 4.2

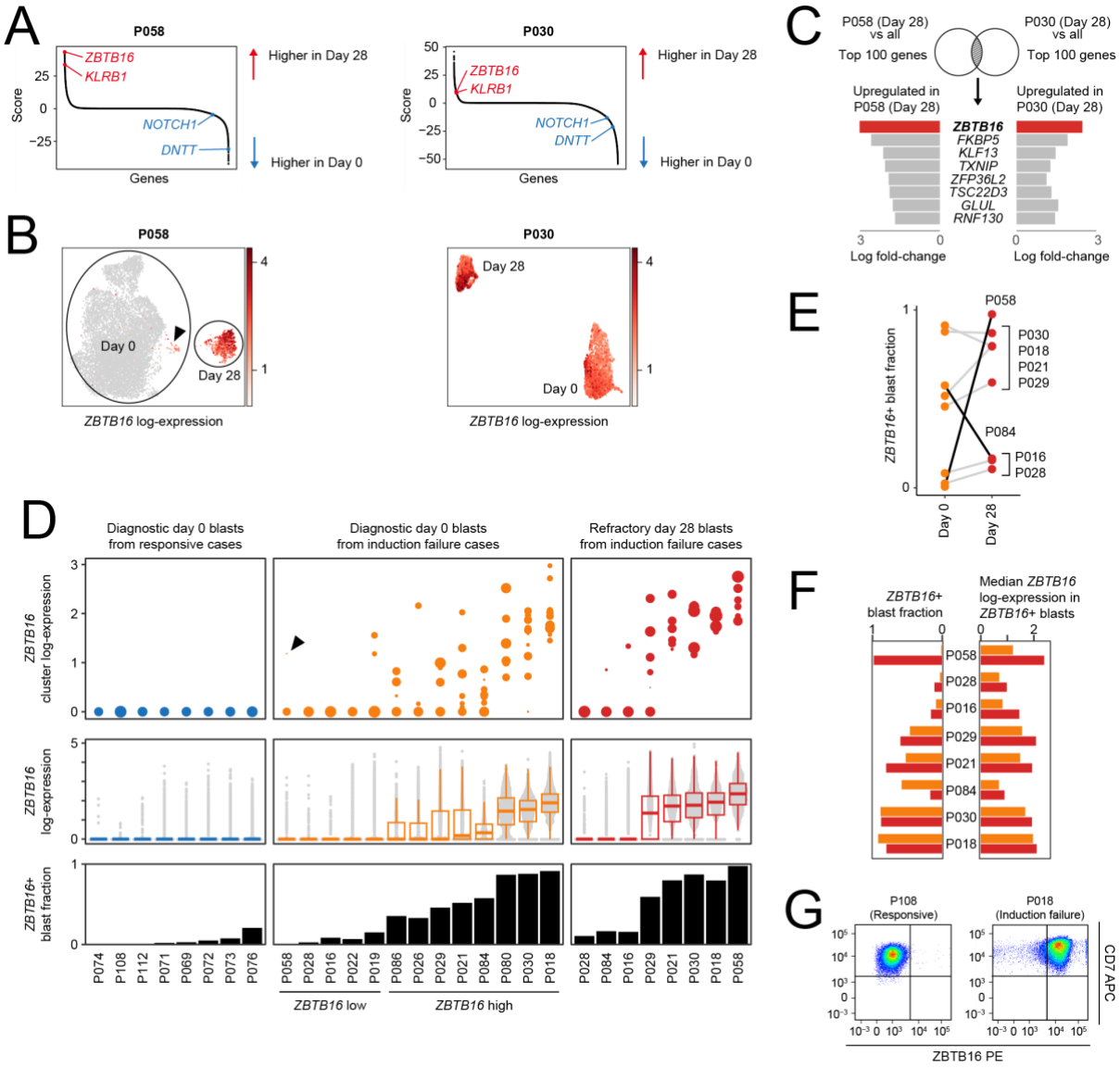

A
P058
P030
ZBTB16
KLRB1
NOTCH1
DNTT
Score
Genes
Higher in Day 28
Higher in Day 0
B
P058
P030
Day 0
Day 28
ZBTB16 log-expression
C
P058 (Day 28) vs all
P030 (Day 28) vs all
Top 100 genes
Upregulated in P058 (Day 28)
Upregulated in P030 (Day 28)
ZBTB16
FKBP5
KLF13
TXNIP
ZFP36L2
TSC22D3
GLUL
RNF130
Log fold-change
D
Diagnostic day 0 blasts from responsive cases
Diagnostic day 0 blasts from induction failure cases
Refractory day 28 blasts from induction failure cases
ZBTB16 cluster log-expression
ZBTB16 log-expression
ZBTB16+ blast fraction
P074 P108 P112 P071 P069 P072 P073 P076
P058 P028 P016 P022 P019
P086 P026 P029 P021 P084 P080 P030 P018
P028 P084 P016 P029 P021 P030 P018 P058
ZBTB16 low
ZBTB16 high
E
ZBTB16+ blast fraction
Day 0
Day 28
P058
P030
P018
P021
P029
P084
P016
P028
F
ZBTB16+ blast fraction
Median ZBTB16 log-expression in ZBTB16+ blasts
P058
P028
P016
P029
P021
P084
P030
P018
G
P108 (Responsive)
P018 (Induction failure)
CD7 APC
ZBTB16 PE

Figure 4.2: Discovery of *ZBTB16* as a marker of refractory T-ALL.

(**A**) Identification of genes which are enriched in day 28 blasts compared to day 0 blasts (and vice versa) for patients P058 and P030, using the *scanpy.tl.rank_genes_group()*, which calculates an enrichment Z-score. Genes with positive scores have higher expression in day 28 blasts, whereas those with negative scores have higher expression in day 0 blasts. Notably, for patient P058, *ZBTB16* had the highest enrichment score out of >30,000 genes in day 28 blasts compared to day 0 blasts.

(**B**) UMAP showing *ZBTB16* expression in day 0 and day 28 samples of P058 and P030. Arrowhead indicates a small cluster of *ZBTB16*+ blasts in the day 0 sample of P058.

(**C**) The top 100 upregulated genes of the day 28 blasts of P058 and P030 compared to blasts all other samples were identified using the *scanpy.tl.rank_genes_group()* function from the Scanpy package. Overlapping these two sets of top 100 genes yielded 8 genes, of which *ZBTB16* had the highest log-fold increase.

(**D**) *ZBTB16* expression in blasts across all samples in the discovery cohort. Top: Median *ZBTB16* expression of each cluster of blasts within each sample; size of circle indicates cluster size; arrowheads indicate small clusters of day 0 blasts from P058 expressing *ZBTB16*. Middle: Box plot of *ZBTB16* expression at single cell resolution. Box plots show the first and third quartiles (boxes), as well as median values (central lines). Whiskers extend to the most extreme values within 1.5 times the interquartile range above and below the boxes. Bottom: Fraction of blasts expressing *ZBTB16*.

(**E**) Line plot showing change in *ZBTB16*+ blast fraction between day 28 and day 0 samples of the same patient.

(**F**) *ZBTB16* expression in blasts from induction failure patients with paired day 0 (orange) and day 28 samples (red). Left: Fraction of blasts expressing *ZBTB16*. Right: Median *ZBTB16* expression in *ZBTB16*+ blasts.

(**G**) Flow cytometry measurement of ZBTB16 in blasts from diagnostic samples of responsive patient P108 and induction failure patient P018.

4.4 Validation of *ZBTB16* signal in bulk transcriptomes of T-ALL

Single cell transcriptomes from 21 patients suggest that *ZBTB16* expression in blasts is associated with refractory T-ALL. To validate these findings in larger cohorts, I looked towards bulk transcriptomes. In addition to measuring *ZBTB16* mRNA transcript levels, I also scored bulk transcriptomes using a 29-gene module which represents the *ZBTB16*+ refractory T-ALL phenotype. This gene module was derived by Dr Holly Whitfield.

First, I evaluated the validity of the gene module in single cell and bulk transcriptomes of the discovery cohort. The 29 individual genes in this gene module were strongly expressed only in *ZBTB16*+ blasts from induction failure patients in the single cell transcriptome data (**Figure 4.3A**). Bulk transcriptome sequencing was performed on diagnostic samples for 19 out of 21 patients in the discovery cohort (not available for P073 and P069). Scoring using the gene module, but excluding *ZBTB16* itself, showed a correlation with *ZBTB16* mRNA transcript levels in these bulk transcriptomes (**Figure 4.3B**), indicating the validity of the gene module. Furthermore, diagnostic samples which had high *ZBTB16* expression in single cell transcriptome sequencing could also be identified in bulk transcriptome sequencing data, using either *ZBTB16* levels alone or the module score (**Figure 4.3C**). These results provide evidence that both *ZBTB16* mRNA transcript levels and the module score are suitable proxy measures to quantify the amount of *ZBTB16*+ blasts in bulk transcriptomes.

Next, I looked at diagnostic bulk transcriptomes from a cohort of 52 patients with T-ALL provided by colleagues at the Princess Maxima Center in The Netherlands. Diagnostic samples from patients with induction failure (day 28 MRD $\geq$ 5%) showed significantly higher *ZBTB16* mRNA transcript levels as well as module score, compared to the responsive group (**Figure 4.3D**).

Finally, I examined 1,335 diagnostic bulk transcriptomes from the Children's Oncology Group (COG) AALL0434 trial, which represents the largest sequencing cohort to date (Pölönen et al., 2024). I restricted my analysis to 1,175 samples with >60% blast content and for which induction outcome was known. As expected, induction failure patients had significantly higher *ZBTB16* mRNA transcript levels and module scores (**Figure 4.3E**). Among the responsive patients, higher *ZBTB16* levels and higher module scores in diagnostic bulk transcriptomes were associated with elevated residual disease at the end of induction.

4.5 *ZBTB16* expression is associated with poor survival in T-ALL

Given that induction failure is predictive of worse survival in T-ALL (O'Connor et al., 2017), I sought to investigate the relationship between the presence of *ZBTB16*+ blasts and survival outcome in the COG AALL0434 cohort data (Pölönen et al., 2024). Patients with high *ZBTB16* mRNA transcript level or high module score were found to have worse overall survival and event free survival (**Figure 4.4A**).

The ETP immunophenotype has initially emerged as a predictor of high-risk T-ALL (Coustan-Smith et al., 2009). However, subsequent studies using modern chemotherapy protocols showed no significant difference in survival (Patrick et al., 2014; Wood et al., 2023). Multivariate analysis by Cox proportional hazards modelling of *ZBTB16* expression with ETP status revealed that *ZBTB16* has a statistically significant effect on survival in the COG AALL0434 cohort data, unlike ETP status (**Figure 4.4B-C**). This was regardless of whether *ZBTB16* was considered as a categorical variable or continuous variable, and regardless of whether the near-ETP immunophenotype was included in the analysis.

During the course of my PhD, bone marrow progenitor-like (BMP-like) blasts have been proposed by another group as a high-risk T-ALL blast subtype, independent of ETP immunophenotype (Xu et al., 2024). Multivariate analysis by Cox proportional hazards demonstrated that our *ZBTB16* gene module has a superior predictive ability on survival in the COG AALL0434 cohort data, compared to any of the BMP-like gene signatures proposed by the authors (**Figure 4.4D**). Overall, analysis of T-ALL bulk transcriptomes from the COG AALL0434 cohort provided evidence that the presence of *ZBTB16*+ blasts at diagnosis not only portends induction failure, but more importantly is predictive of poorer survival.

Figure 4.3

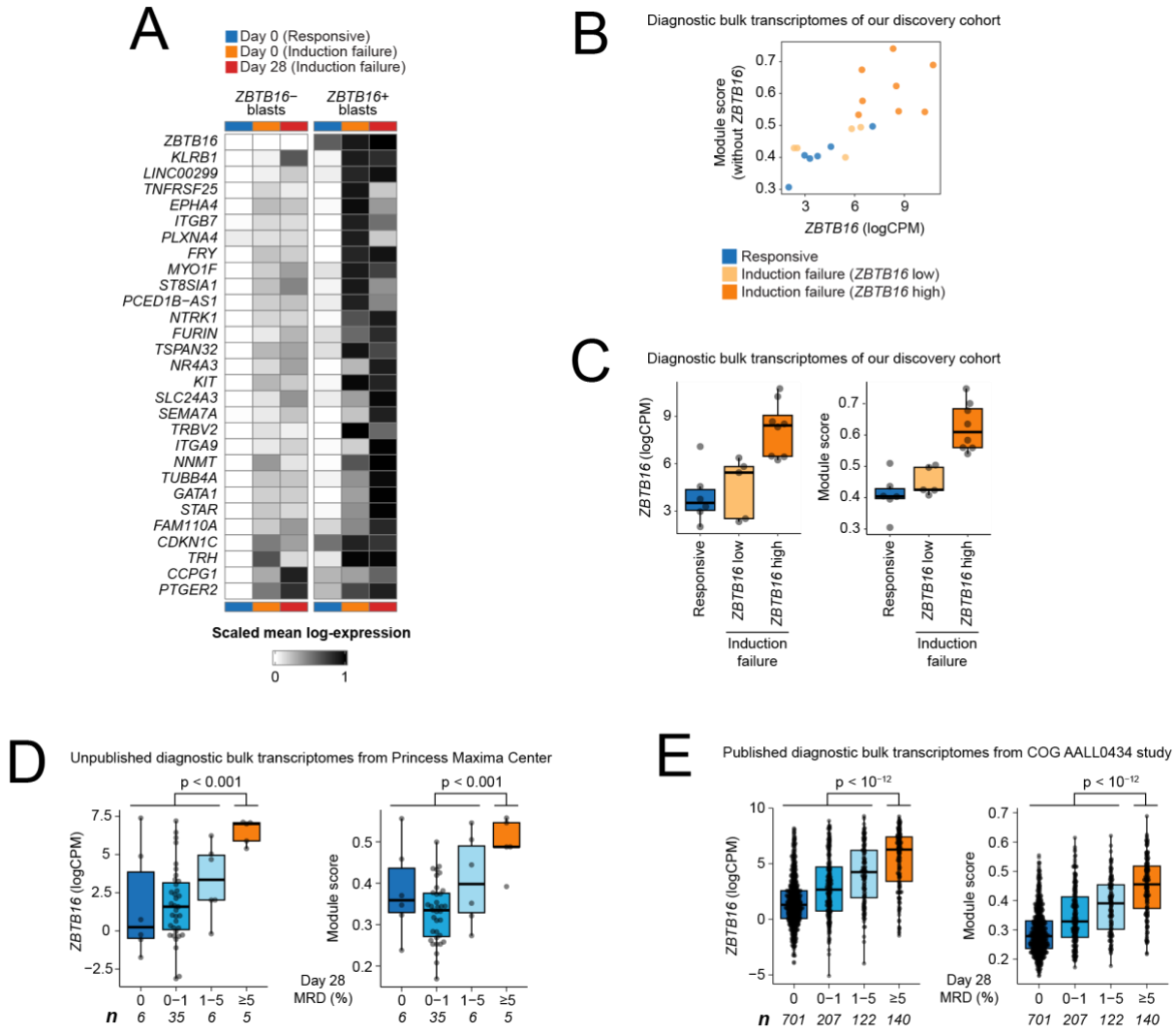

A
Day 0 (Responsive)
Day 0 (Induction failure)
Day 28 (Induction failure)
ZBTB16− blasts
ZBTB16+ blasts
ZBTB16
KLRB1
LINC00299
TNFRSF25
EPHA4
ITGB7
PLXNA4
FRY
MYO1F
ST8SIA1
PCED1B−AS1
NTRK1
FURIN
TSPAN32
NR4A3
KIT
SLC24A3
SEMA7A
TRBV2
ITGA9
NNMT
TUBB4A
GATA1
STAR
FAM110A
CDKN1C
TRH
CCPG1
PTGER2
Scaled mean log-expression
0
1
B
Diagnostic bulk transcriptomes of our discovery cohort
Module score (without ZBTB16)
ZBTB16 (logCPM)
Responsive
Induction failure (ZBTB16 low)
Induction failure (ZBTB16 high)
C
Diagnostic bulk transcriptomes of our discovery cohort
ZBTB16 (logCPM)
Module score
Responsive
ZBTB16 low
ZBTB16 high
Induction failure
D
Unpublished diagnostic bulk transcriptomes from Princess Maxima Center
p < 0.001
ZBTB16 (logCPM)
Module score
Day 28 MRD (%)
n
E
Published diagnostic bulk transcriptomes from COG AALL0434 study
p < 10−12
ZBTB16 (logCPM)
Module score
Day 28 MRD (%)
n

Figure 4.3: Validation of *ZBTB16* signal in bulk transcriptomes.

(**A**) A 29-gene module for the refractory *ZBTB16*+ leukaemia cell state was derived using a pseudobulk differential gene expression analysis approach. The expression of individual genes in this gene module is plotted as a heatmap for *ZBTB16*+ and *ZBTB16*− blasts, taken from day 0 samples of responsive patients (blue), from day 0 samples of induction failure patients (orange), and from day 28 samples of induction failure patients (red). Min-max scaling was performed across each row on the heatmap.

(**B**) Scatter plot showing the correlation between *ZBTB16* expression and our *ZBTB16* module score (without the *ZBTB16* gene itself) in bulk transcriptomes of our discovery cohort. Dots were coloured by whether a sample was from a responsive patient (blue) or a patient with induction failure that was identified as *ZBTB16* low (yellow) or *ZBTB16* high (orange) based on **Figure 4.2D**.

(**C**) Box plots showing *ZBTB16* expression and module score in bulk transcriptomes of our discovery cohort, grouped by whether a sample was from a responsive patient (blue) or a patient with induction failure that was identified as *ZBTB16* low (yellow) or *ZBTB16* high (orange) based on **Figure 4.2D**.

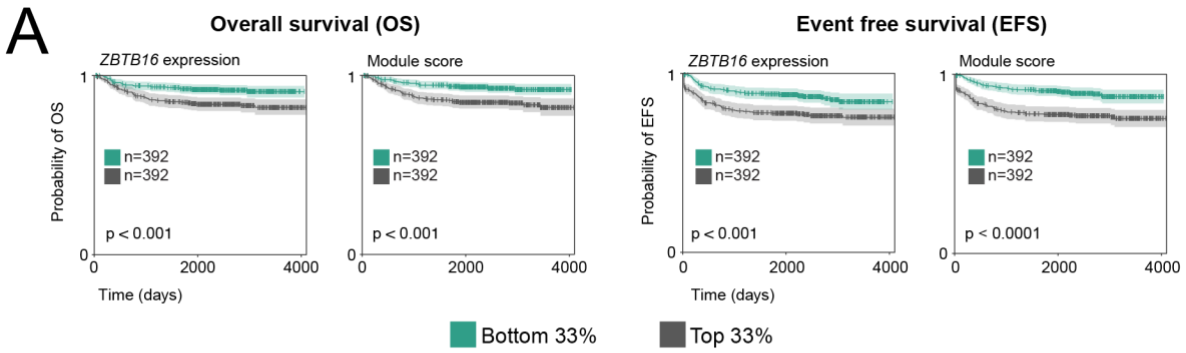
(**D**) Box plots showing *ZBTB16* expression and module score in bulk transcriptomes of an unpublished Princess Maxima Center cohort, plotted against day 28 MRD.

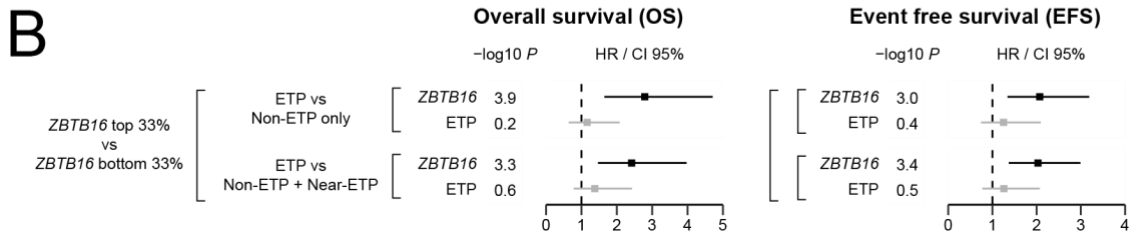
(**E**) Box plots showing *ZBTB16* expression and module score in bulk transcriptomes of the published COG AALL0434 study, plotted against day 28 MRD.

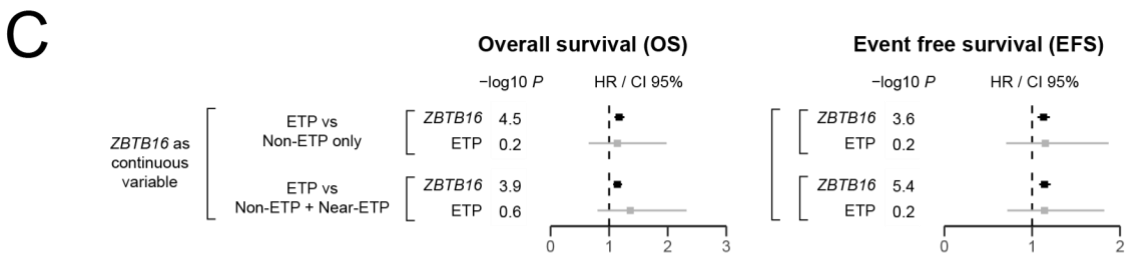
Box plots show the first and third quartiles (boxes), as well as median values (central lines). Whiskers extend to the most extreme values within 1.5 times the interquartile range above and below the boxes. All box plot *P* values were calculated by one-sided Wilcoxon rank-sum test.

MRD, minimal residual disease.

Figure 4.4


A
Overall survival (OS)
Event free survival (EFS)
ZBTB16 expression
Module score
Probability of OS
Probability of EFS
n=392
n=392
p < 0.001
p < 0.0001
Time (days)
Bottom 33%
Top 33%


B
Overall survival (OS)
Event free survival (EFS)
−log10 P
HR / CI 95%
ZBTB16 top 33%
vs
ZBTB16 bottom 33%
ETP vs
Non-ETP only
ETP vs
Non-ETP + Near-ETP
ZBTB16 3.9
ETP 0.2
ZBTB16 3.3
ETP 0.6
ZBTB16 3.0
ETP 0.4
ZBTB16 3.4
ETP 0.5


C
Overall survival (OS)
Event free survival (EFS)
−log10 P
HR / CI 95%
ZBTB16 as
continuous
variable
ETP vs
Non-ETP only
ETP vs
Non-ETP + Near-ETP
ZBTB16 4.5
ETP 0.2
ZBTB16 3.9
ETP 0.6
ZBTB16 3.6
ETP 0.2
ZBTB16 5.4
ETP 0.2

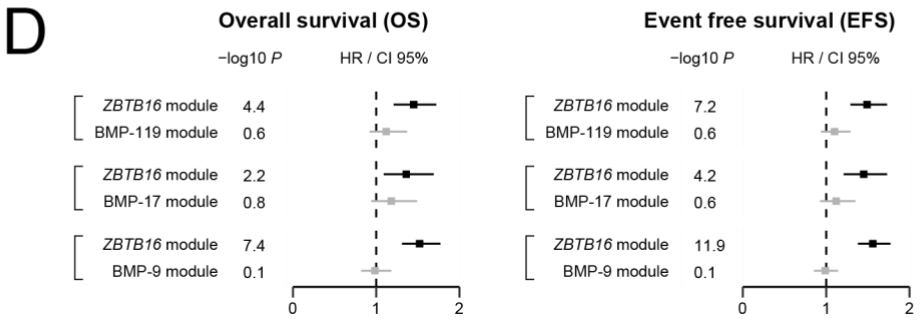

D
Overall survival (OS)
Event free survival (EFS)
−log10 P
HR / CI 95%
ZBTB16 module 4.4
BMP-119 module 0.6
ZBTB16 module 2.2
BMP-17 module 0.8
ZBTB16 module 7.4
BMP-9 module 0.1
ZBTB16 module 7.2
BMP-119 module 0.6
ZBTB16 module 4.2
BMP-17 module 0.6
ZBTB16 module 11.9
BMP-9 module 0.1

Figure 4.4: *ZBTB16* signal in bulk transcriptomes is predictive of survival in T-ALL.

(**A**) Overall survival (OS) and event free survival (EFS) in the COG ALL0434 cohort. Data was stratified into the top and bottom 33% of samples based on either *ZBTB16* expression or module score to test the association of these groups with survival using the Kaplan-Meier method. Error bands show 95% confidence intervals (CI 95%).

(**B**) Cox proportional hazard models were generated to compare *ZBTB16* expression against the immunophenotype-defined ETP status in the published COG AALL0434 study. *ZBTB16* was considered as a categorical variable, where the groups were defined as the top and bottom 33% as in **Figure 4.4A**, while ETP status was considered either excluding or including the "Near-ETP" group. Hazard ratios (HR) and 95% confidence intervals (CI 95%) were plotted, with statistically significant ones ($P < 0.05$) in black and non-significant ones in grey.

(**C**) Cox proportional hazard models were generated to compare *ZBTB16* expression as a continuous variable against the immunophenotype-defined ETP status in the published COG AALL0434 study. ETP status was considered either excluding or including the "Near-ETP" group. Hazard ratios (HR) and 95% confidence intervals (CI 95%) were plotted, with statistically significant ones ($P < 0.05$) in black and non-significant ones in grey.

(**D**) Cox proportional hazard models were generated to compare the *ZBTB16*-derived gene module against the various "BMP-like" gene signatures in bulk transcriptomes of the published COG AALL0434 study. Hazard ratios (HR) and 95% confidence intervals (CI 95%) were plotted, with statistically significant ones ($P < 0.05$) in black and non-significant ones in grey.

4.6 *ZBTB16* signal in an extended cohort of T-ALL single cell transcriptomes

To further validate the relationship between *ZBTB16* expression and refractory T-ALL, I analysed scRNA-seq of an extension cohort of 37 unselected T-ALL patients from GOSH (**Figure 4.5A**). I obtained a total of 333,487 high quality cells, comprising 218,599 leukaemia blasts and 114,888 normal cells (**Figure 4.5B**).

I interrogated *ZBTB16* expression across single leukaemia blast transcriptomes from the validation cohort (**Figure 4.6**). Out of 29 patients who were responsive to induction treatment, only 4 of them had *ZBTB16*+ clusters of blasts at diagnosis. Conversely, of the 6 patients with induction failure, 4 of them had *ZBTB16*+ clusters of blasts at diagnosis. In fact, for patient T014, *ZBTB16*+ blast clusters were found at all four timepoints: day 0, day 7, day 28 and at relapse. Therefore, single leukaemia blast transcriptomes from the extension cohort provide further support that *ZBTB16* expression is associated with refractory disease.

Figure 4.5

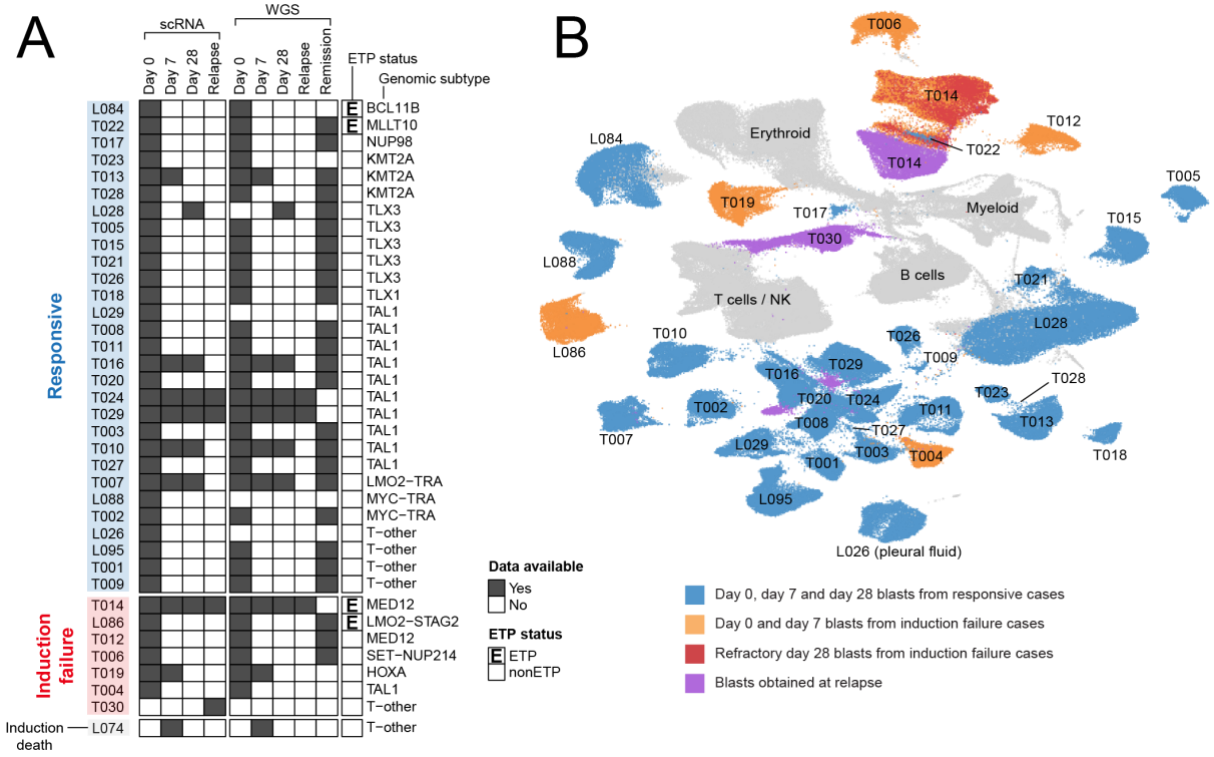


Figure 4.5: Overview of extension cohort scRNA-seq.

(**A**) Heatmap indicating sequencing data generated for each patient from the extension cohort and their clinical information (induction outcome, ETP status, genomic subtype).

(**B**) UMAP of 333,487 cells from the extension cohort, including 218,599 leukaemia blasts (coloured) and 114,888 normal cells (grey). Day 0, day 7 and day 28 blasts from patients who responded to induction treatment (blue); day 0 and day 7 blasts from patients with induction failure (orange); day 28 refractory blasts from patients with induction failure (red); blasts found at relapse (purple).

scRNA, single cell RNA sequencing; WGS, whole genome sequencing; ETP, early T cell precursor.

Figure 4.6

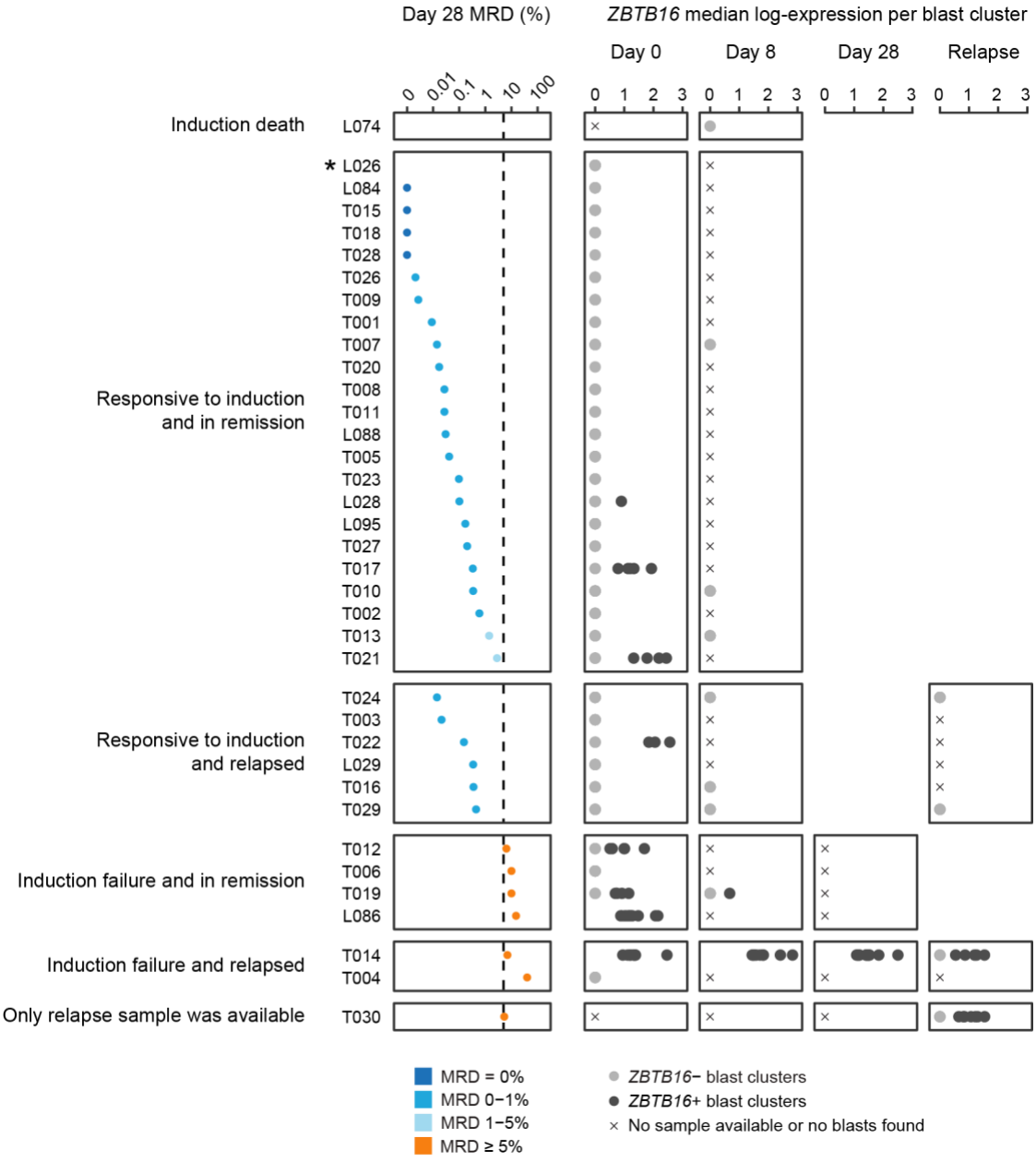

Day 28 MRD (%)
ZBTB16 median log-expression per blast cluster
Day 0
Day 8
Day 28
Relapse
0 0.01 0.1 1 10 100
0 1 2 3
Induction death
L074
Responsive to induction and in remission
* L026
L084
T015
T018
T028
T026
T009
T001
T007
T020
T008
T011
L088
T005
T023
L028
L095
T027
T017
T010
T002
T013
T021
Responsive to induction and relapsed
T024
T003
T022
L029
T016
T029
Induction failure and in remission
T012
T006
T019
L086
Induction failure and relapsed
T014
T004
Only relapse sample was available
T030
MRD = 0%
MRD 0–1%
MRD 1–5%
MRD ≥ 5%
ZBTB16− blast clusters
ZBTB16+ blast clusters
× No sample available or no blasts found

Figure 4.6: *ZBTB16* expression and clinical outcome in the extension scRNA-seq cohort

Longitudinal samples (including day 7, day 28 and relapse where available) were obtained from GOSH. The children are grouped along the y-axis according to whether they responded to induction and whether they eventually relapsed. The first panel from left shows day 28 MRD (%), coloured by MRD group. The following panels show median *ZBTB16* expression of each cluster of blasts within each sample across different timepoints; filled circles are coloured light grey if the median *ZBTB16* expression in that cluster is zero. L026 is a child with T-lymphoblastic lymphoma who was found by CT scan to have responded completely at day 28.

MRD, minimal residual disease.

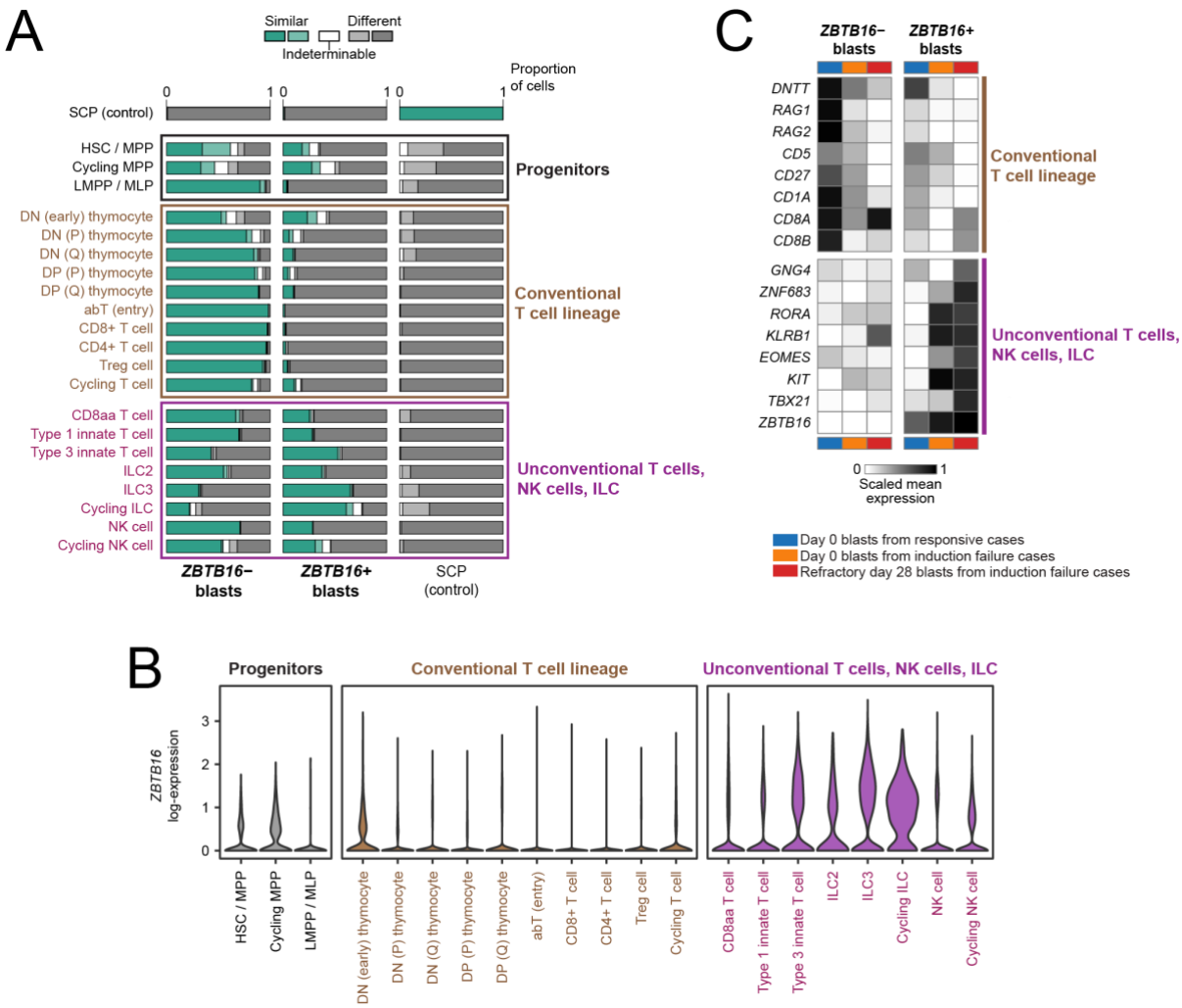
4.7 Leukaemia cell state of *ZBTB16*+ T-ALL blasts

Given that *ZBTB16*+ blasts are associated with refractory T-ALL and worse survival, they could also represent a distinct leukaemia cell state. To investigate this, I generated a logistic regression model trained on *ZBTB16*+ blasts and *ZBTB16*− blasts. Schwann cell precursors (SCP) were included as a negative control (Kildisiute et al., 2021). I then applied this logistic regression model to query T cell subsets and developmental stages from a single cell transcriptome atlas of fetal immune cell development (Suo et al., 2022). I found that various cell states of unconventional T cells, NK cells and innate lymphoid cells (ILC) resembled *ZBTB16*+ blasts, while conventional T cells and their developmental stages resembled *ZBTB16*− blasts (**Figure 4.7A**). This is consistent with the higher expression of *ZBTB16* in unconventional T cells, NK cells and ILC, compared to cells of the conventional T cell lineage (**Figure 4.7B**).

Plotting the expression of well-characterised marker genes of T cells, NK cells and ILCs highlights the disparity between *ZBTB16*+ and *ZBTB16*− blasts (**Figure 4.7C**). Whereas *ZBTB16*− blasts expressed higher levels of *DNTT*, *RAG1* and *RAG2*, which are enzymes involved in somatic recombination of T cell receptors, *ZBTB16*+ blasts expressed markers associated with unconventional T cells, NK cells and ILCs, such as *RORA* and *KLRB1*. Overall, these results suggest that *ZBTB16*+ blasts have a non-canonical, innate-like lymphocyte cell state, possibly driven by the ZBTB16 transcription factor.

The non-canonical leukaemia cell state of *ZBTB16*+ T-ALL blasts may account for their refractory phenotype. These blasts may express lower levels of the glucocorticoid receptor (*NR3C1*) and be less responsive to corticosteroids given during induction. Alternatively, they may be less proliferative than canonical T-ALL blasts and thus less vulnerable to cytotoxic chemotherapy which target rapidly-dividing cells. To investigate these two hypotheses, I plotted the expression of *NR3C1* and *MKI67* (a marker of cell proliferation), as well as cell cycle phase, against *ZBTB16* expression across my scRNA-seq cohort (**Figure 4.8**). Although *ZBTB16* expression was not associated with lower *NR3C1* levels, it was associated with lower *MKI67* levels and a smaller proportion of blasts in the G2/M phase of the cell cycle. Therefore, *ZBTB16*+ blasts might represent a quiescent leukaemia cell state.

Figure 4.7


A
Similar
Different
Indeterminable
Proportion of cells
SCP (control)
HSC / MPP
Cycling MPP
LMPP / MLP
Progenitors
DN (early) thymocyte
DN (P) thymocyte
DN (Q) thymocyte
DP (P) thymocyte
DP (Q) thymocyte
abT (entry)
CD8+ T cell
CD4+ T cell
Treg cell
Cycling T cell
Conventional T cell lineage
CD8aa T cell
Type 1 innate T cell
Type 3 innate T cell
ILC2
ILC3
Cycling ILC
NK cell
Cycling NK cell
Unconventional T cells, NK cells, ILC
ZBTB16− blasts
ZBTB16+ blasts
SCP (control)
C
ZBTB16− blasts
ZBTB16+ blasts
DNTT
RAG1
RAG2
CD5
CD27
CD1A
CD8A
CD8B
Conventional T cell lineage
GNG4
ZNF683
RORA
KLRB1
EOMES
KIT
TBX21
ZBTB16
Unconventional T cells, NK cells, ILC
Scaled mean expression
Day 0 blasts from responsive cases
Day 0 blasts from induction failure cases
Refractory day 28 blasts from induction failure cases

B
Progenitors
Conventional T cell lineage
Unconventional T cells, NK cells, ILC
ZBTB16 log-expression
HSC / MPP
Cycling MPP
LMPP / MLP
DN (early) thymocyte
DN (P) thymocyte
DN (Q) thymocyte
DP (P) thymocyte
DP (Q) thymocyte
abT (entry)
CD8+ T cell
CD4+ T cell
Treg cell
Cycling T cell
CD8aa T cell
Type 1 innate T cell
Type 3 innate T cell
ILC2
ILC3
Cycling ILC
NK cell
Cycling NK cell

Figure 4.7: Comparing *ZBTB16*+ T-ALL blasts to T cell developmental stages

(**A**) A logistic regression model was trained using *ZBTB16*+ blasts, *ZBTB16*− blasts, and Schwann cell precursors (SCP) which serve as a negative control. This model was used to calculate the similarity of query cell types (y-axis: conventional T cells, unconventional T cells, NK cells, ILC) to reference cell types (x-axis).

(**B**) *ZBTB16* expression across stages of T cell development and various T cell subtypes (progenitors, conventional T cells, unconventional T cells, NK cells, ILC), obtained from a published single cell atlas of the normal developing fetal immune system.

(**C**) Heatmap showing expression of well-characterised cell type marker genes across *ZBTB16*+ blasts and *ZBTB16*− blasts, taken from day 0 samples of responsive patients (blue), from day 0 samples of induction failure patients (orange), and from day 28 samples of induction failure patients (red). Expression values are log-normalised gene expression averaged across blasts within each group. Genes are known markers for conventional T cell lineages (top) as well as unconventional T cells, NK cells and ILC (bottom).

HSC, haematopoietic stem cell; MPP, multipotent progenitor; LMPP, lymphoid-primed multipotent progenitor; MLP, multi-lymphoid progenitor; DN, double-negative; DP, double-positive; ILC, innate lymphoid cell; NK cell, natural killer cell.

Figure 4.8

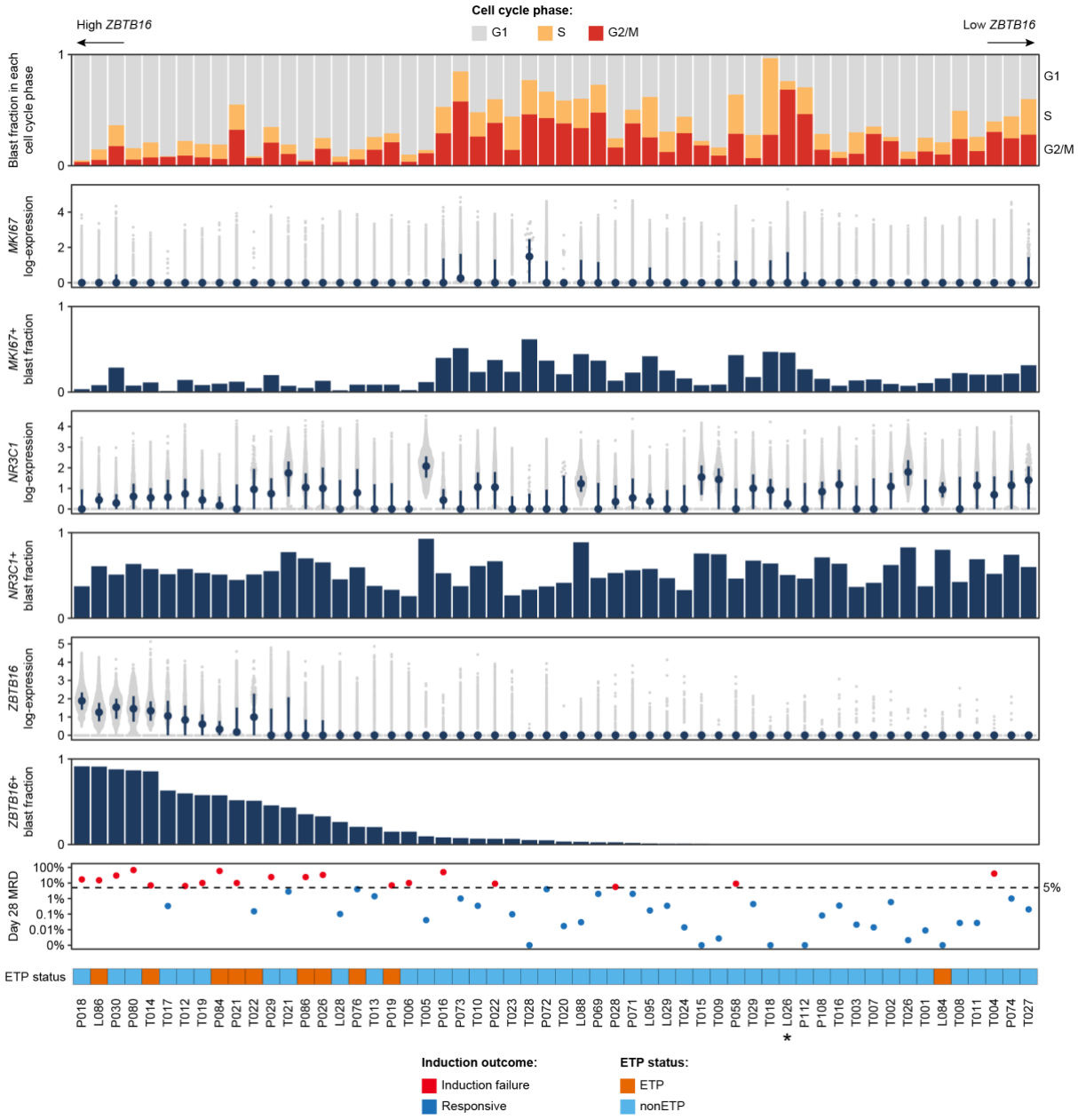

Cell cycle phase:
G1
S
G2/M
High ZBTB16
Low ZBTB16
Blast fraction in each cell cycle phase
MKI67 log-expression
MKI67+ blast fraction
NR3C1 log-expression
NR3C1+ blast fraction
ZBTB16 log-expression
ZBTB16+ blast fraction
Day 28 MRD
100%
10%
1%
0.1%
0.01%
0%
5%
ETP status
P018 L086 P030 P080 T014 T017 T012 T019 P084 P021 T022 P029 T021 P086 P026 L028 P076 T013 P019 T006 T005 P016 P073 T010 P022 T023 T028 P072 T020 L088 P069 P028 P071 L095 L029 T024 T015 T009 P058 P029 T018 L026 P112 P108 T016 T003 T007 T002 T026 T001 L084 T008 T011 T004 P074 T027
*
Induction outcome:
Induction failure
Responsive
ETP status:
ETP
nonETP

Figure 4.8: Cell cycle phase and *NR3C1* expression in *ZBTB16*+ T-ALL blasts

Expression of *ZBTB16*, *NR3C1* (glucocorticoid receptor) and *MKI67* (a marker of cell proliferation) in single leukaemia blasts, and the fraction of blasts which express these genes, are plotted across 58 diagnostic samples from my scRNA-seq cohort. For each sample, the median expression and interquartile range are shown by the large central dot and vertical line, respectively. Samples are arranged in order from highest to lowest *ZBTB16*+ blast fraction. Fraction of blasts in each cell cycle phase (G1, S, G2/M) is displayed at the top. Day 28 MRD and ETP status are indicated. L026 is a child with T-lymphoblastic lymphoma who was found by CT scan to have responded completely at day 28.

MRD, minimal residual disease; ETP, early T cell precursor.

4.8 Drug sensitivity of *ZBTB16*+ T-ALL blasts

Refractory disease in *ZBTB16*+ T-ALL may be due to intrinsic resistance to specific drugs, including those used during induction: dexamethasone, PEGylated L-asparaginase, vincristine and daunorubicin. To assess if *ZBTB16*+ T-ALL blasts may be resistant to specific drugs, I turned to a recently published dataset of drug sensitivity in primary leukaemia blasts with concomitant bulk transcriptome sequencing (Lee et al., 2023). Drug sensitivity was quantified using LC50, which is the lethal concentration of drug needed to kill 50% of leukaemia blasts. Raw LC50 values were log-transformed and normalised by the authors to lie between 0 (most sensitive) and 1 (most resistant). The study comprised 805 children with ALL, including 16 cases of ETP and 113 cases of non-ETP T-ALL. Not all drugs were tested on every leukaemia case, resulting in variable sample sizes among the drugs.

Of the 17 drugs analysed, I did not find any statistically significant difference in drug sensitivity (false discovery rate = 0.05) when comparing between high and low *ZBTB16* expression (**Figure 4.9A**), or between high and low *ZBTB16*-derived module score (**Figure 4.9B**). However, I did observe that T-ALL with high module score had higher normalised LC50 values for dexamethasone (more resistant), thought this was not statistically significant. A substantial number of cases had normalised LC50 for dexamethasone at the upper limit of 1. It is possible that if higher concentrations of dexamethasone were tested, it might reveal a true difference in sensitivity to dexamethasone.

Next, I investigated if *ZBTB16* expression (**Figure 4.9C**) or its associated module score (**Figure 4.9D**) is related to drug sensitivity in a continuous rather than stepwise fashion. I found that the module score was positively correlated with normalised LC50 for prednisolone, nelarabine, mercaptopurine and panobinostat, and these correlations were found to be statistically significant (false discovery rate = 0.05). Additionally, there also appeared to be a positive relationship between module score and normalised LC50 for dexamethasone, as well as a negative relationship between module score and normalised LC50 for venetoclax, although these were not statistically significant. Overall, these results suggest there may be differences in drug sensitivities in *ZBTB16*+ blasts, but the evidence from this study is weak.

Figure 4.9A-B

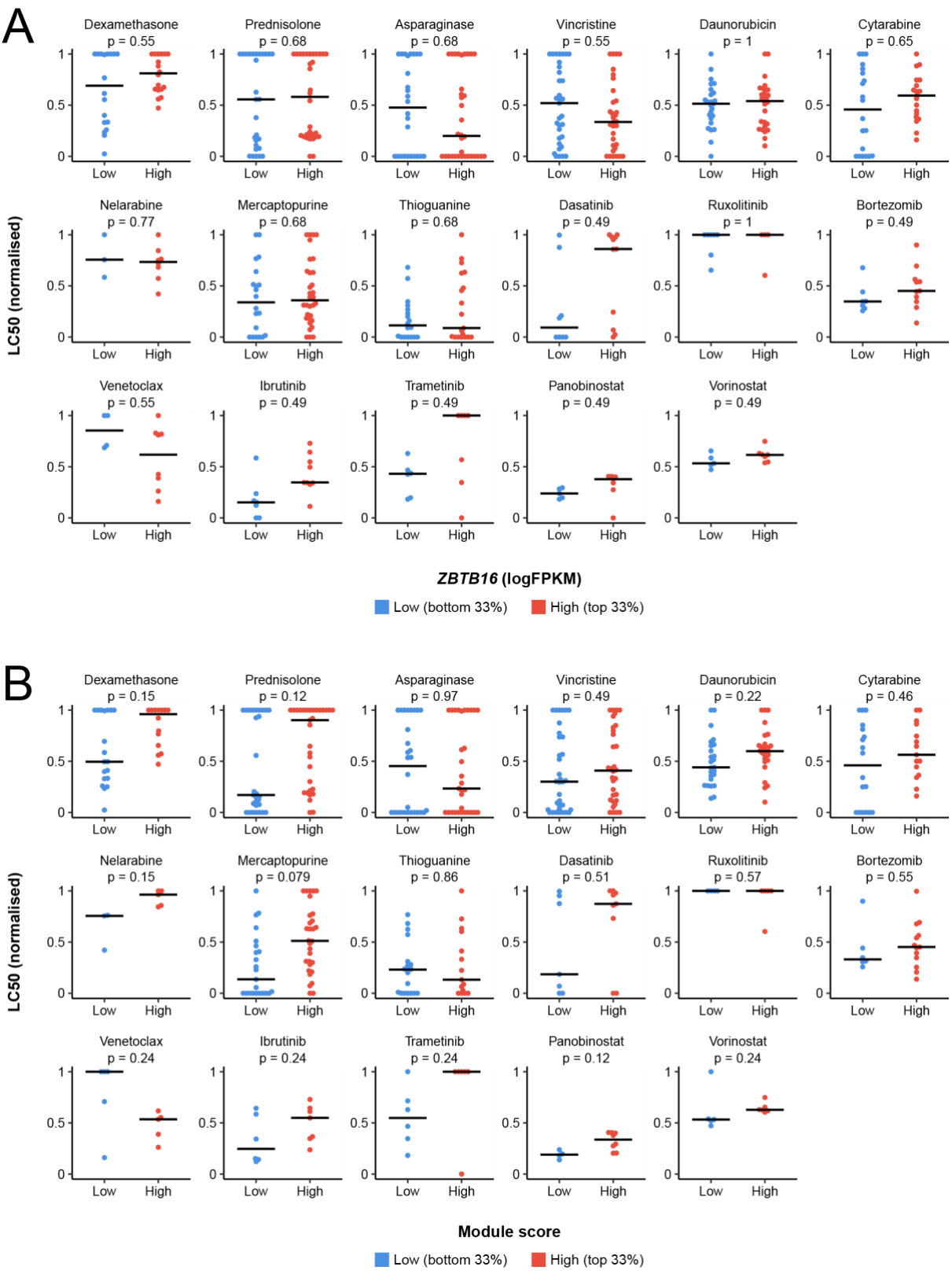

A
Dexamethasone
p = 0.55
Prednisolone
p = 0.68
Asparaginase
p = 0.68
Vincristine
p = 0.55
Daunorubicin
p = 1
Cytarabine
p = 0.65
Nelarabine
p = 0.77
Mercaptopurine
p = 0.68
Thioguanine
p = 0.68
Dasatinib
p = 0.49
Ruxolitinib
p = 1
Bortezomib
p = 0.49
Venetoclax
p = 0.55
Ibrutinib
p = 0.49
Trametinib
p = 0.49
Panobinostat
p = 0.49
Vorinostat
p = 0.49
LC50 (normalised)
Low
High
ZBTB16 (logFPKM)
Low (bottom 33%)
High (top 33%)
B
Dexamethasone
p = 0.15
Prednisolone
p = 0.12
Asparaginase
p = 0.97
Vincristine
p = 0.49
Daunorubicin
p = 0.22
Cytarabine
p = 0.46
Nelarabine
p = 0.15
Mercaptopurine
p = 0.079
Thioguanine
p = 0.86
Dasatinib
p = 0.51
Ruxolitinib
p = 0.57
Bortezomib
p = 0.55
Venetoclax
p = 0.24
Ibrutinib
p = 0.24
Trametinib
p = 0.24
Panobinostat
p = 0.12
Vorinostat
p = 0.24
LC50 (normalised)
Module score
Low (bottom 33%)
High (top 33%)

Figure 4.9C-D

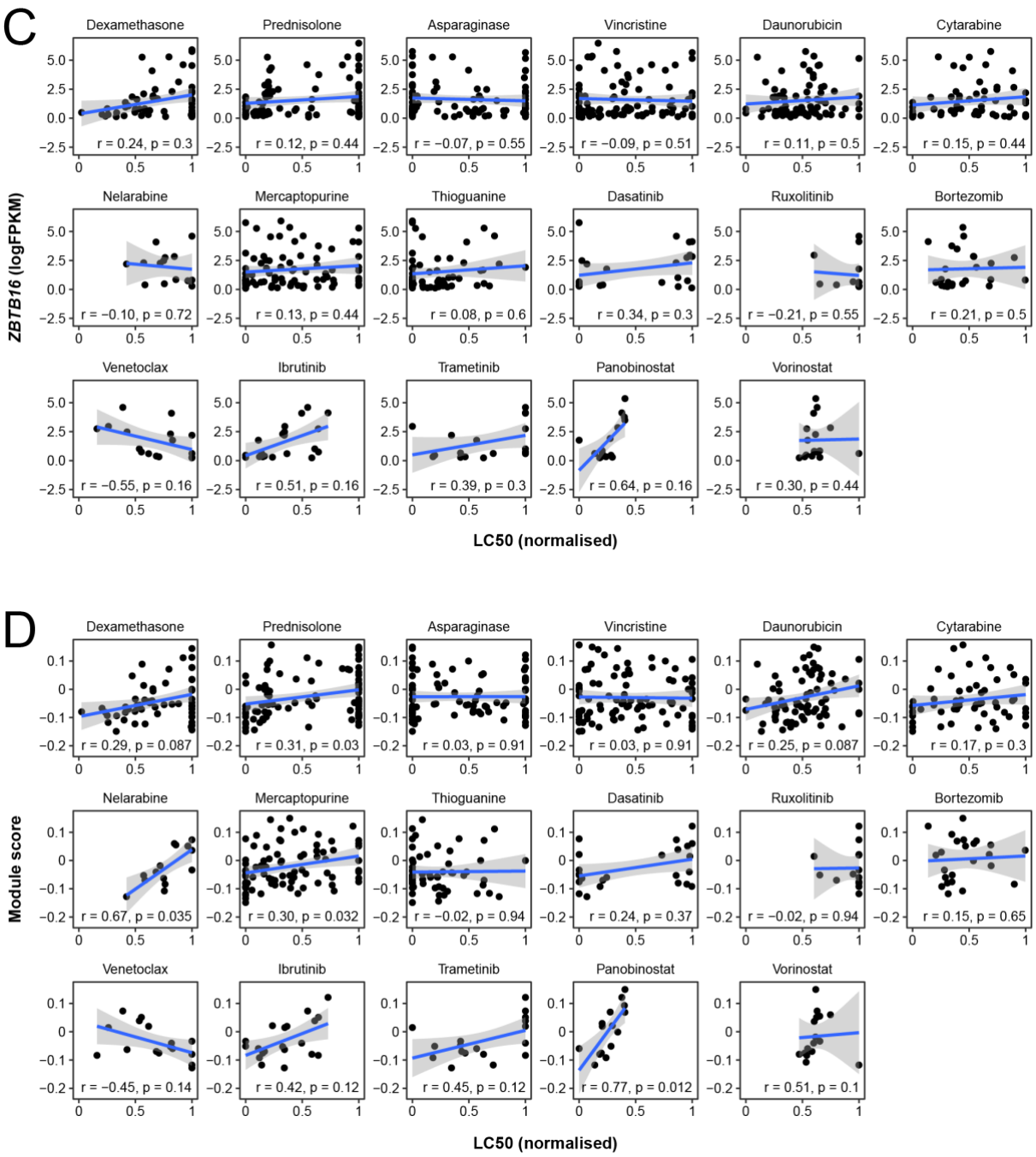

C
Dexamethasone
r = 0.24, p = 0.3
Prednisolone
r = 0.12, p = 0.44
Asparaginase
r = −0.07, p = 0.55
Vincristine
r = −0.09, p = 0.51
Daunorubicin
r = 0.11, p = 0.5
Cytarabine
r = 0.15, p = 0.44
Nelarabine
r = −0.10, p = 0.72
Mercaptopurine
r = 0.13, p = 0.44
Thioguanine
r = 0.08, p = 0.6
Dasatinib
r = 0.34, p = 0.3
Ruxolitinib
r = −0.21, p = 0.55
Bortezomib
r = 0.21, p = 0.5
Venetoclax
r = −0.55, p = 0.16
Ibrutinib
r = 0.51, p = 0.16
Trametinib
r = 0.39, p = 0.3
Panobinostat
r = 0.64, p = 0.16
Vorinostat
r = 0.30, p = 0.44
ZBTB16 (logFPKM)
LC50 (normalised)
D
Dexamethasone
r = 0.29, p = 0.087
Prednisolone
r = 0.31, p = 0.03
Asparaginase
r = 0.03, p = 0.91
Vincristine
r = 0.03, p = 0.91
Daunorubicin
r = 0.25, p = 0.087
Cytarabine
r = 0.17, p = 0.3
Nelarabine
r = 0.67, p = 0.035
Mercaptopurine
r = 0.30, p = 0.032
Thioguanine
r = −0.02, p = 0.94
Dasatinib
r = 0.24, p = 0.37
Ruxolitinib
r = −0.02, p = 0.94
Bortezomib
r = 0.15, p = 0.65
Venetoclax
r = −0.45, p = 0.14
Ibrutinib
r = 0.42, p = 0.12
Trametinib
r = 0.45, p = 0.12
Panobinostat
r = 0.77, p = 0.012
Vorinostat
r = 0.51, p = 0.1
Module score
LC50 (normalised)

Figure 4.9: Drug sensitivity of *ZBTB16*+ T-ALL blasts.

(**A**) Swarm plots showing normalised LC50 in T-ALL with low (bottom 33% of samples) and high (top 33% of samples) *ZBTB16* expression for each of the tested drugs. Horizontal bars indicate median normalised LC50 for each group.

(**B**) Swarm plots showing normalised LC50 in T-ALL with low (bottom 33% of samples) and high (top 33% of samples) module score for each of the tested drugs. Horizontal bars indicate median normalised LC50 for each group.

(**C**) Correlation plots between *ZBTB16* expression and normalised LC50 for each of the tested drugs. Blue line represents linear regression line and grey shaded area indicates 95% confidence interval.

(**D**) Correlation plots between module score and normalised LC50 for each of the tested drugs. Blue line represents linear regression line and grey shaded area indicates 95% confidence interval.

All swarm plot *P* values were calculated with two-sided Wilcoxon rank-sum test and adjusted by Benjamini-Hochberg correction for multiple hypothesis testing. All correlation plots display Spearman coefficients and *P* values adjusted by Benjamini-Hochberg correction for multiple hypothesis testing.

LC50, lethal concentration to kill 50% of leukaemia blasts.

4.9 Derivation of potential cell surface targets on T-ALL blasts

Our extensive single cell transcriptome cohort of 55 individuals provides a rich dataset to search for new therapeutic targets, especially cell surface proteins which may be targetted by monoclonal antibodies or chimeric antigen receptor (CAR) T cells. Together with Dr Holly Whitfield, I derived a list of genes which are expressed by T-ALL blasts but not by normal T cells, and filtered to keep those which encode for cell surface proteins (**Figure 4.10A**). Moreover, as *ZBTB16*+ blasts were shown earlier to be the source of refractory disease, I also derived another list of genes which are specifically upregulated in refractory *ZBTB16*+ blasts but not normal T cells, and filtered to keep those encoding for cell surface proteins.

I examined the expression of these target genes in cells from our T-ALL scRNA-seq data, as well as in cells from a published single cell atlas of the normal immune system (Domínguez Conde et al., 2022) (**Figure 4.10B**). I found the expression profiles of both the targets against T-ALL and the targets against refractory *ZBTB16*+ T-ALL to be comparable or even better than cell surface targets that are already under investigation (Chiesa et al., 2023; Maciocia et al., 2022; Oh et al., 2024; Pan et al., 2025; Sánchez-Martínez et al., 2019). For instance, although CD7 has been trialled as a target for CAR T cell therapy due to its high expression in T-ALL blasts, it is also ubiquitously expressed in normal T cells, as well as NK cells and ILCs. While anti-CD7 CAR T cells may be modified to prevent their surface CD7 from being recognised and hence avoid fratricide (Chiesa et al., 2023; Oh et al., 2024), depletion of CD7+ T cells remains a major side effect. Instead, I propose targets such as *FLT3*, *FZD6* and *CD109* which are expressed in T-ALL but not in normal T cells. Additionally, targets like *IL9R* appear more highly expressed in *ZBTB16*+ blasts, which are associated with refractory T-ALL.

Figure 4.10

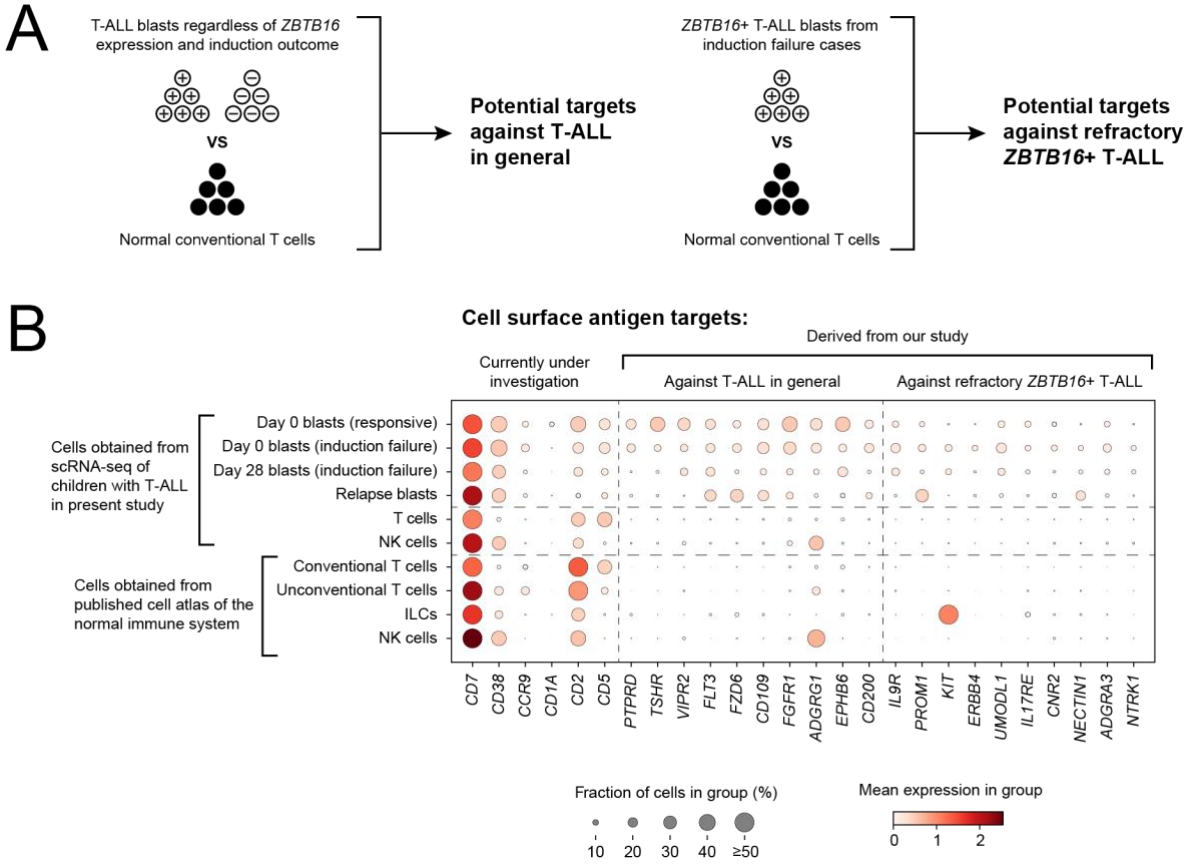

A
T-ALL blasts regardless of ZBTB16 expression and induction outcome
vs
Normal conventional T cells
Potential targets against T-ALL in general
ZBTB16+ T-ALL blasts from induction failure cases
vs
Normal conventional T cells
Potential targets against refractory ZBTB16+ T-ALL
B
Cell surface antigen targets:
Derived from our study
Currently under investigation
Against T-ALL in general
Against refractory ZBTB16+ T-ALL
Cells obtained from scRNA-seq of children with T-ALL in present study
Day 0 blasts (responsive)
Day 0 blasts (induction failure)
Day 28 blasts (induction failure)
Relapse blasts
T cells
NK cells
Cells obtained from published cell atlas of the normal immune system
Conventional T cells
Unconventional T cells
ILCs
NK cells
CD7 CD38 CCR9 CD1A CD2 CD5 PTPRD TSHR VIPR2 FLT3 FZD6 CD109 FGFR1 ADGRG1 EPHB6 CD200 IL9R PROM1 KIT ERBB4 UMODL1 IL17RE CNR2 NECTIN1 ADGRA3 NTRK1
Fraction of cells in group (%)
10 20 30 40 ≥50
Mean expression in group
0 1 2

Figure 4.10: Potential cell surface targets on T-ALL blasts.

(**A**) Derivation of two lists of potential antigen targets for (1) T-ALL blasts in general and (2) refractory *ZBTB16*+ blasts specifically. Differential expression analysis between blasts and normal cells across our T-ALL scRNA-seq cohort provided blast-specific genes that were further filtered for cell surface targets.

(**B**) Dot plot showing expression of potential cell surface targets on cells from our T-ALL scRNA-seq cohort (day 0 blasts, day 28 blasts, relapse blasts, normal T cells, normal NK cells), as well as cells from a published single cell atlas of the normal immune system which include: conventional T cells (T CD4/CD8, T naive/CM CD8, T naive/CM CD4, T naive/CM CD4 activated, Tfh, Tregs, T effector/EM CD4, Trm Th1/Th17, Trm gut CD8, Tem/emra CD8, Trm/em CD8), unconventional T cells (MAIT, Trm Tgd, Tgd CRTAM+), ILCs, and NK cells. Cell surface antigen targets include those derived from our study against T-ALL blasts in general and against refractory *ZBTB16*+ T-ALL blasts, as well as those which are currently under investigation.

4.10 Discussion

In this chapter, I examined single cell transcriptomes of leukaemia blasts from a cohort of 55 children with T-ALL. For a number of children, multiple timepoints were analysed, including at the end of induction treatment (day 28), which represents a critical point where treatment response is evaluated. Through my investigations, I discovered *ZBTB16* as an unexpected marker of refractory T-ALL blasts. I showed that evidence of *ZBTB16*+ blasts at diagnosis portends induction failure in independent bulk transcriptome sequencing cohorts and is predictive of worse survival in the largest T-ALL cohort of over 1,300 patients. *ZBTB16* expression delineates a non-canonical lymphoblast cell state that resembles that of unconventional T cells, NK cells and ILC, as opposed to conventional T cells.

The results of my investigation raise the exciting possibility of solving the clinical conundrum of risk stratification in T-ALL. Despite previous extensive work by other groups, there is currently no marker in the clinic for predicting at diagnosis which patient with T-ALL will have a worse prognosis. Although various markers have been proposed previously, including specific genetic mutations or immunophenotype categories, none have proven to be sufficiently robust to be incorporated into current clinical practice (Conter et al., 2016; Coustan-Smith et al., 2009; Jenkinson et al., 2013; Petit et al., 2018; Seki et al., 2017; Taj et al., 2022).

The fact that I have found a single marker gene, as opposed to a set of genes, allows for easy implementation in clinical flow cytometry, which is routinely used to diagnose, characterise and monitor treatment in T-ALL (Pagliaro et al., 2024). Collaborators at GOSH demonstrated that ZBTB16, despite being an intracellular protein, can be measured by flow cytometry, and are expanding their investigation into newly diagnosed T-ALL cases. Prospective studies measuring ZBTB16 by flow cytometry and monitoring patients in the long term will be necessary to establish the ability of ZBTB16 to predict induction failure and survival.

Previous groups have attempted to link refractory T-ALL to a particular lymphoblast cell state. The ETP immunophenotype has been a notable example in the past 15 years (Coustan-Smith et al., 2009) and was incorporated into the 2017 World Health Organization (WHO) classification as a provisional subtype (Arber et al., 2016). However, subsequent data from the COG AALL0434 cohort showed no significant difference in survival outcome in ETP T-ALL (Wood et al., 2023). This agrees with my analyses, where I showed that only *ZBTB16* expression and not ETP status, when modelled together, was associated with survival.

More recently, the BMP-like leukaemia cell state was proposed a predictor of refractory disease in T-ALL and a potential replacement for the ETP definition (Xu et al., 2024). Although the authors demonstrated worse survival in patients whose leukaemia transcriptomes scored highly for BMP-like gene signatures, when their gene signatures were modelled in a direct comparison with our *ZBTB16* gene module, I found that only the *ZBTB16* gene module and not the BMP-like gene signatures, was significantly associated with poorer survival. These results suggest that the *ZBTB16*+ cell state supersedes the BMP-like cell state.

ZBTB16 (zinc finger and BTB domain containing 16), also known as *PLZF* (promyelocytic leukaemia zinc finger) was first discovered in a patient with acute promyelocytic leukaemia, where it formed a fusion gene with *RARA* (retinoic acid receptor alpha) as a result of a t(11;17) translocation (Chen et al., 1993). The ZBTB16 protein has previously been found to be expressed in a subset of T-ALL (Jeon et al., 2012) and in a subset of peripheral T cell lymphoma (McGregor et al., 2014). However, prior to my doctoral study, *ZBTB16* expression has not be associated with poorer clinical outcome in T-ALL.

The discovery of *ZBTB16*+ blasts in refractory T-ALL raises important biological questions: (1) what causes *ZBTB16* to be expressed in a subset of T-ALL and (2) why ZBTB16 is associated with treatment resistance in T-ALL. Although I may be able to speculate the pathways involved by examining corelated genes in single cell leukaemia transcriptomes, other sequencing modalities and functional investigations are required to establish the biological mechanisms involving ZBTB16 in T-ALL.

The driving cause of *ZBTB16* expression could be a somatic mutation, an epigenetic change, or perhaps even chance itself. *ZBTB16-ABL1* gene fusions have been detected in T-ALL and these leukaemias are sensitive to tyrosine kinase inhibitors (Chen et al., 2018), suggesting that the leukaemogenic effects are due to the ABL1 kinase rather than the ZBTB16 transcription factor. Apart from being a gene fusion partner, *ZBTB16* is not otherwise known to be a cancer gene in T-ALL. Over-expression of *ZBTB16* could be caused by mutations in an intergenic region, such as in the case for *TAL1* and *BCL11B* (Mansour et al., 2014; Montefiori et al., 2021). Such driver events may be difficult to detect and hence explain why *ZBTB16* is not currently a known cancer gene in T-ALL. If allele specific expression of *ZBTB16* is detected, it may suggest a cis-acting mechanism (e.g. a mutation which generates a novel enhancer) instead of a trans-acting mechanism (e.g. an upstream transcription factor which binds to the *ZBTB16* enhancer). Alternatively, the cause of *ZBTB16* expression may be epigenetic, such as methylation, histone modification or chromatin conformation (Kloetgen et al., 2020; Touzart et al., 2021). Additionally, noncoding RNA may also be involved in the

regulation of *ZBTB16*. For instance, the let-7 family of microRNAs target *Zbtb16* mRNA to direct the terminal differentiation fate of NKT cells in mice (Pobezinsky et al., 2015).

ZBTB16 is a lineage-defining transcription factor that is required for the development of unconventional T cells, NK cells and ILC (Constantinides et al., 2014; Klose et al., 2014; Koay et al., 2016; Kovalovsky et al., 2008; Savage et al., 2008). *ZBTB16*+ blasts could thus represent of a distinct T cell or lymphocyte lineage that is unlike canonical stages of conventional T cell development. Experimental over-expression of ZBTB16 is sufficient to induce an unconventional T cell-like phenotype even in mature conventional T cells (Kovalovsky et al., 2010). It is possible then that *ZBTB16* expression might drive T-ALL blasts to adopt a non-canonical phenotype.

Analysis of my scRNA-seq data suggests that *ZBTB16* expression is associated with reduced cell proliferation in T-ALL. Given that cytotoxic chemotherapy targets rapidly-dividing cancer cells, this could be a possible mechanism for the intrinsic resistance of *ZBTB16*+ T-ALL blasts. The ZBTB16 protein has previously been shown to restrict the proliferation of myeloid progenitors (Doulatov et al., 2009) and it has also been found to inhibit tumour growth in breast and lung cancers (He et al., 2020; Wang et al., 2024).

Data from drug sensitivity assays of primary T-ALL blasts (Lee et al., 2023) suggest that evidence of *ZBTB16*+ blasts may possibly be associated with resistance to dexamethasone. An earlier study in ALL cell lines (including T-ALL) has linked *ZBTB16* expression with glucocorticoid resistance (Wasim et al., 2010). However, *ZBTB16*+ blasts do not appear to have reduced expression of the glucocorticoid receptor. Thus, if *ZBTB16*+ blasts are resistant to glucocorticoid therapy, it is likely occurring downstream of the glucocorticoid receptor.

For now, the precise mechanism by which *ZBTB16* expression leads to refractory T-ALL remains elusive. Chromatin immunoprecipitation and sequencing (ChIP-seq) for the ZBTB16 transcription factor may delineate the target genes it regulates and provide insights into the mechanism of refractory disease. This may also help identify additional markers to define *ZBTB16*+ blasts, thereby improving clinical risk stratification and providing additional pharmacological targets to treat the leukaemia. Alternatively, there could be other factors which cooperate with *ZBTB16* to produce the refractory phenotype. This may explain the presence of a small number of individuals in my study cohort with *ZBTB16*+ blasts that responded to induction treatment, although this could equally be ascribed to random chance. Finally, it is possible that *ZBTB16* is simply a non-functional marker of refractory T-ALL and the cause of treatment resistance may lie upstream of *ZBTB16* itself.

Despite decades of T-ALL research, the association between *ZBTB16* and refractory disease has eluded discovery, up till now. Several factors have contributed to this serendipitous finding. Firstly, there has been greater understanding and appreciation of unconventional T cells and ILC in recent years, whereas earlier T-ALL research were based on models of conventional T cell development. Next, transcriptome sequencing at the single cell resolution offers greater sensitivity and specificity to detect the *ZBTB16* signal, such as in the diagnostic sample of P058, where only 1% of blasts were *ZBTB16*+. Furthermore, we included day 28 samples of patients with induction failure, which by definition contain the actual refractory blasts that were not killed by treatment. Finally, we did not confine our study to specific T-ALL subtypes (e.g. ETP T-ALL), nor did we enrich samples for blasts which may otherwise bias the data. The sequencing strategy in our study may provide a blueprint for future single cell transcriptome sequencing projects in leukaemia.

Chapter 5: Origin of refractory blasts in T-ALL

5.1 Introduction

T-ALL has higher rates of refractory disease compared to B-ALL (Goldberg et al., 2003). Induction failure, defined as the presence of ≥5% blasts following the initial 28 days of chemotherapy treatment (O'Connor et al., 2017), is the first sign of refractory T-ALL and a powerful predictor of worse clinical prognosis. Only half of T-ALL patients with induction failure survive beyond 5 years (Raetz et al., 2023). In the previous chapter, I showed that refractory T-ALL is driven by *ZBTB16*+ blasts. However, the origin of refractory *ZBTB16*+ blasts is not known.

A recent landmark study of the genomic and transcriptomic profiles of ~1,300 individuals with T-ALL, including refractory T-ALL, identified high-risk genomic subtypes of T-ALL (Pölönen et al., 2024). It is possible that these high-risk genomic subtypes strongly express *ZBTB16*. Alternatively, T-ALL can be classified by its differentiation state, which may be defined by its progress through the stages of TCR rearrangement (Asnafi et al., 2003). Thus, it is also possible that *ZBTB16* expression in T-ALL may be related to TCR rearrangement status.

Noncoding mutations may drive the aberrant expression of cancer genes. For instance, sequence insertion mutations in an intergenic region 7.5 kb upstream of the *TAL1* gene has been shown to create a novel super-enhancer that drives *TAL1* expression in around 5% of T-ALL (Mansour et al., 2014). Additional noncoding mutations relating to T-ALL genes have also been found in *LMO1* and *LMO2* (Hu et al., 2017), *MYC* (Herranz et al., 2014) and *BCL11B* (Montefiori et al., 2021). It is not known if a noncoding mutation drives *ZBTB16* expression in T-ALL or any other cancer.

Genomic sequencing of leukaemia across longitudinal timepoints has identified leukaemia evolution pathways and provide insights into treatment resistance. Notably, relapsed ALL was found to have acquired mutations in *NT5C2*, which were not found at diagnosis, thereby conferring resistance to thiopurine drugs used in the maintenance treatment phase (Tzoneva et al., 2013). In other cases, relapsed leukaemia represents an independent leukaemia lineage derived from a pre-leukaemic clone and lacks known chemotherapy resistance mutations (O'Connor et al., 2024). Phylogenetic analysis of refractory *ZBTB16*+ T-ALL may uncover the origin of refractory blasts, which may be different from relapse blasts.

Here, I examined the origin of refractory *ZBTB16*+ blasts. First, I analysed data from genomic sequencing of my T-ALL cohort by annotating driver mutations and genomic subtype. From this, I then explored the possibility that T-ALL genomic subtype may be associated with the refractory *ZBTB16*+ phenotype. I also considered whether TCR rearrangement state is related to *ZBTB16* expression in T-ALL blasts. Third, I searched for evidence that *ZBTB16* expression may be driven by a noncoding mutation. Finally, I reconstructed the leukaemia phylogeny of two individuals, P058 and P030, to demonstrate how refractory blasts at the end of induction treatment is genetically related to blasts at diagnosis.

5.2 Overview and quality control of genomic sequencing

Key to the analyses presented in this chapter are somatic mutations called from whole genome sequencing (WGS). Additionally, T cell receptor (TCR) rearrangement state was determined from both single cell T cell receptor sequencing (scTCR-seq) and single cell RNA sequencing (scRNA-seq). Data from both the discovery and extension cohorts were combined, giving a total of 58 individuals with T-ALL. Samples were collected across longitudinal timepoints (day 0, day 7, day 28, relapse, remission) where available and profiled by one or more sequencing modalities (scRNA-seq, scTCR-seq, WGS, bulk RNA-seq) (**Figure 5.1A**).

Of these 58 individuals in the full cohort, only 54 had WGS performed on a diagnostic sample (**Figure 5.1B**). Of these 54, four individuals (L026, L028, L029, L088) had their read alignment files (BAM files) stored on a clinical database and were not available to me. Instead, variants of these four individuals were obtained by Dr Angus Hodder from the clinical database. As I could not access the WGS BAM files for these four individuals for further analysis, their variants were only used to determine the genomic subtype of their leukaemia. Of the remaining 50 individuals whose WGS BAM files were available to me, 39 also had WGS done on a remission sample and variant calling was run with a matched normal to remove germline variants. One individual (T024) did not have a remission sample, but had a day 28 sample with <0.1% blasts and this sample was used a matched normal for variant calling. Three individuals (P030, T014, T029) did not have a remission sample, but had a day 28 sample with <50% blasts. Variant calling was run unmatched for these three individuals, and germline variants could be separated from somatic variants based on their differences in allele frequency. Finally, seven individuals (P021, P080, P084, L084, T004, T019, T023), did not have any sample which had <50% blasts. Hence, variant calling for them was run unmatched and it was impossible to distinguish somatic variants from germline variants on the basis of

allele frequency alone. For these individuals, variants which occur on known leukaemia genes or known noncoding mutation hotspots were reported as leukaemia driver mutations.

Overall, WGS was performed on 70 samples which contain blasts and variant callers were run on these samples as tumour samples. However, some of these samples had low blast content (from clinical flow cytometry) and their sequencing depth may be insufficient for variant callers to reliably call somatic mutations. Variant callers typically require at least 3 mutant reads to call a somatic mutation. I calculated the average number of mutant reads in diploid regions based on the blast content and sequencing depth of each sample (Tarabichi et al., 2021). For seven samples, this value is less than 3 (**Figure 5.1C**) and it is unlikely that variant callers are able to reliably call somatic mutations in these samples, unless the blast content was underestimated by clinical flow cytometry.

Figure 5.1

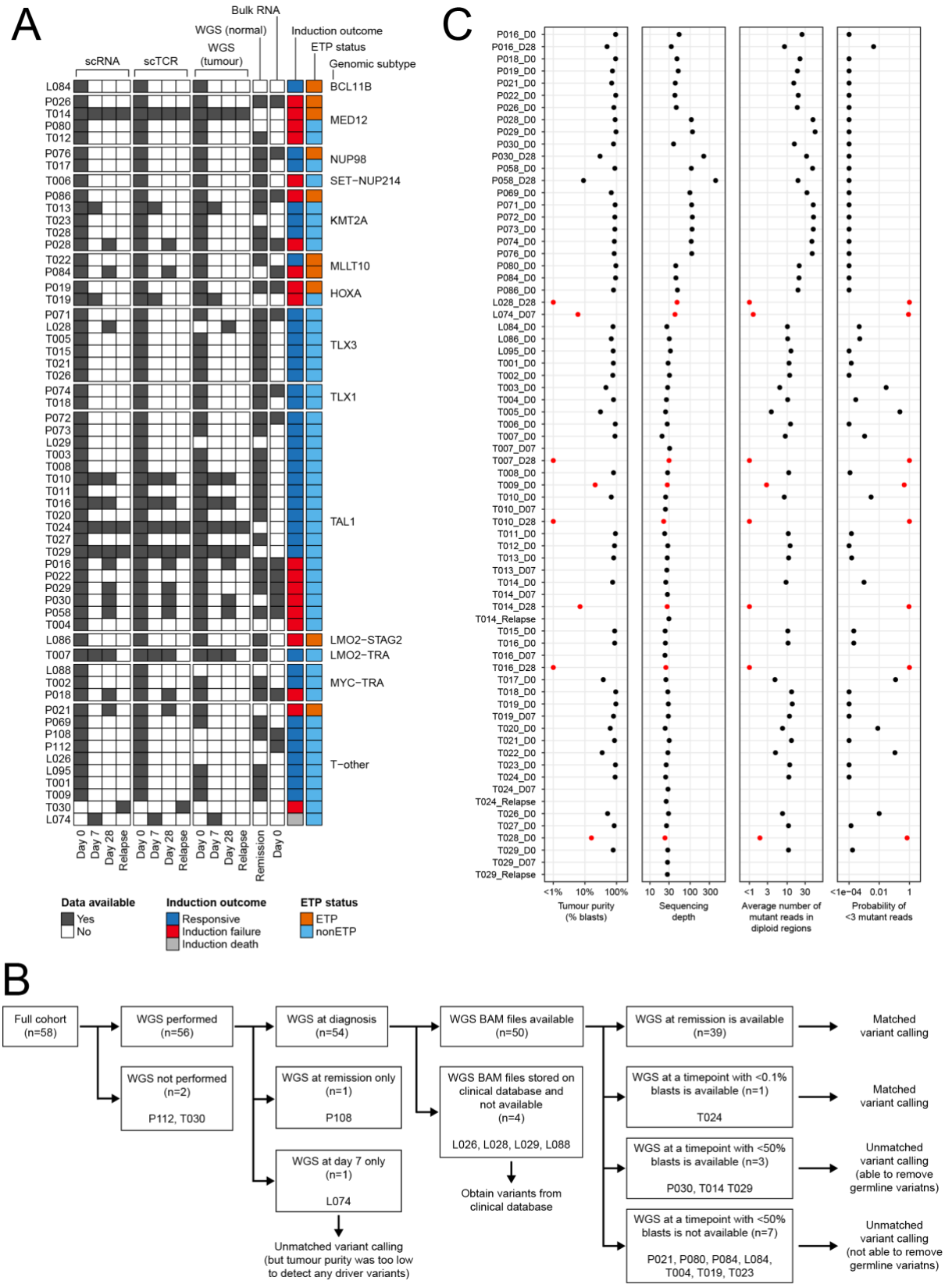

A
Bulk RNA
WGS (normal)
Induction outcome
WGS (tumour)
ETP status
scRNA
scTCR
Genomic subtype
BCL11B
MED12
NUP98
SET−NUP214
KMT2A
MLLT10
HOXA
TLX3
TLX1
TAL1
LMO2−STAG2
LMO2−TRA
MYC−TRA
T−other
Day 0
Day 7
Day 28
Relapse
Remission
Data available
Yes
No
Induction outcome
Responsive
Induction failure
Induction death
ETP status
ETP
nonETP
C
Tumour purity (% blasts)
Sequencing depth
Average number of mutant reads in diploid regions
Probability of <3 mutant reads
B
Full cohort (n=58)
WGS performed (n=56)
WGS not performed (n=2)
P112, T030
WGS at diagnosis (n=54)
WGS at remission only (n=1)
P108
WGS at day 7 only (n=1)
L074
Unmatched variant calling (but tumour purity was too low to detect any driver variants)
WGS BAM files available (n=50)
WGS BAM files stored on clinical database and not available (n=4)
L026, L028, L029, L088
Obtain variants from clinical database
WGS at remission is available (n=39)
Matched variant calling
WGS at a timepoint with <0.1% blasts is available (n=1)
T024
Matched variant calling
WGS at a timepoint with <50% blasts is available (n=3)
P030, T014 T029
Unmatched variant calling (able to remove germline variatns)
WGS at a timepoint with <50% blasts is not available (n=7)
P021, P080, P084, L084, T004, T019, T023
Unmatched variant calling (not able to remove germline variatns)

Figure 5.1: Overview of genomic sequencing data.

(**A**) Overview of sequencing modalities and associated clinical metadata. A sequencing modality is indicated as data available only if the alignment file is available for my analysis.

(**B**) Quality control information of WGS data of samples which contain blasts. First panel from left: tumour purity (blast percentage found on clinical flow cytometry). Second panel: average sequencing depth in WGS. Third panel: average number of mutant reads in diploid regions. Fourth panel: Probability of obtaining less than 3 mutant reads on WGS in diploid regions. Dots in red are those with less than 3 mutant reads on average in diploid regions and indicate samples for which it is unlikely to find somatic mutations on WGS due to low tumour purity and/or insufficient sequencing depth.

scRNA, single cell RNA sequencing; scTCR, single cell T cell receptor sequencing; WGS, whole genome sequencing; ETP, early T cell precursor; BAM, binary alignment map.

5.3 Driver mutations and genomic subtypes

Using WGS data, I called all classes of somatic mutations: single nucleotide variants (SNV), sequence insertions and deletions (indel), structural variants and copy number alterations. Examination of SNV and indel mutation burdens across 40 individuals, for which matched variant calling was performed, revealed one individual, T012, with a hypermutator phenotype (**Figure 5.2**). T012 was subsequently found to possess germline mutations in the DNA mismatch repair gene *PMS2* – a splice site mutation and a p.P246fs*3 frameshift mutation.

Copy number analysis identified T-ALL genes with copy number gains and losses, as well as copy-neutral loss of heterozygosity (LOH) where the loss of one allele was followed by the duplication of the remaining allele such that the total copy number remains unchanged. The most common copy number alterations in my T-ALL cohort are: chromosome 9p loss and copy-neutral LOH (targetting *CDKN2A*), chromosome 12p loss (*ETV6* and *CDKN1B*), and chromosome 17q gain (its role in T-ALL is unknown) (**Figure 5.3A-D**).

Overlapping all classes of somatic mutation calls with the genomic footprints of T-ALL genes as well as known loci of T-ALL noncoding driver mutations yielded a catalogue of driver mutations across 50 individuals with available WGS data (**Figure 5.4**). The most commonly mutated gene was the cell cycle regulator *CDKN2A*, consistent with previous studies (Liu et al., 2017). Inactivation of this gene typically occurred in a two-hit process – a copy-neutral LOH event on chromosome 9p followed by a focal deletion of the gene – although other classes of mutations were also seen in my data.

The genomic subtype of each T-ALL was determined from the catalogue of driver mutations. Canonical T-ALL subtypes, defined by translocations and other structural variants involving their associated transcription factors (*TAL1*, *TLX3*, *TLX1*, *KMT2A*, *MLLT10*, *NUP98*, *SET* and the *HOXA* genes), were found in my cohort. The TAL1 subtype was the most common genomic subtype in my data. This includes those with *TAL1-STIL* fusion and those driven by a noncoding mutation which generate a new super-enhancer for *TAL1*. The TAL1 subtype was found to have co-occurring mutations in *LMO1*, *LMO2* and *PTEN*. Loss of function mutation in *MED12* is characteristic of a newly defined subtype of T-ALL (Pölönen et al., 2024). I found evidence of *MED12* inactivating mutations (either splice site or stop gain) in four individuals (P026, T012, T014, P080). The T-ALL genomic subtype of five individuals (P021, L095, P069, T001, T009) could not be determined even though driver mutations were detected in their leukaemia WGS data, and they were classified as T-other.

Figure 5.2

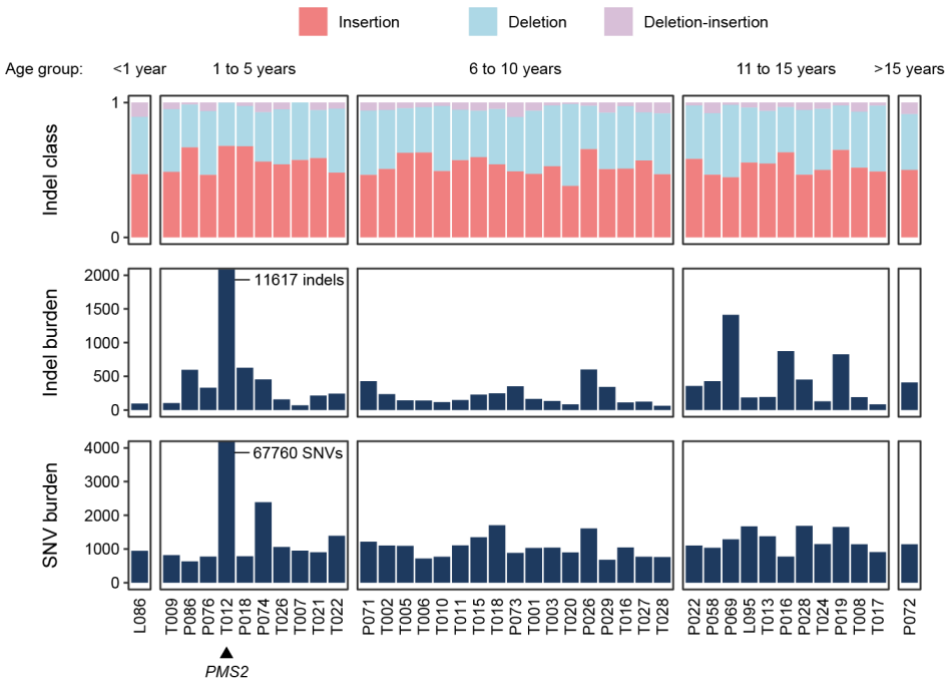


Figure 5.2: Mutational burden in T-ALL genomes at diagnosis.

Top: Proportion of indel mutations belonging to each indel class. Middle: Indel mutation burden. Bottom: SNV mutation burden. Individuals are faceted by their age group. Only 40 individuals, for which matched variant calling of SNVs and indels were performed, are shown here. T012 has a hypermutator phenotype due to germline mutations in the DNA mismatch repair gene *PMS2* (splice mutation and p.P246fs*3 frameshift mutation).

Figure 5.3

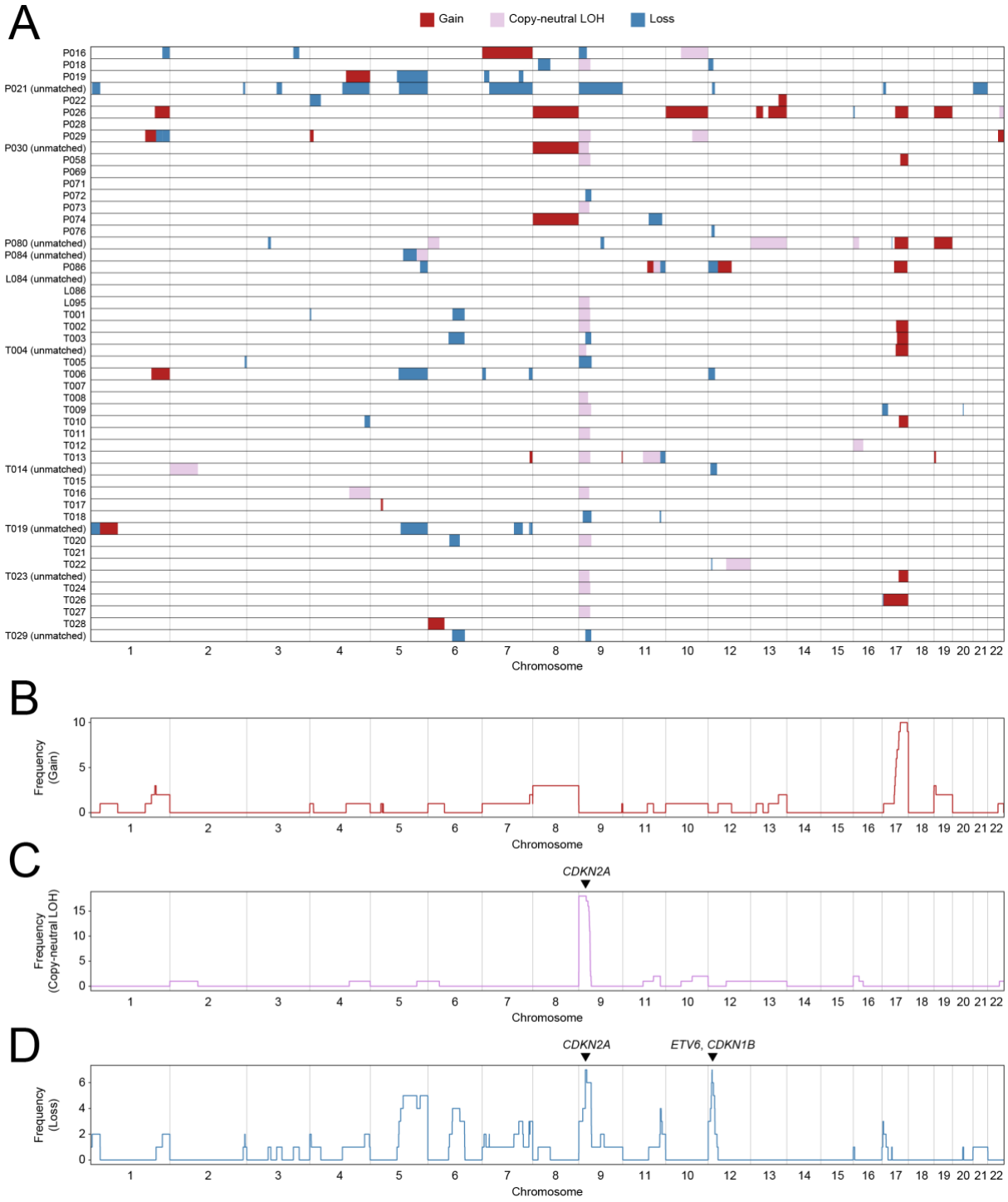

A
Gain
Copy-neutral LOH
Loss
P016
P018
P019
P021 (unmatched)
P022
P026
P028
P029
P030 (unmatched)
P058
P069
P071
P072
P073
P074
P076
P080 (unmatched)
P084 (unmatched)
P086
L084 (unmatched)
L086
L095
T001
T002
T003
T004 (unmatched)
T005
T006
T007
T008
T009
T010
T011
T012
T013
T014 (unmatched)
T015
T016
T017
T018
T019 (unmatched)
T020
T021
T022
T023 (unmatched)
T024
T026
T027
T028
T029 (unmatched)
1 2 3 4 5 6 7 8 9 11 10 12 13 14 15 16 17 18 19 20 21 22
Chromosome
B
Frequency (Gain)
10
5
0
1 2 3 4 5 6 7 8 9 11 10 12 13 14 15 16 17 18 19 20 21 22
Chromosome
C
CDKN2A
Frequency (Copy-neutral LOH)
15
10
5
0
1 2 3 4 5 6 7 8 9 11 10 12 13 14 15 16 17 18 19 20 21 22
Chromosome
D
CDKN2A
ETV6, CDKN1B
Frequency (Loss)
6
4
2
0
1 2 3 4 5 6 7 8 9 11 10 12 13 14 15 16 17 18 19 20 21 22
Chromosome

Figure 5.3: Copy number landscape in T-ALL genomes at diagnosis.

(**A**) Copy number altered segments across chromosomes 1 to 22 in T-ALL genomes of 50 individuals at diagnosis. Copy number profiles for 40 individuals with a matched normal were obtained using ASCAT and Battenberg. For the remaining 10 individuals without a matched normal ('unmatched'), they were obtained using PURPLE.

(**B**) Histogram showing frequency of gains across chromosomes 1 to 22.

(**C**) Histogram showing frequency of copy-neutral LOH across chromosomes 1 to 22.

(**D**) Histogram showing frequency of losses across chromosomes 1 to 22.

LOH, loss of heterozygosity.

Figure 5.4

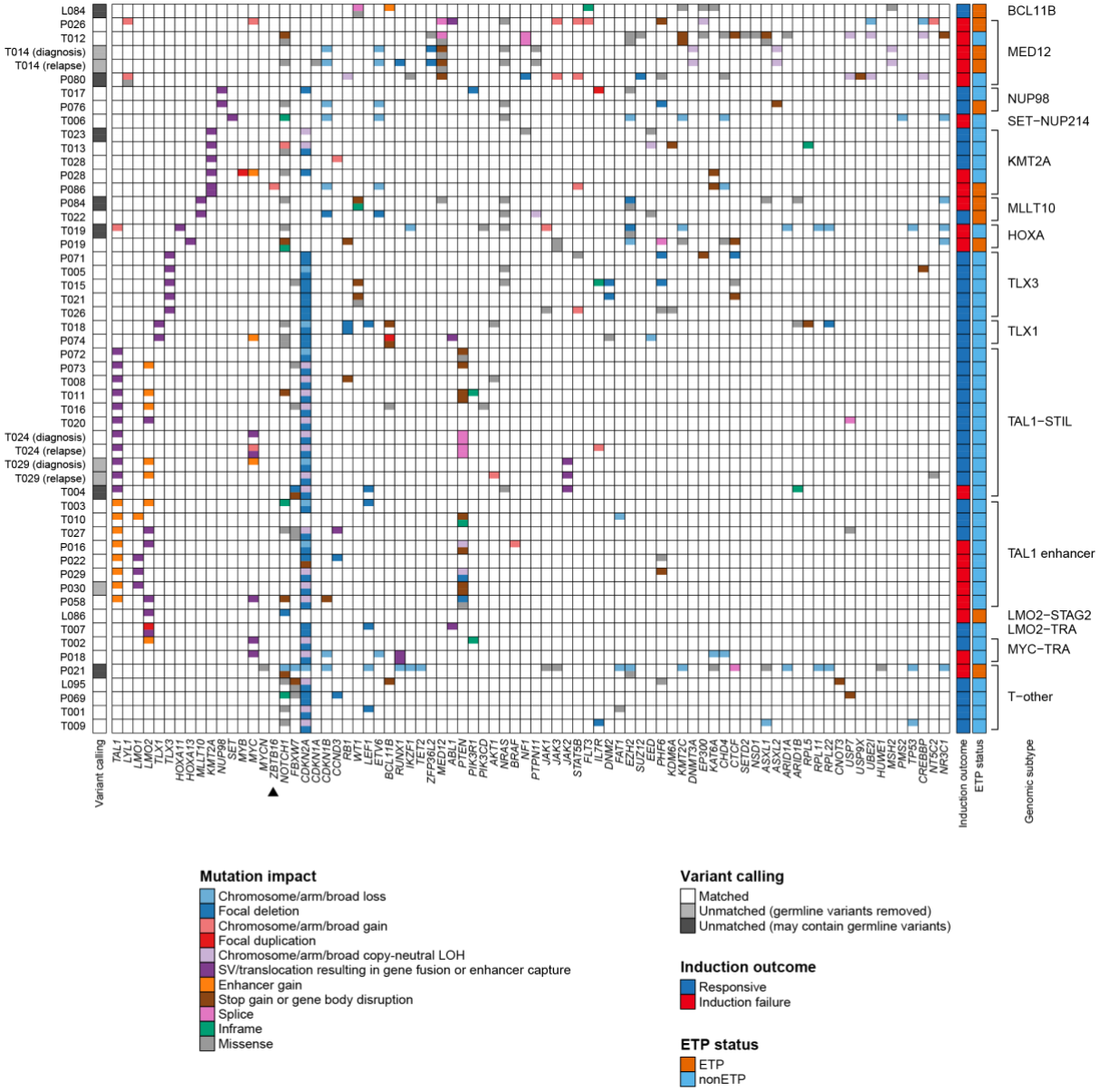

L084
P026
T012
T014 (diagnosis)
T014 (relapse)
P080
T017
P076
T006
T023
T013
T028
P028
P086
P084
T022
T019
P019
P071
T005
T015
T021
T026
T018
P074
P072
P073
T008
T011
T016
T020
T024 (diagnosis)
T024 (relapse)
T029 (diagnosis)
T029 (relapse)
T004
T003
T010
T027
P016
P022
P029
P030
P058
L086
T007
T002
P018
P021
L095
P069
T001
T009
BCL11B
MED12
NUP98
SET−NUP214
KMT2A
MLLT10
HOXA
TLX3
TLX1
TAL1−STIL
TAL1 enhancer
LMO2−STAG2
LMO2−TRA
MYC−TRA
T−other
Variant calling
Induction outcome
ETP status
Genomic subtype
TAL1
LYL1
LMO1
LMO2
TLX1
TLX3
HOXA13
MLLT10
KMT2A
NUP98
SET
MYB
MYC
MYCN
ZBTB16
NOTCH1
FBXW7
CDKN2A
CDKN1A
CDKN1B
CCND3
RB1
WT1
LEF1
ETV6
BCL11B
RUNX1
IKZF1
TET2
ZFP36L2
MED12
ABL1
PTEN
PIK3R1
PIK3CD
AKT1
NRAS
BRAF
NF1
PTPN11
JAK1
JAK3
JAK2
STAT5B
FLT3
IL7R
DNM2
FAT1
EZH2
SUZ12
EED
PHF6
KDM6A
KMT2C
DNMT3A
EP300
KAT6A
CHD4
CTCF
SETD2
NSD1
ASXL1
ASXL2
ARID1A
ARID1B
RPL5
RPL11
RPL22
CNOT3
USP7
USP9X
UBE2T
HUWE1
MSH2
PMS2
TP53
CREBBP
KDM5C
NR3C1
Mutation impact
Chromosome/arm/broad loss
Focal deletion
Chromosome/arm/broad gain
Focal duplication
Chromosome/arm/broad copy-neutral LOH
SV/translocation resulting in gene fusion or enhancer capture
Enhancer gain
Stop gain or gene body disruption
Splice
Inframe
Missense
Variant calling
Matched
Unmatched (germline variants removed)
Unmatched (may contain germline variants)
Induction outcome
Responsive
Induction failure
ETP status
ETP
nonETP

Figure 5.4: Driver mutations in T-ALL genomes at diagnosis and relapse.

Driver mutations in 50 individuals for which WGS BAM files were available for me to run variant calling. For majority of samples, matched variant calling was performed and only somatic variants are shown here. Where unmatched variant calling was done, if a sample with low blast content was available, germline variants could be identified and removed, otherwise the driver mutation catalogue for that individual may contain germline variants. In samples with more than two mutations in the same gene, only the two with the highest priority (from top to bottom of mutation impact) are shown. The only somatic mutation in *ZBTB16* that was found in my cohort was a copy number gain in P086.

LOH, loss of heterozygosity; SV, structural variant; ETP, early T cell precursor.

5.4 Relationship between refractory *ZBTB16*+ blasts and T-ALL genomic subtype

By analysing single leukaemia blast expression of *ZBTB16* at diagnosis across different genomic subtypes in my cohort (**Figure 5.5**), I found that certain T-ALL subtypes tended to present with *ZBTB16*+ blasts at diagnosis, such as the MED12, NUP98 and MLLT10 subtypes; these subtypes are associated with early differentiation states of T cell development and have higher frequency of the ETP immunophenotype. *ZBTB16*+ blasts were also enriched in a case of *STAG2-LMO2* rearranged T-ALL, which has recently been described as a high-risk T-ALL subtype (Kimura et al., 2024). *ZBTB16*+ blasts could also arise from late differentiation state genomic subtypes like those driven by enhancer mutations on *TAL1*. Apart from *ZBTB16* copy number gain in P086, which may have accounted for *ZBTB16*+ blasts in this individual, no *ZBTB16* somatic mutations were found in my cohort (**Figure 5.4**).

Within the *TAL1* mutated subtype, it has recently been described that some cases possess additional cooperative mutations in the Myc signalling pathway (*MYC*, *NMYC*) or the Ras signalling pathway (*KRAS*, *NRAS*, *BRAF*, *NF1*, *PTPN11*) (O'Connor et al., 2023). I found five of such cases in my cohort (**Figure 5.5**): T004 (*NRAS* missense mutation), T029 (*MYC* enhancer gain), T024 (*MYC* translocation), P016 (*BRAF* copy number gain) and P058 (*MYC* translocation). There was no association with either *ZBTB16* expression or induction failure, although this was a very small sample size.

To extend beyond the limited cohort size of my scRNA-seq data, I turned towards a recently published bulk RNA-seq dataset of over 1,300 individuals with T-ALL from the COG AALL0434 cohort. (Pölönen et al., 2024), restricting my analysis to 1,175 diagnostic samples with >60% blast content. I analysed *ZBTB16* expression, as a measure of the presence of *ZBTB16*+ blasts, across 15 genomic subtypes defined by the authors (**Figure 5.6A**), I found high *ZBTB16* expression in the LMO2 gd-like and STAG2/LMO2 subtypes, consistent with results from my scRNA-seq data.

Certain genomic subtypes (ETP-like, TAL1 ab-like, TAL1 DP-like, TLX3, NKX2-1) were further subdivided by the authors into fine genomic subtypes based on their co-occurring genomic lesion and their disease risk group (very high risk, high risk, low risk, very low risk). I examined *ZBTB16* expression across these risk-defining genomic subtypes (**Figure 5.6B**). *ZBTB16* expression was elevated in the very high risk LMO2 gd-like and KMT2A ETP-like subtypes, which have high rates of induction failure. In contrast, *ZBTB16* expression was lower in the SPI1 subtype, despite it being also classified as very high risk. This could be because the SPI1 subtype has good induction response and its poor disease prognosis is attributed to early relapse from secondary myeloid malignancies (Pölönen et al., 2024), whereas *ZBTB16*+ blasts are refractory towards induction.

Plotting *ZBTB16* expression against day 28 residual disease burden (which represents the actual clinical response to induction treatment) for each risk-defining genomic subtype reveals heterogeneity within some of these subtypes (MLLT10 ETP-like, Rare ETP-like, ZFP36L2 ETP-like, TAL1 DP-like other), where high *ZBTB16* expression at diagnosis identifies leukaemias within these genomic subtypes which will be refractory to induction treatment (**Figure 5.6C**). Overall, my analyses suggest that *ZBTB16*+ blasts are strongly associated with LMO2 gd-like and STAG2/LMO2 subtypes, but may also arise from diverse genomic subtypes of T-ALL.

Figure 5.5

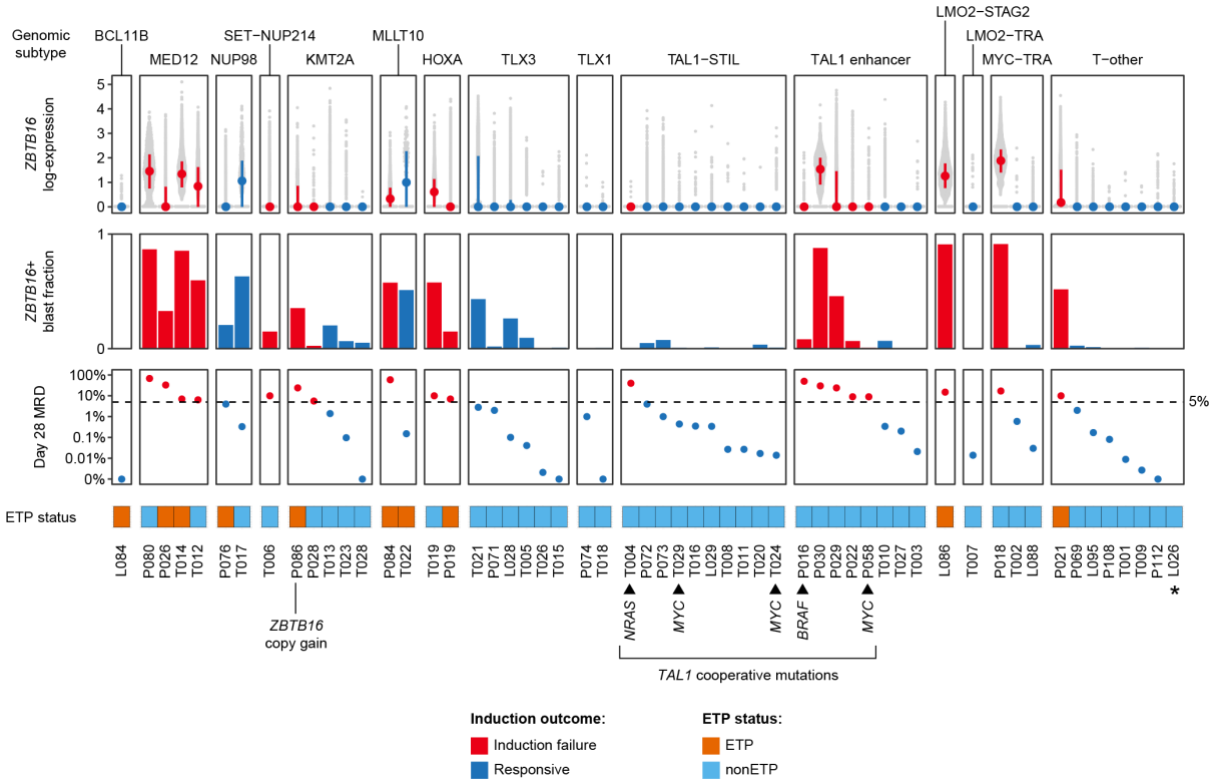


Figure 5.5: Single leukaemia blast *ZBTB16* expression across genomic subtypes.

ZBTB16 expression in single leukaemia blasts and *ZBTB16*+ blast fraction were plotted across 58 diagnostic samples from my scRNA-seq cohort, faceted by genomic subtype. For each sample, median expression and interquartile range are shown by the large central dot and vertical line, respectively. The TAL1 subtype was divided into those with *TAL1-STIL* fusion and those driven by enhancer gain mutations on *TAL1*. Day 28 MRD and ETP status are indicated. Five cases of *TAL1* mutated subtype have cooperative mutations in the Myc or Ras signalling pathways. The leukaemia of P086 has a copy number gain in *ZBTB16*. L026 is a child with T-lymphoblastic lymphoma who was found by CT scan to have responded completely at day 28.

MRD, minimal residual disease; ETP, early T cell precursor.

Figure 5.6A-B

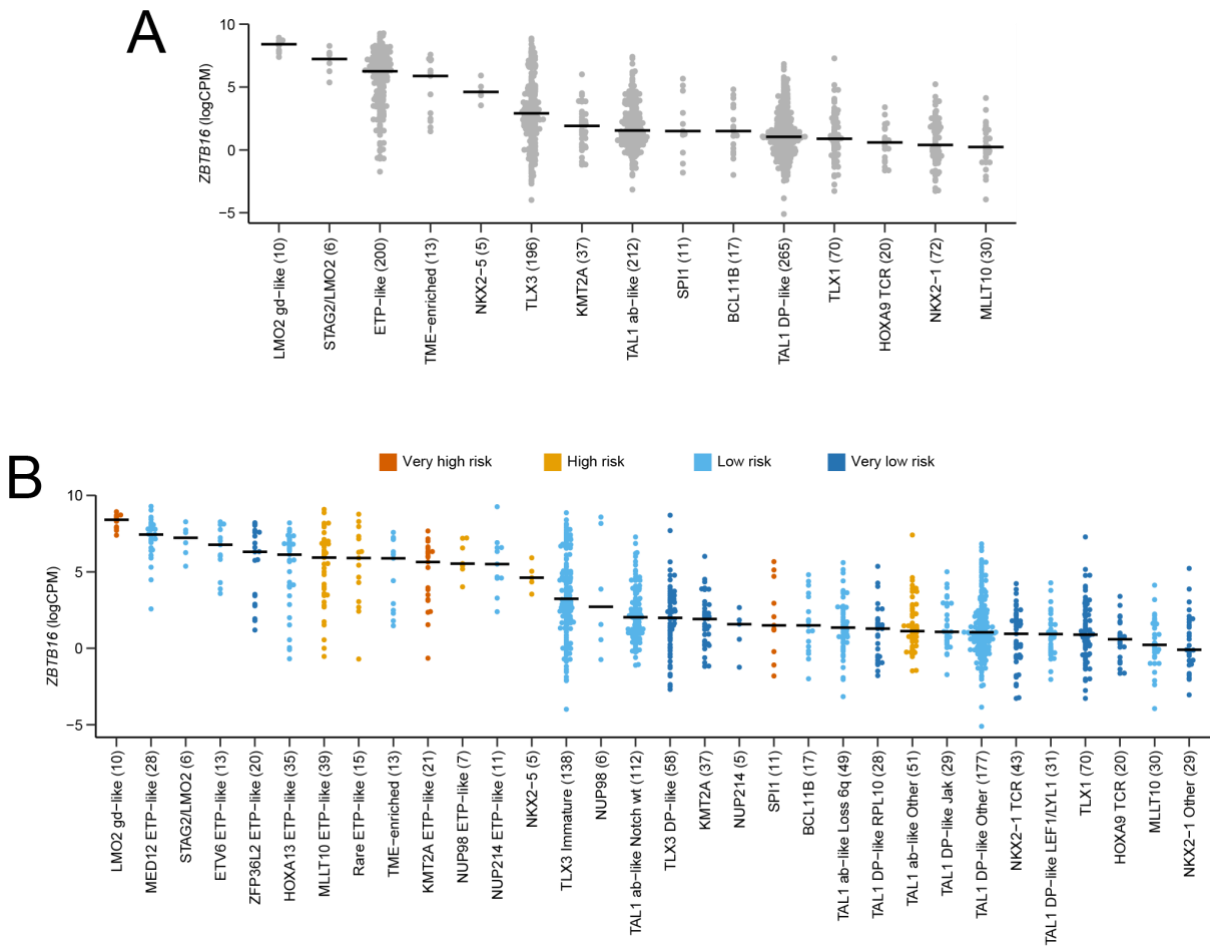

A
ZBTB16 (logCPM)
10
5
0
−5
LMO2 gd−like (10)
STAG2/LMO2 (6)
ETP−like (200)
TME−enriched (13)
NKX2−5 (5)
TLX3 (196)
KMT2A (37)
TAL1 ab−like (212)
SPI1 (11)
BCL11B (17)
TAL1 DP−like (265)
TLX1 (70)
HOXA9 TCR (20)
NKX2−1 (72)
MLLT10 (30)
B
Very high risk
High risk
Low risk
Very low risk
ZBTB16 (logCPM)
LMO2 gd−like (10)
MED12 ETP−like (28)
STAG2/LMO2 (6)
ETV6 ETP−like (13)
ZFP36L2 ETP−like (20)
HOXA13 ETP−like (35)
MLLT10 ETP−like (39)
Rare ETP−like (15)
TME−enriched (13)
KMT2A ETP−like (21)
NUP98 ETP−like (7)
NUP214 ETP−like (11)
NKX2−5 (5)
TLX3 Immature (138)
NUP98 (6)
TAL1 ab−like Notch wt (112)
TLX3 DP−like (58)
KMT2A (37)
NUP214 (5)
SPI1 (11)
BCL11B (17)
TAL1 ab−like Loss 6q (49)
TAL1 DP−like RPL10 (28)
TAL1 ab−like Other (51)
TAL1 DP−like Jak (29)
TAL1 DP−like Other (177)
NKX2−1 TCR (43)
TAL1 DP−like LEF1/LYL1 (31)
TLX1 (70)
HOXA9 TCR (20)
MLLT10 (30)
NKX2−1 Other (29)

Figure 5.6C

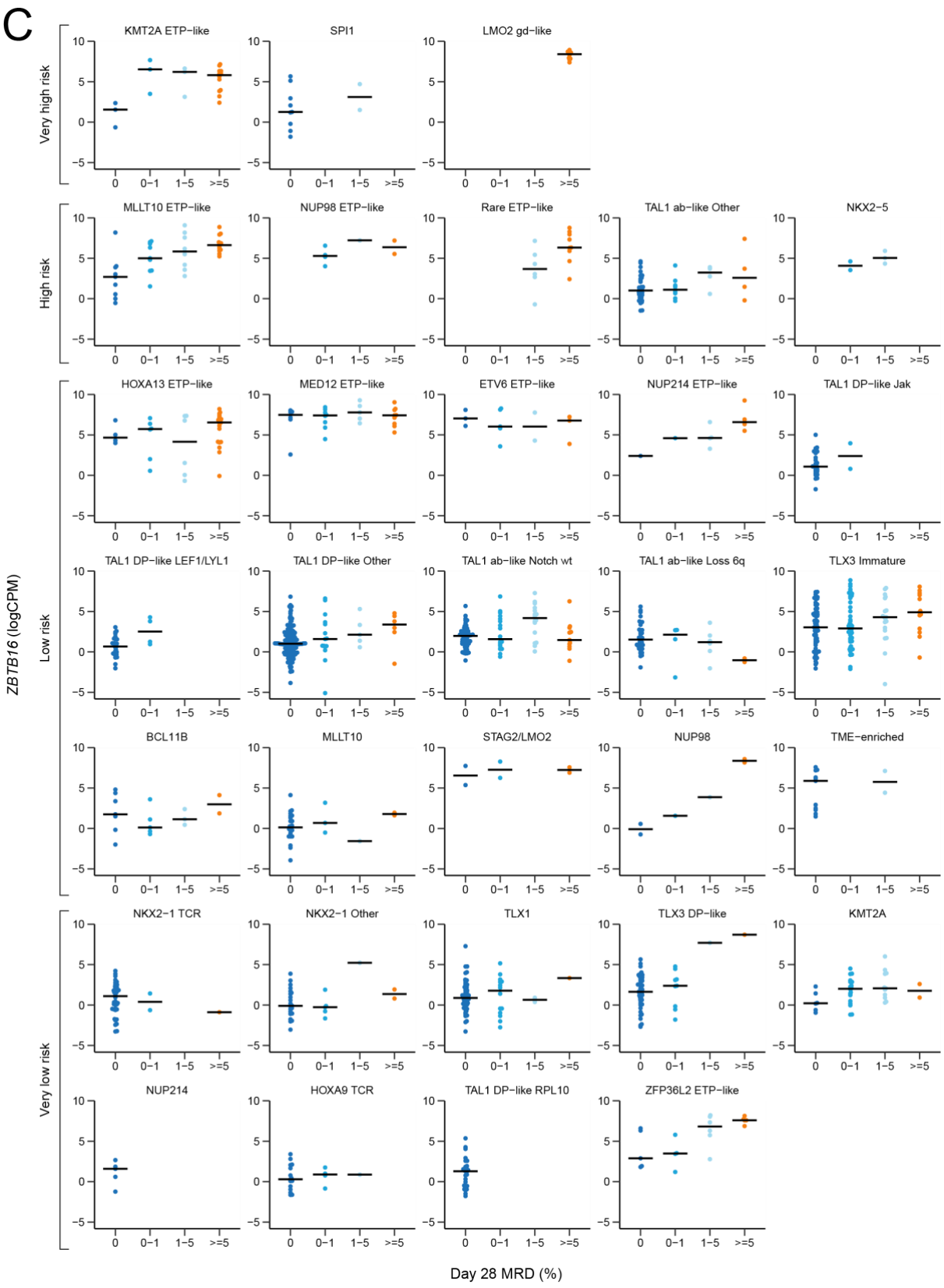

C
Very high risk
High risk
Low risk
Very low risk
ZBTB16 (logCPM)
KMT2A ETP−like
SPI1
LMO2 gd−like
MLLT10 ETP−like
NUP98 ETP−like
Rare ETP−like
TAL1 ab−like Other
NKX2−5
HOXA13 ETP−like
MED12 ETP−like
ETV6 ETP−like
NUP214 ETP−like
TAL1 DP−like Jak
TAL1 DP−like LEF1/LYL1
TAL1 DP−like Other
TAL1 ab−like Notch wt
TAL1 ab−like Loss 6q
TLX3 Immature
BCL11B
MLLT10
STAG2/LMO2
NUP98
TME−enriched
NKX2−1 TCR
NKX2−1 Other
TLX1
TLX3 DP−like
KMT2A
NUP214
HOXA9 TCR
TAL1 DP−like RPL10
ZFP36L2 ETP−like
10
5
0
−5
0
0−1
1−5
>=5
Day 28 MRD (%)

Figure 5.6: *ZBTB16* expression across genomic subtypes in the COG AALL0434 study.

(**A**) Swarm plots showing *ZBTB16* expression across 15 genomic subtypes in 1,175 individuals from the COG AALL0434 cohort. Horizontal bars represent median *ZBTB16* expression values for each subtype. The subtypes are sorted from highest to lowest median values. Number of individuals in each subtype are indicated in parentheses.

(**B**) Swarm plots showing *ZBTB16* expression across 32 risk-defining genomic subtypes in 1,175 individuals from the COG AALL0434 cohort. Horizontal bars represent median *ZBTB16* expression values for each subtype. The subtypes are sorted from highest to lowest median values. Number of individuals in each subtype are indicated in parentheses. Dots are coloured by risk categories (very high risk, high risk, low risk, very low risk) defined by the authors of the COG AALL0434 study.

(**C**) Swarm plots showing *ZBTB16* expression plotted against day 28 MRD categories and faceted by the 32 risk-defining genomic subtypes in 1,175 individuals from the COG AALL0434 cohort. Horizontal bars represent median *ZBTB16* expression values for each day 28 MRD category within each subtype.

MRD, minimal residual disease.

5.5 Relationship between refractory *ZBTB16*+ blasts and TCR rearrangement state

T cells rearrange and express their individual TCR chains (*TRA*, *TRB*, *TRG*, *TRD*) in a stepwise order as they develop (**Figure 1.1**). T cells expressing alpha-beta TCR (ab-TCR) will rearrange *TRB* first, followed by *TRA*. T cells with gamma-delta TCR (gd-TCR) constitute a distinct lineage from those with ab-TCR; the sequence of *TRG* and *TRD* rearrangements in these cells is unclear. Therefore, TCR rearrangement state of T-ALL blasts can be used as a marker of its T cell lineage and its stage of developmental arrest.

I determined TCR rearrangement state by detecting TCR transcripts expressed by T-ALL blasts, using both scRNA-seq and scTCR-seq. I analysed TCR transcripts from scRNA-seq using TRUST4 (Song et al., 2021) and TCR transcripts from scTCR-seq using Dandelion (Suo et al., 2024), as described in my methods chapter. The presence of gd-TCR can only be detected on scRNA-seq, as scTCR-seq only captures ab-TCR, not gd-TCR.

I found that *ZBTB16*+ blasts were associated with T-ALL which do not express any TCR, or those which express gd-TCR (*TRD* and/or *TRG*) (**Figure 5.7**). *ZBTB16*+ blasts can also sometimes arise from cells expressing both *TRB* and *TRA*, and more rarely from cells expressing *TRB* only. Blasts with *TRB* only also express *PTCRA*, which encodes a substitute for the *TRA* chain to pair up with the rearranged *TRB* chain, as part of T cell development that I described earlier in my introduction chapter. *PTCRA* is not expressed in blasts with other TCR states. *DNTT*, which encodes an enzyme involved in the process of TCR rearrangement, is expressed in blasts across all TCR states.

To investigate the relationship between *ZBTB16*+ blasts and TCR state in a larger cohort, I analysed data of 1,175 individuals from the COG AALL0434 cohort (Pölönen et al., 2024), using TCR states annotated by the authors (**Figure 5.8A-B**). *ZBTB16* expression was on average higher in T-ALL with no TCR and those with gd-TCR, although a wide range of *ZBTB16* expression levels were observed in T-ALL across the different TCR rearrangement states. Among T-ALL with no TCR and those with gd-TCR, *ZBTB16* expression was higher in the leukaemias which were refractory to induction (**Figure 5.8C**). Therefore, while *ZBTB16*+ blasts are associated with early precursor stages (no TCR) and gd-TCR, they represent a cell state that is not entirely captured by TCR rearrangement state alone.

Figure 5.7

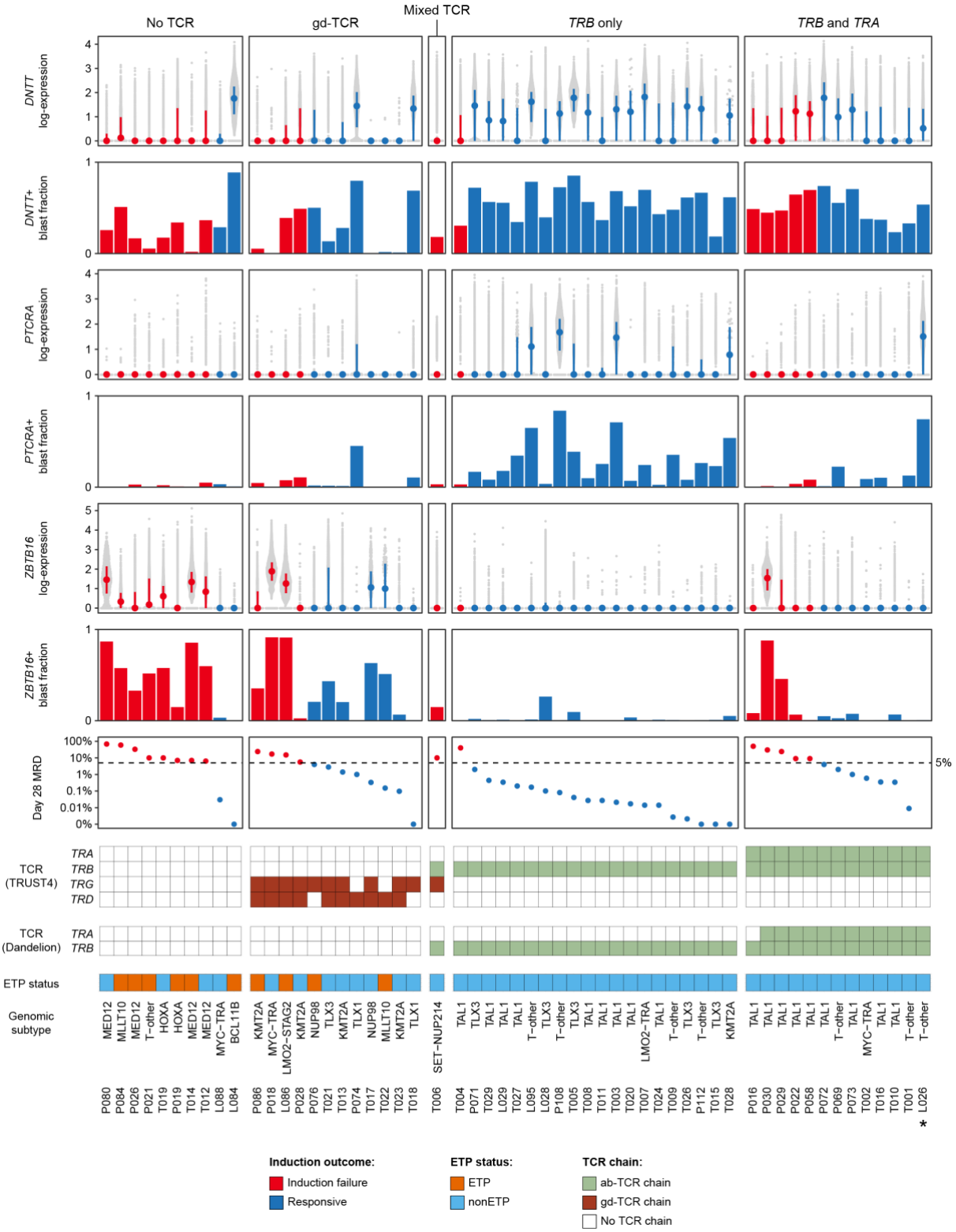

No TCR
gd-TCR
Mixed TCR
TRB only
TRB and TRA
DNTT log-expression
DNTT+ blast fraction
PTCRA log-expression
PTCRA+ blast fraction
ZBTB16 log-expression
ZBTB16+ blast fraction
Day 28 MRD
100%
10%
1%
0.1%
0.01%
0%
5%
TCR (TRUST4)
TRA
TRB
TRG
TRD
TCR (Dandelion)
ETP status
Genomic subtype
MED12 MED12 MED12 T−other HOXA HOXA MED12 MED12 MYC−TRA BCL11B
KMT2A MYC−TRA LMO2−STAG2 KMT2A NUP98 TLX3 KMT2A TLX1 NUP98 MLLT10 KMT2A TLX1
SET−NUP214
TAL1 TLX3 TAL1 TAL1 TAL1 T−other TLX3 T−other TLX3 TAL1 TAL1 TAL1 LMO2−TRA TAL1 T−other TLX3 T−other TLX3 KMT2A
TAL1 TAL1 TAL1 TAL1 TAL1 T−other TAL1 MYC−TRA TAL1 TAL1 T−other T−other
P080 P084 P026 P021 T019 P019 T014 T012 L088 L084
P086 P018 L086 P028 P076 T021 T013 P074 T017 T022 T023 T018
T006
T004 P071 T029 L029 T027 L095 L028 P108 T005 T008 T011 T003 T020 T007 T024 T009 T026 P112 T015 T028
P016 P030 P029 P022 P058 P072 P069 P073 T002 T016 T010 T001 L026 *
Induction outcome:
Induction failure
Responsive
ETP status:
ETP
nonETP
TCR chain:
ab-TCR chain
gd-TCR chain
No TCR chain

Figure 5.7: Single leukaemia blast *ZBTB16* expression across TCR rearrangement states.

Expression of *ZBTB16*, *PTCRA* and *DNTT* in single leukaemia blasts, and the fraction of blasts which express these genes, were plotted across 58 diagnostic samples from my scRNA-seq cohort, faceted by TCR rearrangement state categories. For each sample, the median expression and interquartile range are shown by the large central dot and vertical line, respectively. Day 28 MRD, ETP status and genomic subtype are indicated. The *TRA* chain of P016 leukaemia uses a *TRDV* gene segment in place of *TRAV*; this was detected by TRUST4 but not by Dandelion. L026 is a child with T-lymphoblastic lymphoma who was found by CT scan to have responded completely at day 28.

TCR, T cell receptor; MRD, minimal residual disease; ETP, early T cell precursor.

Figure 5.8

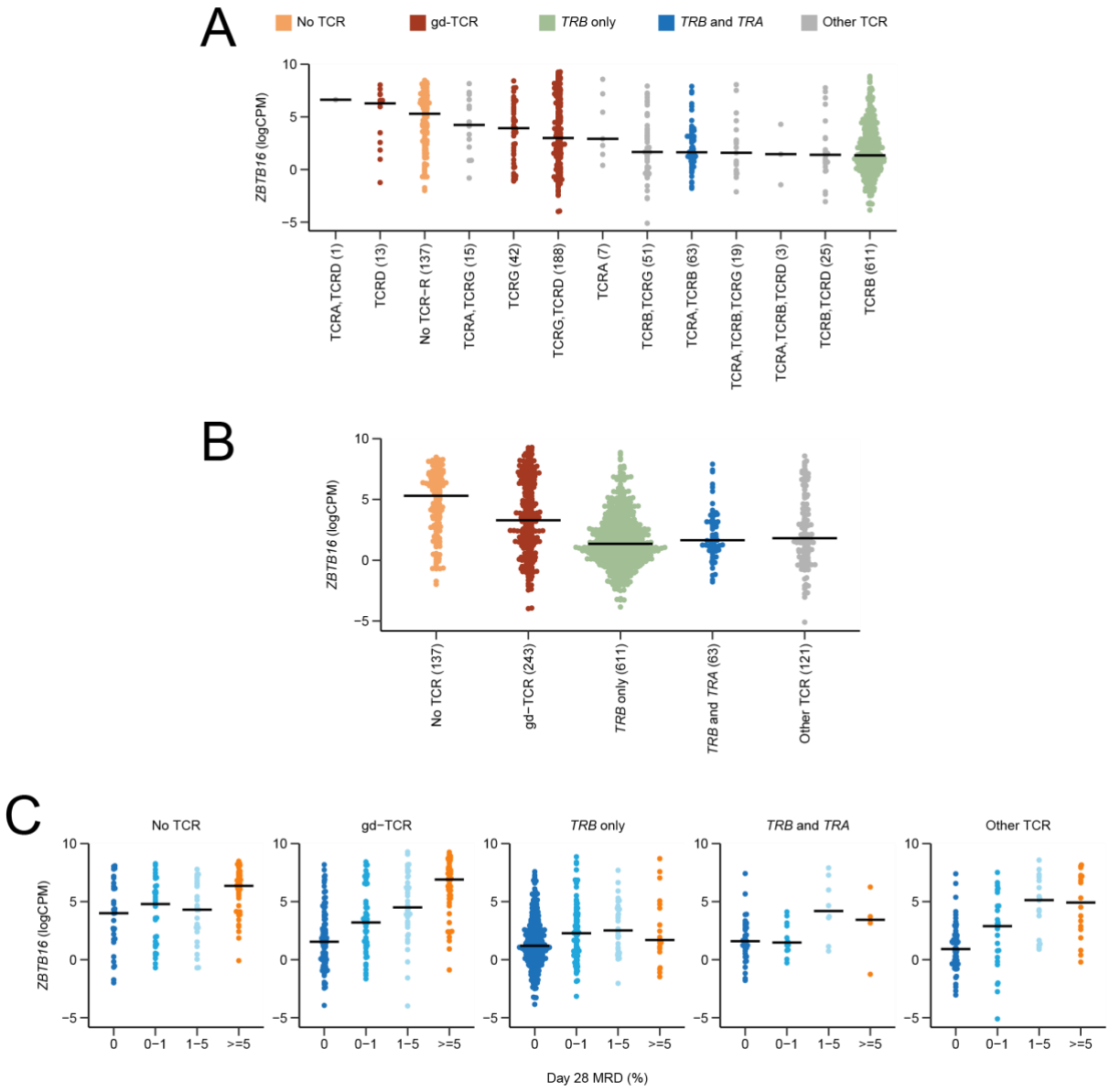

A
No TCR
gd-TCR
TRB only
TRB and TRA
Other TCR
ZBTB16 (logCPM)
TCRA,TCRD (1)
TCRD (13)
No TCR (137)
TCRA,TCRG (15)
TCRG (42)
TCRG,TCRD (188)
TCRA (7)
TCRB,TCRG (51)
TCRA,TCRB (63)
TCRA,TCRB,TCRG (19)
TCRA,TCRB,TCRD (3)
TCRB,TCRD (25)
TCRB (611)
B
ZBTB16 (logCPM)
No TCR (137)
gd−TCR (243)
TRB only (611)
TRB and TRA (63)
Other TCR (121)
C
No TCR
gd−TCR
TRB only
TRB and TRA
Other TCR
ZBTB16 (logCPM)
0
0−1
1−5
>=5
Day 28 MRD (%)

Figure 5.8: *ZBTB16* expression across TCR rearrangement states in the COG AALL0434 study.

(**A**) Swarm plots showing *ZBTB16* expression across TCR rearrangement states in 1,175 individuals from the COG AALL0434 cohort. TCR rearrangement states were annotated by the authors and indicate which TCR chains have been rearranged. TCR rearrangement states are categorised according to their corresponding stage of T cell development and coloured as such. Horizontal bars represent median *ZBTB16* expression values for each TCR state. The TCR states are sorted from highest to lowest median values. Number of individuals in each TCR state are indicated in parentheses.

(**B**) Swarm plots showing *ZBTB16* expression across TCR rearrangement state categories in 1,175 individuals from the COG AALL0434 cohort. TCR rearrangement state categories correspond to stages of T cell development. Horizontal bars represent median *ZBTB16* expression values for each TCR state. Number of individuals in each TCR state are indicated in parentheses.

(**C**) Swarm plots showing *ZBTB16* expression plotted against day 28 MRD categories and faceted by the TCR rearrangement state categories in 1,175 individuals from the COG AALL0434 cohort. Horizontal bars represent median *ZBTB16* expression values for each day 28 MRD category within each TCR rearrangement state category.

TCR, T cell receptor; MRD, minimal residual disease.

5.6 Lack of evidence of noncoding mutations driving *ZBTB16* expression

Aberrant expression of leukaemia genes may be driven by noncoding mutations. These may in turn be detected by searching for recurrently mutated hotspots in the noncoding genome. Refractory T-ALL is associated with high *ZBTB16* expression and it is plausible that *ZBTB16* expression, at least in a subset of cases, may be driven by noncoding mutations. I examined the *ZBTB16* genomic locus and its surrounding region for potential hotspot mutations (**Figure 5.9A-B**). As a positive control, I interrogated the region surrounding the *TAL1* gene, for which a well characterised noncoding mutation hotspot exists which generates a de novo super-enhancer (Mansour et al., 2014). Whereas the noncoding mutation hotspot near the *TAL1* locus could be identified, no such hotspot was found for the *ZBTB16* locus.

Although no noncoding mutation hotspot was found in the genomic region around the *ZBTB16* gene, it does not rule out the possibility that *ZBTB16* expression is driven by one. To this end, I examined for monoallelic expression of *ZBTB16* as evidence that such a noncoding mutation potentially exists (**Figure 5.10A-D**). I performed this analysis on *TAL1* and *LMO1* as positive controls where there are known mutations driving their monoallelic expression in T-ALL. I also performed this analysis on the *CD3E* gene as a negative control, as it is ubiquitously expressed in T-ALL but its expression is not driven by a noncoding mutation. For this analysis, heterozygous single nucleotide polymorphisms (SNPs) were called from WGS of the matched normal sample and their allelic ratios in bulk RNA-seq of the target gene was measured. I found evidence of monoallelic expression of *TAL1* in leukaemias with the *TAL1* enhancer mutation, and of *LMO1* in leukaemias with *LMO1-TRB* translocations, but there was no evidence of monoallelic expression of *CD3E* or *ZBTB16*. However, because I only searched for monoallelic expression of *ZBTB16* in 4 samples where *ZBTB16* is highly expressed, it is possible that monoallelic expression might occur in other samples. Interestingly, although the *TAL1* mutation status of P108 was not known as there was no WGS of its tumour sample, this analysis suggests that there is likely to be a noncoding mutation near the *TAL1* locus in this patient. Therefore, *ZBTB16* expression in T-ALL is more likely to be driven by a trans regulatory mechanism (e.g. transcription factor binding) rather than a cis regulatory mechanism (e.g. noncoding mutation which generates a de novo enhancer)

Figure 5.9

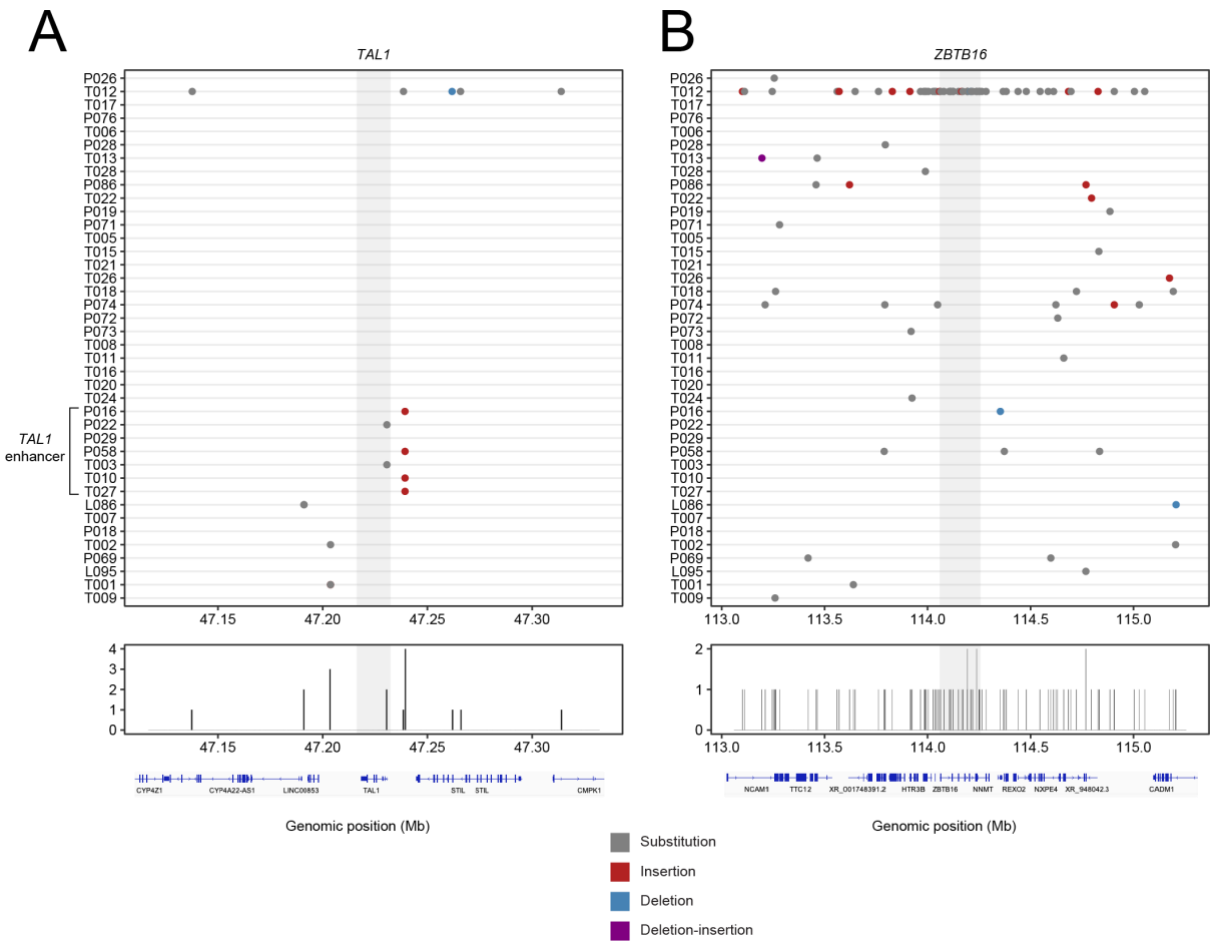


Figure 5.9: Somatic mutation landscape around the *TAL1* and *ZBTB16* loci.

Landscape of somatic mutations around the *TAL1* locus (**A**) and the *ZBTB16* locus (**B**). All classes of sequence mutations (substitution, insertion, deletion, deletion-insertion) called in WGS of diagnostic T-ALL samples are plotted. Top panel: Mutations are plotted along the genomic coordinate (x-axis) for each patient (y-axis) and coloured according to mutation class. The gene bodies of *TAL1* and *ZBTB16* are highlighted as a grey background. T-ALL driven by a mutation which generates a de novo *TAL1* enhancer are indicated. Middle panel: Histogram of mutation frequency along genomic coordinate to identify mutation hotspots. The gene bodies of *TAL1* and *ZBTB16* are highlighted as a grey background. Bottom panel: Gene track showing *TAL1*, *ZBTB16* and neighbouring genes.

Figure 5.10A-B

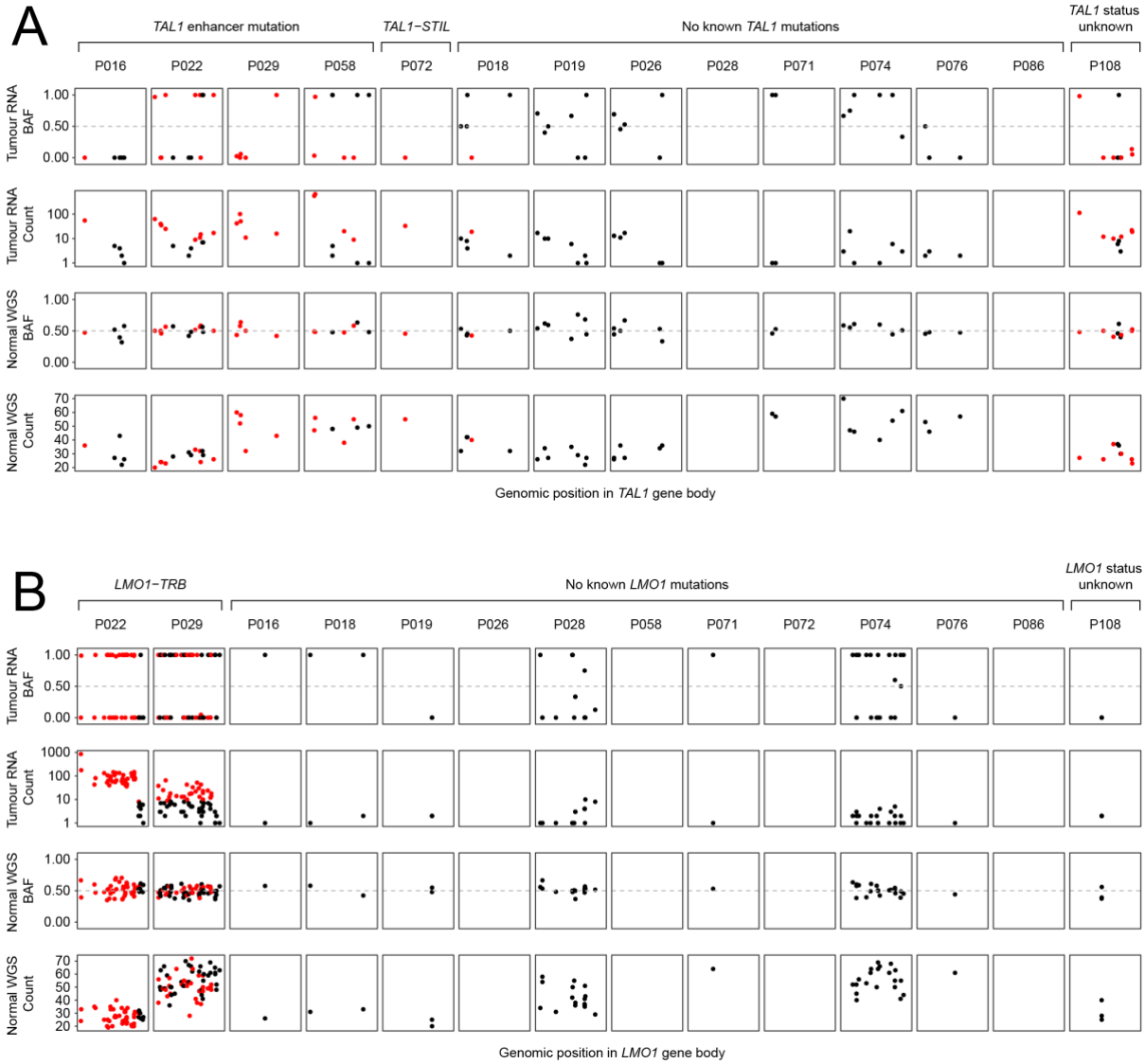


Figure 5.10C-D

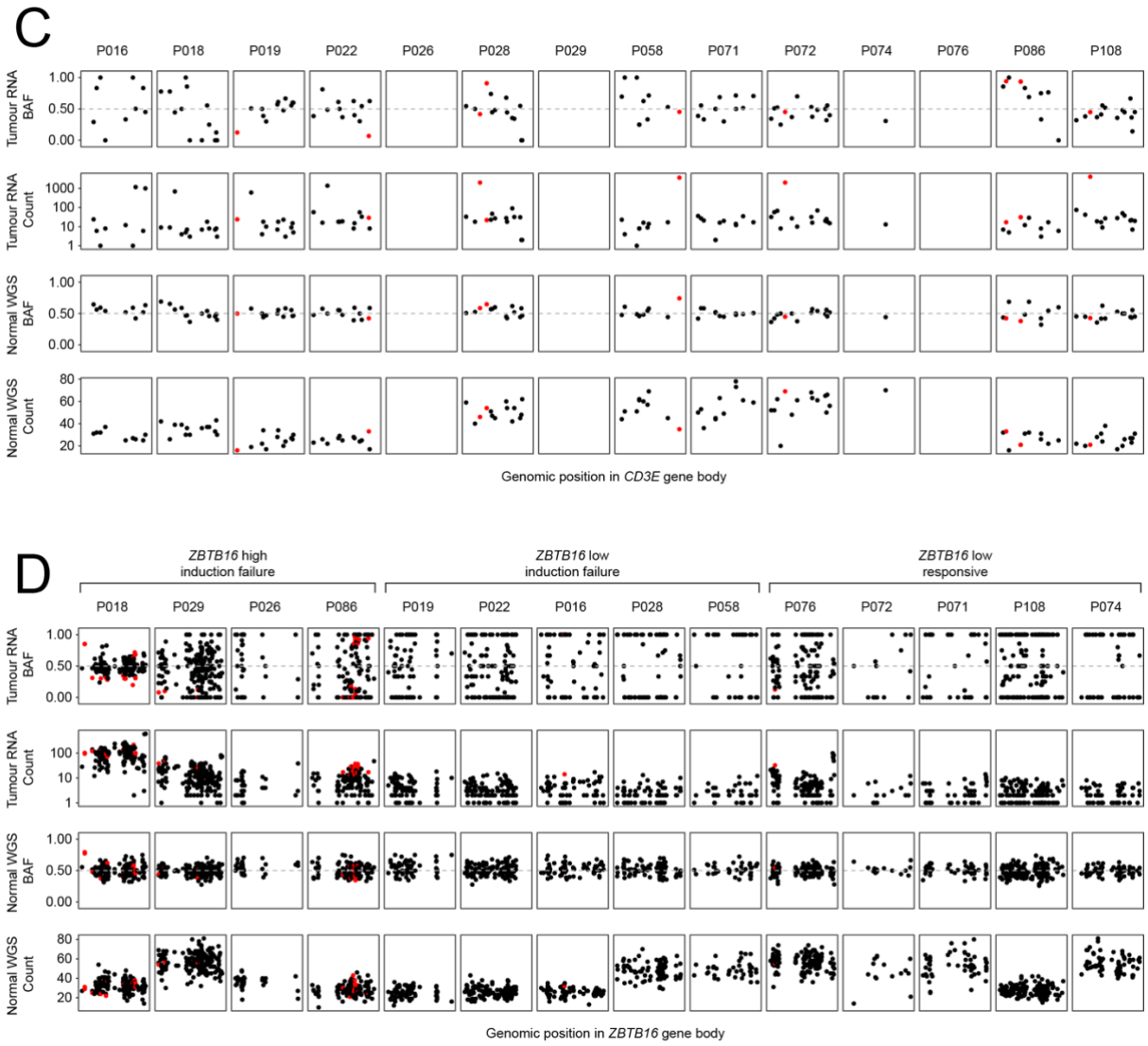


Figure 5.10: Allele specific expression in T-ALL genes and *ZBTB16*.

Detection of allele specific expression in *TAL1* (**A**), *LMO1* (**B**)*, CD3E* (**C**) and *ZBTB16* (**D**). Each dot represents a heterozygous SNP derived from WGS of the matched normal sample which has coverage in the gene of interest on bulk RNA-seq of the tumour sample. Dots in red are SNPs with a statistically significant deviation of the tumour bulk RNA-seq BAF from 0.5, calculated using a two-sided exact binomial test with Benjamini-Hochberg multiple hypothesis correction (false discovery rate of 0.01). The mutational status of *TAL1* and *LMO1* in P108 are not known as WGS of the tumour sample was not available.

Top panel: BAF in bulk RNA-seq of tumour sample.
Second panel: Read count in bulk RNA-seq of tumour sample.
Third panel: BAF in WGS of matched normal sample.
Bottom panel: Read count in WGS of matched normal sample.

BAF, B allele frequency.

5.7 Origin of refractory *ZBTB16*+ blasts in P058

P058 is a child whose day 0 and day 28 leukaemia blasts clustered separately on a UMAP (**Figure 4.1B**). While the day 0 blasts did not express *ZBTB16*, the day 28 blasts were *ZBTB16*+, bearing the refractory T-ALL phenotype which I described in the previous chapter. It is possible that these blasts acquired a refractory *ZBTB16*+ transcriptome as a result of cancer cell plasticity, and are otherwise genetically identical. Alternatively, the day 28 blasts could represent a distinct refractory clone which was positively selected for by induction chemotherapy and outcompeted the original diagnostic clone to become the dominant clone at day 28. Under this alternative hypothesis, the day 0 and day 28 blasts will exhibit genetic differences. Therefore, I aimed to reconstruct the leukaemia phylogeny of P058 using three different sets of somatic mutations: copy number alterations, SNVs and TCR rearrangements.

The copy number profile of the high tumour purity day 0 sample of P058 (88% blasts by clinical flow cytometry) was determined from WGS, revealing copy-neutral LOH at chromosome 9p and gain at chromosome 17q (**Figure 5.11A-B**). In contrast, tumour purity of the day 28 sample of P058 (9% blasts by clinical flow cytometry) was too low to fit its copy number profile using its WGS data alone, despite being sequenced to 429x depth (**Figure 5.11C**). Instead, I phased SNPs on the chromosome 9p and 17q segments, using WGS data from the day 0 sample, to identify the major and minor allelic genotypes at these SNPs (Trinh et al., 2022). By examining WGS data with allelic phasing, I found that although the chromosome 9p copy-neutral LOH was present at day 28, the chromosome 17q gain was unexpectedly absent (**Figure 5.11D-E**). Hence, the dominant clones at day 0 and at day 28 are not genetically identical as they bear different copy number states. However, this does not rule out the possibility that the dominant clone at day 28 is derived from the dominant clone at day 0.

Analysis of somatic SNVs provides many more genetic markers to resolve the phylogenetic relationship between the day 0 and day 28 leukaemia of P058. Following SNV calling by CaVEMan (Jones et al., 2016) and filtering as described in the methods chapter, I obtained 813 and 766 SNVs from the day 0 and day 28 samples, respectively. I took the union of these two sets SNVs, arriving at a total of 1210 SNVs. Based on the distribution of the variant allele frequencies (VAF) of the SNVs, I estimated the day 0 sample tumour purity to be 84%, consistent with 88% blasts reported by clinical flow cytometry (**Figure 5.12A**). In contrast, the VAF distribution at day 28 (**Figure 5.12B**) suggested a tumour purity of 5%, which was lower than 9% blasts by clinical flow cytometry. By plotting the VAF at these two timepoints against each other, it was clear there were multiple leukaemia clones (**Figure 5.12C**).

To account for differences in tumour purity and tumour copy number profile, I calculated the cancer cell fraction (CCF) of each SNV at day 0 (**Figure 5.12D**) and day 28 (**Figure 5.12E**), as described in the methods chapter (Tarabichi et al., 2021). SNVs with a CCF of 1.0 are clonal in that sample and those less than 1.0 are subclonal. By plotting the day 0 CCF against the day 28 CCF (**Figure 5.12F**), I could infer three clusters of SNVs: (1) SNVs which are clonal at both timepoints, (2) SNVs private to the day 0 sample, and (3) SNVs private to the day 28 sample.

The leukaemia at day 0 and day 28 in patient P058 are genetically different and can be related to each other by three different possibilities (**Figure 5.13A**). Firstly, the day 28 clone could have arisen from day 0 clone, and in the process acquired a resistant phenotype along with additional somatic mutations (**Figure 5.13A**). Alternatively, the day 0 and day 28 clones could be completely independent leukaemias with completely different sets of somatic mutations. Finally, the day 0 and day 28 leukaemias could be derived from a common precursor. They would thus share some somatic mutations, but each with have its own set of private mutations.

By combining multiple lines of somatic mutation evidence – including SNVs, copy number alterations, TCR rearrangements and cancer driver mutations – I found that the leukaemia phylogeny of P058 belongs to the third case, where the responsive clone at day 0 and the refractory clone at day 28 share a common precursor (**Figure 5.13B**). Interestingly, the two leukaemia clones in this child expressed different somatically rearranged TCR-β chains, which is highly unusual. Both clones do share the same somatic rearrangement on their TCR-γ locus at the genome level, although this TCR chain did not have detectable expression at the transcriptome level.

Both clones shared the same initiating driver lesion which generates a *TAL1* super enhancer (Mansour et al., 2014), among other shared driver mutations (**Figure 5.13B**). However, the responsive clone at day 0 possessed additional mutations – *NOTCH1*, *PTEN* and *MYC-TRA* rearrangement – which were absent in the refractory day 28 clone. There were no driver mutations private to the day 28 clone. Therefore, the only known factor which might explain the refractory phenotype of the day 28 clone was its high *ZBTB16* expression compared to the day 0 clone, which did not express it (**Figure 5.13C**).

By analysing single cell transcriptomes of leukaemia blasts, I demonstrated the presence of a small distinct cluster of *ZBTB16*+ blasts in the day 0 sample of P058 (**Figure 5.13C**). Moreover, by projecting WGS-derived somatic mutations onto single cell leukaemia

transcriptomes, I further showed that this small cluster of *ZBTB16*+ blasts at day 0 belongs to the same phylogenetic clone as the refractory *ZBTB16*+ blasts at day 28.

Firstly, I used TRUST4 (Song et al., 2021) to reconstruct TCR sequences and determine the TCR gene usage of individual leukaemia blasts using reads from scRNA-seq. I found that the small *ZBTB16*+ cluster shared the same TCR-β gene usage as the refractory *ZBTB16*+ blasts at day 28, which was different from the responsive *ZBTB16*− blasts at day 0 (**Figure 5.13D**). Next, although all blast at day 0 and day 28 shared the copy-neutral LOH at chromosome 9p, the chromosome 17q gain was only detected in the *ZBTB16*− blasts at day 0, and absent in the small cluster of *ZBTB16*+ blasts at day 0 (**Figure 5.13E**). Finally, refractory *ZBTB16*+ blasts at day 28 and responsive *ZBTB16*− blasts at day 0 showed evidence in their scRNA-seq reads that they possessed different sets of SNVs, although there was insufficient aggregate coverage in the small *ZBTB16*+ cluster to conclude which set of SNVs it possessed (**Figure 5.13F**).

Taken together, the evidence of somatic mutations in single cell leukaemia transcriptomes suggest that not only are the responsive *ZBTB16*− blasts at day 0 and refractory *ZBTB16*+ blasts at day 28 genetically distinct, but also refractory *ZBTB16*+ blasts were present at diagnosis and comprise around 1% of the blast population there.

Figure 5.11A-C

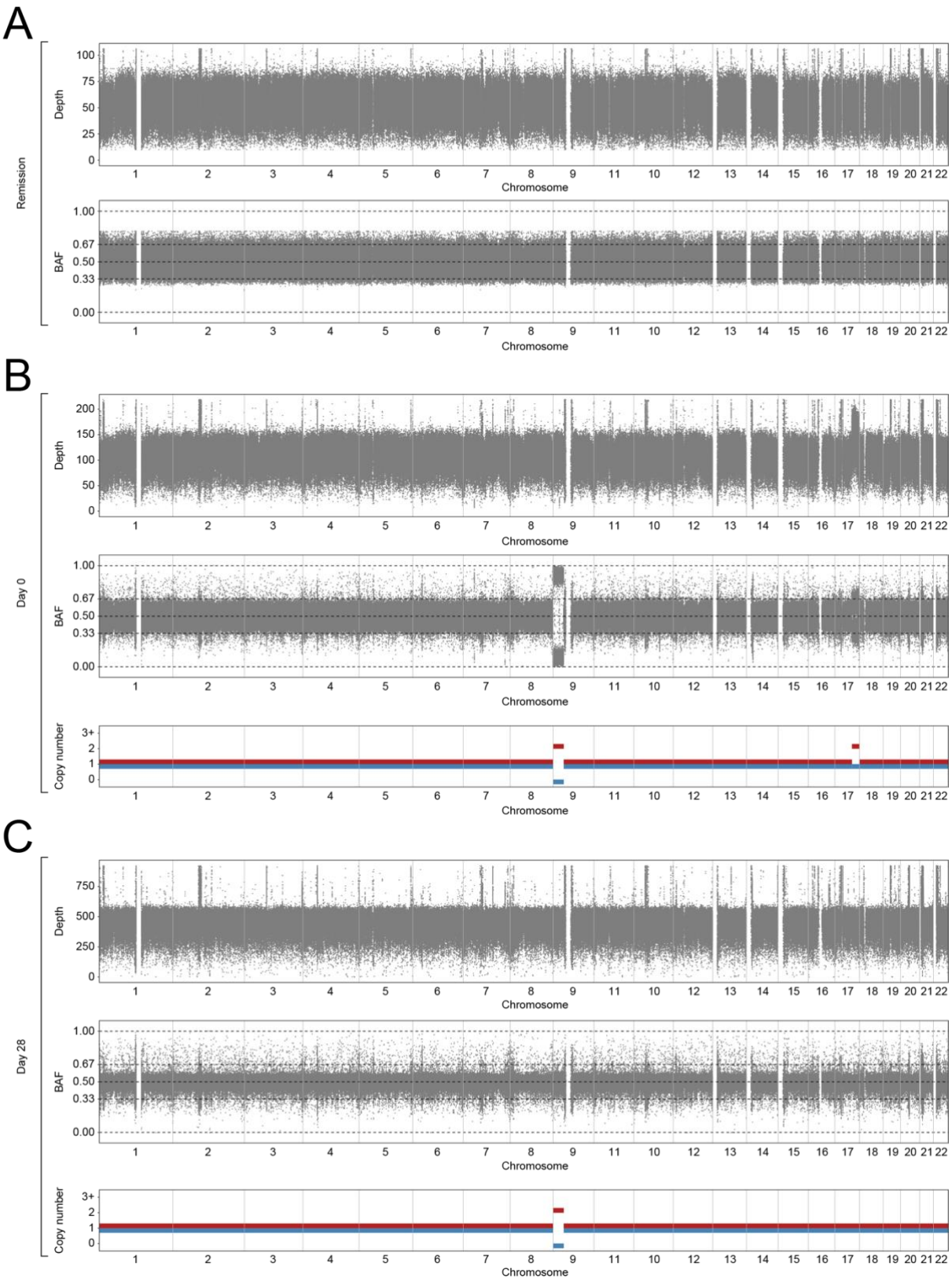


Figure 5.11D-E

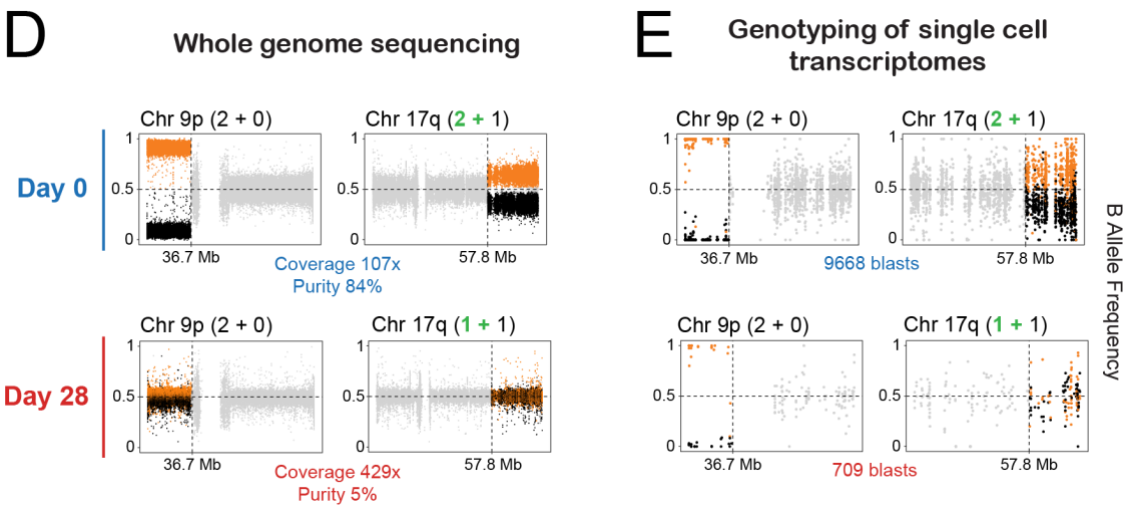


Figure 5.11: Analysis of copy number alterations in P058.

(**A**) P058 remission sample WGS: Read depth (top) and BAF (bottom) of SNPs, plotted across chromosomes 1 to 22.

(**B**) P058 day 0 sample WGS: Read depth (top) and BAF (middle) of SNPs, plotted across chromosomes 1 to 22. Copy number profile (bottom), where line segments indicate integer values of the major (red) and minor (blue) copy number.

(**C**) P058 day 28 sample WGS: Read depth (top) and BAF (middle) of SNPs, plotted across chromosomes 1 to 22. Copy number profile (bottom), where line segments indicate integer values of the major (red) and minor (blue) copy number.

(**D**) Phased BAF of SNPs in WGS, at chromosome 9 and chromosome 17, of the day 0 and day 28 samples. The copy number state of each copy number altered segment is described as X + Y, where X and Y are the major and minor copy numbers respectively.

(**E**) Phased BAF of SNPs in scRNA-seq, aggregated across all leukaemia blasts, at chromosome 9 and chromosome 17, of the day 0 and day 28 samples. The copy number state of each copy number altered segment is described as X + Y, where X and Y are the major and minor copy numbers respectively.

BAF, B allele frequency.

Figure 5.12

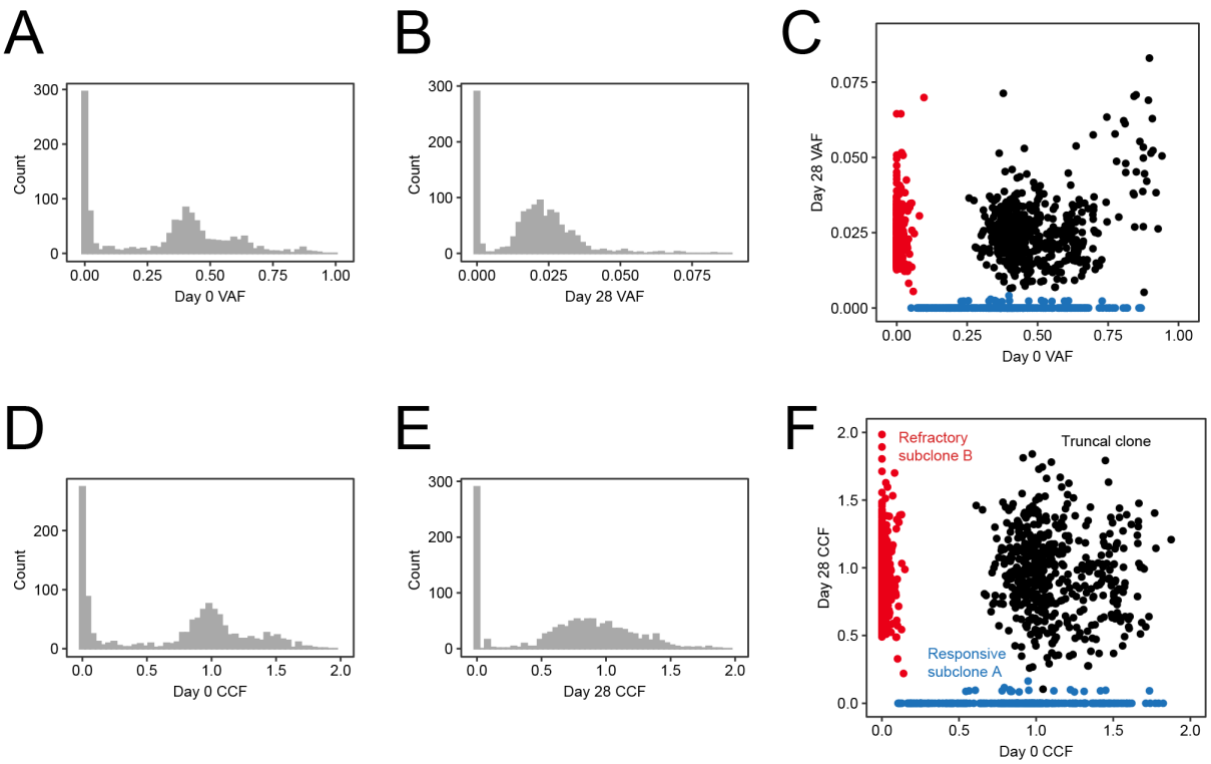


Figure 5.12: Analysis of SNVs in P058.

(**A**) Histogram showing VAF of SNVs at day 0.

(**B**) Histogram showing VAF of SNVs at day 28.

(**C**) Scatterplot showing VAF of SNVs at day 0 (x-axis) and day 28 (y-axis), coloured according to whether the SNV is clonal at day 0 (blue), day 28 (red) or both timepoints.

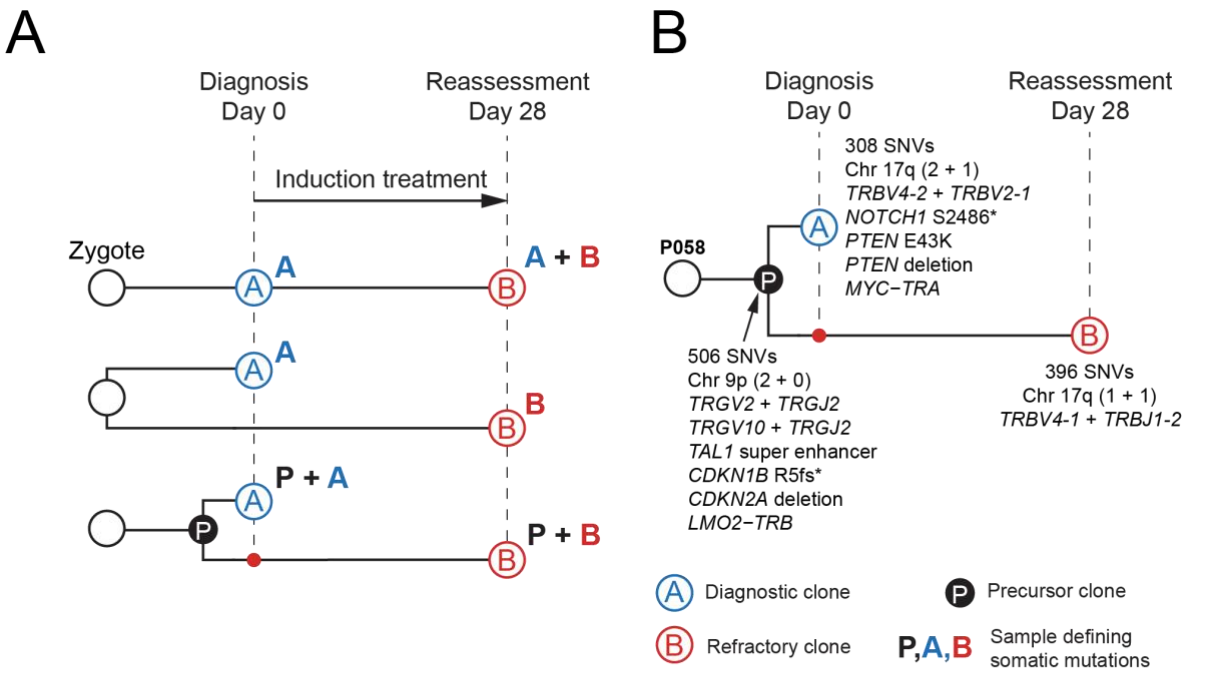
(**D**) Histogram showing CCF of SNVs at day 0.

(**E**) Histogram showing CCF of SNVs at day 28.

(**F**) Scatterplot showing CCF of SNVs at day 0 (x-axis) and day 28 (y-axis), coloured according to whether the SNV is clonal at day 0 (blue), day 28 (red) or both timepoints.

VAF, variant allele frequency; CCF, cancer cell fraction.

Figure 5.13


A
Diagnosis
Day 0
Reassessment
Day 28
Induction treatment
Zygote
A
A + B
A
B
P + A
P + B
B
Diagnosis
Day 0
Reassessment
Day 28
308 SNVs
Chr 17q (2 + 1)
TRBV4-2 + TRBV2-1
NOTCH1 S2486*
PTEN E43K
PTEN deletion
MYC−TRA
P058
506 SNVs
Chr 9p (2 + 0)
TRGV2 + TRGJ2
TRGV10 + TRGJ2
TAL1 super enhancer
CDKN1B R5fs*
CDKN2A deletion
LMO2−TRB
396 SNVs
Chr 17q (1 + 1)
TRBV4-1 + TRBJ1-2
Diagnostic clone
Precursor clone
Refractory clone
P,A,B
Sample defining somatic mutations

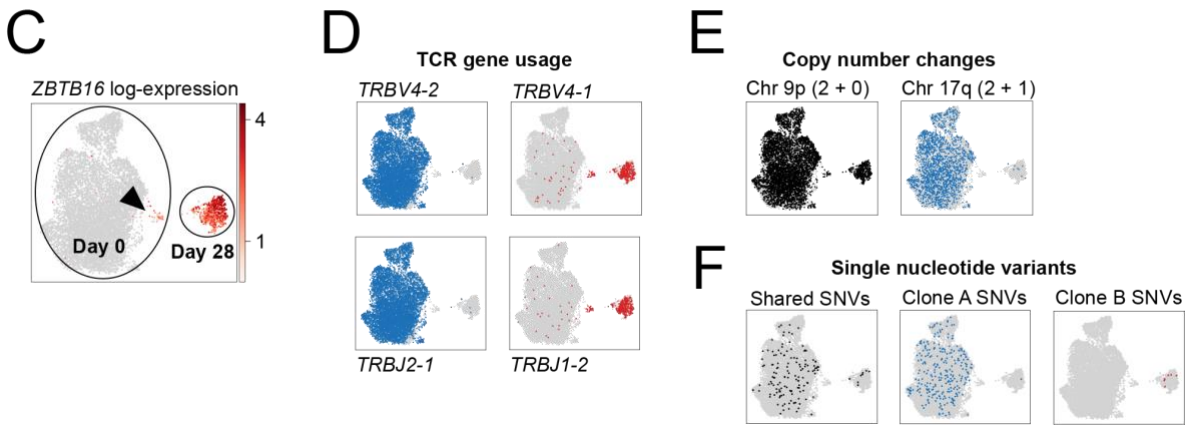

C
ZBTB16 log-expression
4
1
Day 0
Day 28
D
TCR gene usage
TRBV4-2
TRBV4-1
TRBJ2-1
TRBJ1-2
E
Copy number changes
Chr 9p (2 + 0)
Chr 17q (2 + 1)
F
Single nucleotide variants
Shared SNVs
Clone A SNVs
Clone B SNVs

Figure 5.13: Leukaemia phylogenetic reconstruction in P058.

(**A**) Possible leukaemia phylogenies relating clone A at diagnosis (day 0) and clone B at the end of induction treatment (day 28). Top: Clone B directly derives from clone A, and thus possesses all somatic mutations found in clone A, plus additional mutations which were gained in clone B; Middle: Clone B is completely independent of clone A and thus has a completely different set of somatic mutations; Bottom: Clone B and clone A share a common precursor, hence both clones share a common set of somatic mutations found in their common precursor, but each clone has also gained additional mutations of its own.

(**B**) Leukaemia phylogeny in P058, delineating the diagnostic clone A (blue open circle) and the refractory clone B (red open circle). Both clones share a precursor clone (black filled circle). Importantly, clone B was present at diagnosis as a minor clone (small red filled circle). SNVs, copy number alterations, driver mutations and TCR rearrangements are overlayed on top of this phylogeny.

(**C**) UMAP (uniform manifold approximation and projection) showing the expression of *ZBTB16* in day 0 and day 28 blasts from P058. *ZBTB16* expression was largely absent in day 0 blasts, apart from a small cluster indicated by an arrowhead, whereas day 28 blasts strongly expressed *ZBTB16*.

(**D**) UMAPs showing TCR gene usage in day 0 and day 28 blasts from P058, determined by TRUST4 analysis of scRNA-seq data: TCR genes used by clone A (blue) and clone B (red).

(**E**) UMAPs showing the presence of specific copy number alterations (posterior probability ≥0.95) in day 0 and day 28 blasts from P058: chromosome 9 of the precursor clone (black) and chromosome 17 of clone A (blue).

(**F**) UMAPs of day 0 and day 28 blasts from P058, showing the presence of at least one SNV associated with the precursor clone (black), clone A (blue) and clone B (red).

5.8 Origin of refractory *ZBTB16*+ blasts in P030

P030 is another child whose day 28 blasts clustered separately from their day 0 blasts on a UMAP (**Figure 4.1B**). The day 0 and day 28 blasts in P030 could possibly also represent distinct leukaemia clones, similar to P058. No differences in TCR gene usage were seen in the day 0 and day 28 blasts of P030. Hence, I looked towards copy number alterations and SNVs for evidence of multiple leukaemia clones.

To analyse the copy number profile of P030, which did not have a remission sample, I called heterozygous SNPs from WGS of the day 28 sample (**Figure 5.14A**), as it had relatively low 30% blast content by clinical flow cytometry. I then genotyped these heterozygous SNPs in WGS of the day 0 sample (**Figure 5.14B**), which had high 80% blast content by clinical flow cytometry. While copy-neutral LOH at chromosome 9p was strongly implied by both the read depth and BAF evidence, the copy number state of chromosome 8 is less clear. The BAF deviation suggested either chromosome 8 gain or loss of heterozygosity (**Figure 5.14C**), whereas the lack of deviation in read depth pointed towards a diploid state.

To resolve this, the probabilities that each leukaemia blast fits the chromosome 8 loss of heterozygosity, gain of diploid copy number states were calculated from scRNA-seq data, and the most likely copy number state was assigned to each blast (**Figure 5.14D**). This revealed that within both day 0 and day 28 samples, there were three clones of blasts, each with the above three different copy number states. Despite being genetically distinct clones, blasts of all three copy number states expressed *ZBTB16* (**Figure 5.14E**). Therefore, the separate clustering of P030 day 0 and day 28 blasts on the UMAP is unlikely to be attributed to the evolution of copy number clones or *ZBTB16* expression.

Next, I analysed SNVs in P030 to understand the phylogenetic relationship between the three copy number clones. Due to the absence of a remission sample, CaVEMan was run unmatched on WGS of the day 0 and day 28 samples, outputting both germline and somatic SNVs. Germline SNVs were removed using an exact binomial test, as described in the methods chapter. Although I obtained a set of high quality somatic SNVs for analysis (**Figure 5.15A-E**), most of these SNVs were clonal (CCF = 1.0) and the sequencing depth was not sufficient to reliably cluster the subclonal SNVs (CCF < 1.0). Therefore, I could not resolve the phylogenetic relationship between the three copy number clones in P030.

Figure 5.14A-B

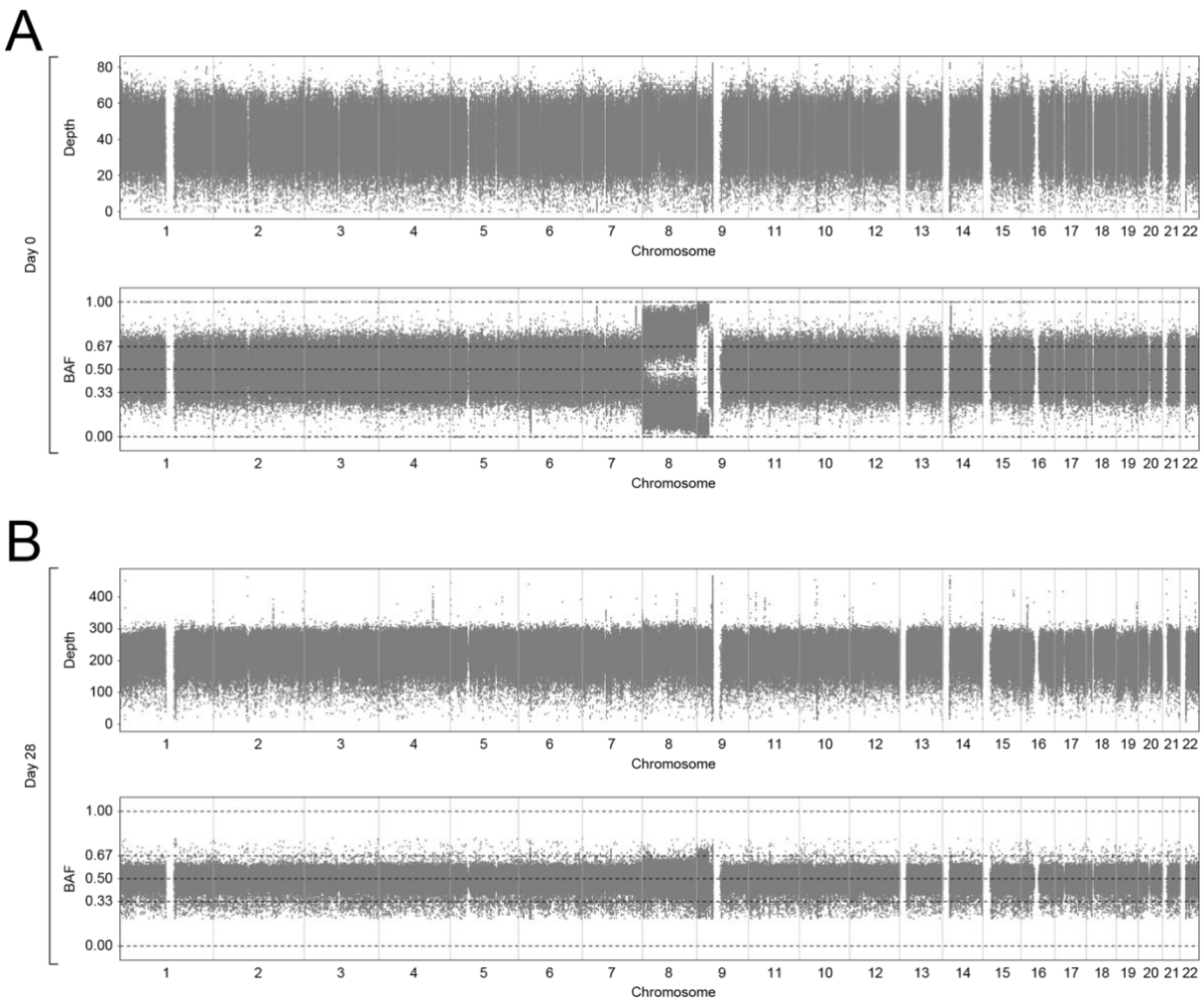


Figure 5.14C-E

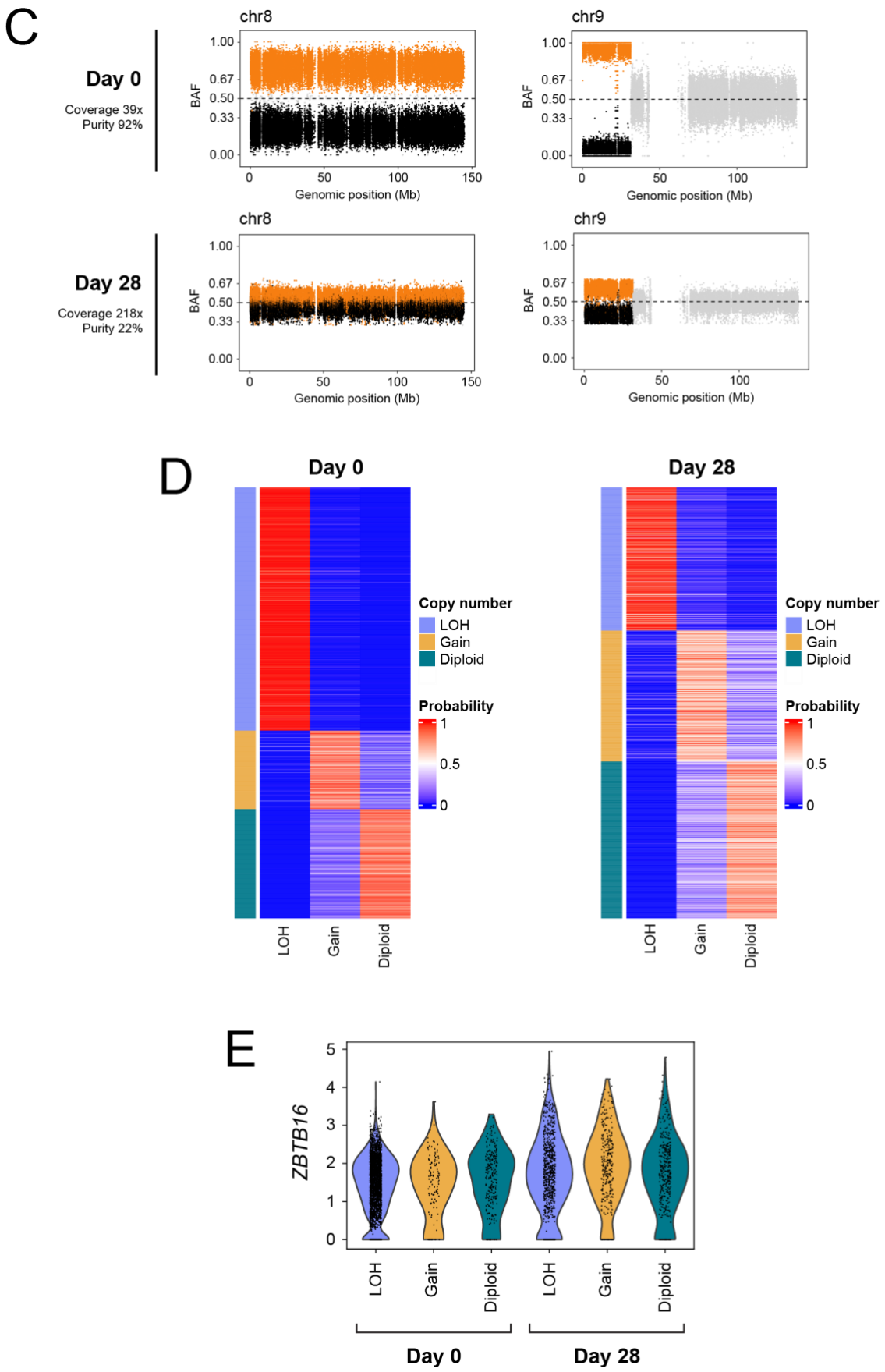


Figure 5.14: Multiple copy number states in P030.

(**A**) P030 day 0 sample WGS: Read depth (top) and BAF (middle) of SNPs, plotted across chromosomes 1 to 22.

(**B**) P030 day 28 sample WGS: Read depth (top) and BAF (middle) of SNPs, plotted across chromosomes 1 to 22.

(**C**) Phased BAF of SNPs in WGS, at chromosome 8 and chromosome 9, of the day 0 and day 28 samples.

(**D**) Heatmap showing the probabilities that each leukaemia blast fits the chromosome 8 LOH, gain or diploid copy number states, using the scRNA-seq data from the day 0 and day 28 samples.

(**E**) Violin plot showing *ZBTB16* expression across blasts from all three copy number states from day 0 and day 28 samples of P030.

BAF, B allele frequency.

Figure 5.15

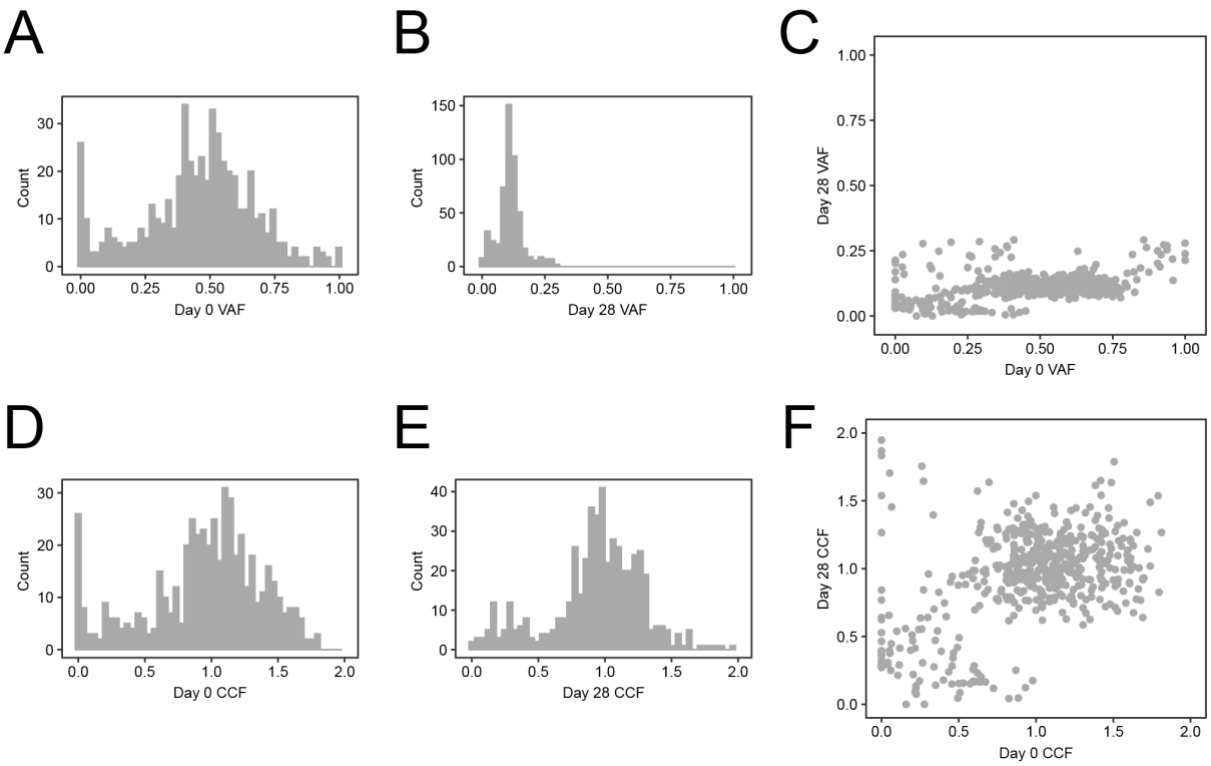


Figure 5.15: Analysis of SNVs in P030.

(**A**) Histogram showing VAF of SNVs at day 0.

(**B**) Histogram showing VAF of SNVs at day 28.

(**C**) Scatterplot showing VAF of SNVs at day 0 (x-axis) and day 28 (y-axis).

(**D**) Histogram showing CCF of SNVs at day 0.

(**E**) Histogram showing CCF of SNVs at day 28.

(**F**) Scatterplot showing CCF of SNVs at day 0 (x-axis) and day 28 (y-axis).

VAF, variant allele frequency; CCF, cancer cell fraction.

5.9 Discussion

In this chapter, I investigated the origin of refractory *ZBTB16*+ blasts in T-ALL. While *ZBTB16*+ blasts were enriched in certain T-ALL genomic subtypes (e.g. LMO2 gd-like and STAG2/LMO2) and in certain TCR rearrangement states (e.g. no TCR and gd-TCR), they could arise from any of the genomic subtypes or TCR rearrangement states. Moreover, within some of the genomic subtypes, *ZBTB16* expression identified the refractory leukaemias from the responsive ones within the same subtype. Thus, *ZBTB16* expression could represent a refractory leukaemia cell state that is independent to its stage of T cell developmental arrest, and independent to its driver genomic lesions.

ZBTB16 expression in T-ALL blasts could be driven by noncoding mutations. However, in my limited size cohort, I did not find a recurrently mutated noncoding hotspot, nor did I find evidence of monoallelic expression of *ZBTB16*. It is possible that interrogation of a much larger genomic cohort, such as the ~1,300 individuals with T-ALL from the COG AALL043 study, may yield potential noncoding mutations related to *ZBTB16* (Pölönen et al., 2024). Currently, the only known mutation involving *ZBTB16* in T-ALL is the fusion between *ZBTB16* and *ABL1* (Chen et al., 2018). It is likely that enhancers on *ZBTB16* drive the expression of *ABL1*, reminiscent of the *TAL1-STIL* fusion, where *TAL1* is driven by *STIL* enhancers (Janssen et al., 1993). Methylation changes may also modulate *ZBTB16* expression. For instance, promoter hypomethylation activates *IRX1* in osteosarcoma and increases its metastatic potential (Lu et al., 2015). Thus, methylation sequencing of T-ALL may provide answers on the origin of *ZBTB16*+ blasts.

Phylogenetic analysis of leukaemia across treatment timepoints sheds insights into the origin of treatment resistant blasts. Refractory blasts may have acquired resistance mechanisms over the course of treatment. Alternatively, refractory blasts may be intrinsically resistant and already present at diagnosis. Phylogenetic analysis of leukaemia in two individuals, P058 and P030, demonstrated that refractory *ZBTB16* blasts were already present at diagnosis.

In the case of P058, while majority of blasts at diagnosis (99%) was a clone that responded to induction chemotherapy, there was a small clone (1%) also present at diagnosis and this clone was selected for by induction treatment. Although both clones shared some driver mutations, including the *TAL1* enhancer mutation, there were no additional mutations in known T-ALL genes in the refractory clone. Instead, the refractory clone was *ZBTB16*+ while the responsive clone was *ZBTB16*−. This supports the finding in my previous chapter that *ZBTB16* is a driver of refractory T-ALL.

In the case of P030, there were at least three leukaemia clones present at both diagnosis and end of induction, defined by their copy number state at chromosome 8. This suggests that no clonal selection occurred. All three clones were *ZBTB16*+, and the separate clustering of blasts between diagnosis and end of induction could not be accounted for by clonal selection or *ZBTB16* expression. This is consistent with the hypothesis that *ZBTB16*+ blasts drive refractory disease in T-ALL.

Key to my phylogenetic analysis was the ability to project somatic mutations onto single cell transcriptome sequencing data. Traditional bulk WGS of tumours at standard sequencing depth (30-100x) may only detect somatic mutations of the truncal clone and major subclones which make up at least 30% of the tumour. However, by projecting somatic mutations called from WGS onto single cell transcriptome data, I could define a clone consisting of around 50 cells and making up 1% of all leukaemia blasts, such as in the diagnostic sample of P058.

The sensitivity of reconstructing phylogeny using single cell transcriptome data is contingent on having sufficient coverage of somatic mutation in single cell transcriptome data. TCR is rearranged and expressed in a subset of T-ALL. The exact set of TCR genes a leukaemia utilises is unique due to somatic recombination of V, D and J gene segments at the TCR loci. Hence, TCR gene usage is a powerful somatic mutation marker for lineage tracing in single cell transcriptomes of T-ALL. Switching of TCR gene usage in T-ALL is very rare, but has been reported in relapse cases (Szczepański et al., 2011). However, to my best knowledge, this has not been seen at the end of induction, where only 28 days have passed, making P058 an extremely unique case.

Copy number alterations can be projected onto single cell transcriptome data by detecting allelic imbalance (Trinh et al., 2022). Due to the large number of germline heterozygous SNPs across the genome (~1 million), their coverage on single cell transcriptome is high, resulting in good sensitivity for resolving small subclones, such as in the case of P058.

SNVs can also be detected on single cell transcriptomes. However, the average T-ALL has around ~1000 clonal SNVs and majority are in noncoding regions of the genome. Thus, their coverage on short-read single cell transcriptome sequencing is low. Long-read single cell transcriptome sequencing may permit greater coverage of SNVs and increase the sensitivity for delineating subclones on single cell data (Monzó et al., 2025).

Overall, this chapter has provided a glimpse into future approaches for cancer genomics. Traditional bulk WGS of tumours generates a catalogue of somatic mutations, which can be used to detect large subclones and reconstruct cancer phylogeny, if multiple samples across time and space are sequenced. However, smaller subclones cannot be detected directly, even if they can be inferred phylogenetically. In contrast, single cell transcriptome sequencing yields gene expression outputs at an individual cell level and can be used to detect a small cluster of cancer cells with a different phenotype from the rest of the tumour. However, the phylogenetic relationship of cells cannot be directly inferred from single cell gene expression data alone. Projecting WGS-derived somatic mutations onto single cell transcriptome sequencing data allows one to directly link genetically defined clones to their gene expression cell states and demonstrate how these cells are phylogenetically related to each other. Therefore, one will be able to make conclusions about how a cancer evolves over time, not only at a genetic level, but also how its cell behaviour changes.

Chapter 6: Conclusion

6.1 Summary of main findings

This dissertation presents transcriptomic and genomic analyses of T-ALL from a cohort of 55 individuals, leveraging on contemporary single cell RNA sequencing (scRNA-seq), single cell T cell receptor sequencing (scTCR-seq) and whole genome sequencing (WGS). With the power of quantitative, high-dimensional, molecular readouts at the resolution of individual cells, I could define precise cell states of both leukaemia cells and non-malignant cells in the tumour microenvironment. Moreover, by using somatic mutations as natural lineage tracing barcodes and detecting them in single leukaemia cells, I could investigate the origin of refractory T-ALL blasts. The main results from each of the chapters were:

(1) T-ALL comprises diverse cell states and cell compositions, in both the leukaemia and non-malignant compartments. Leukaemia cells share transcriptomic resemblance to various stages of T cell development, from early haematopoietic and lymphoid progenitors in the bone marrow to thymocyte stages in the thymus. Sometimes, they may even have features of unconventional T cell lineages or even innate lymphocyte cells (ILC), which are not T cells by definition but share significant overlapping biology. Within the leukaemia microenvironment, I found activated effector T cell states and exhausted T cells, potentially suggesting the presence of adaptive immune responses in T-ALL. Finally, by comparing the bone marrow cell composition at the end of induction treatment, I found that the presence of significant residual blasts (induction failure) was associated with depletion of memory B cells.

(2) Refractory T-ALL is driven by the presence of *ZBTB16*+ blasts. By analysing gene expression of single leukaemia blasts, both at diagnosis and more importantly at the end of induction treatment, which represent true refractory blasts, I found that refractory T-ALL is associated with the presence of *ZBTB16*+ blasts. I validated this finding in an independent scRNA-seq cohort which was generated at a later part of my PhD, as well as independent bulk RNA sequencing cohorts, including the COG AALL0434 cohort of >1,300 patients with T-ALL. Moreover, evidence of *ZBTB16*+ blasts in bulk transcriptomes was associated with poorer overall survival in the COG AALL0434 cohort, and more predictive of survival outcome than other proposed biomarkers. *ZBTB16*+ blasts exhibit a distinct cell state that is unlike canonical stages of T cell development, and instead resemble unconventional T cells and ILCs. Using the scRNA-seq data, I derived several

cell surface targets for immunotherapy, which are uniquely expressed on leukaemia cells and not on normal conventional T cells, thereby avoiding T cell aplasia.

(3) **Refractory *ZBTB16*+ blasts are not associated with particular genomic subtypes or TCR rearrangement states, and can exist as a phylogenetically distinct lineage together with *ZBTB16*− blasts.** *ZBTB16*+ T-ALL can arise from any of the T-ALL genomic subtypes (defined by key driver mutations) and TCR rearrangements states (which represent progression through stages of T cell development). *ZBTB16* expression does not appear to be driven by noncoding mutations in the vicinity of its gene body and there is no evidence of monoallelic expression of *ZBTB16*. Finally, by projecting various somatic mutations, including copy number alterations, single nucleotide variants (SNVs) and TCR rearrangements, onto single leukaemia blast transcriptomes, I demonstrated that in one individual, *ZBTB16*+ blasts at the end of induction could be traced back to a 1% subclone of blasts at diagnosis.

6.2 Risk stratification of T-ALL

Risk stratification is a critical aspect in the treatment of leukaemia. It involves identifying both low-risk patients who will respond with less treatment and high-risk patients who require more aggressive therapy to overcome the disease. Highly toxic chemotherapy is used to kill leukaemia cells, but this often results in acute side effects, and importantly, chronic sequelae including secondary cancers, cognitive impairment, anthracycline-related cardiomyopathy and endocrine dysfunction (Oeffinger et al., 2006). On the other hand, the inability to eliminate leukaemia cells (i.e. induction failure) is a major contributor of disease mortality, with half of induction failure T-ALL patients dying within 5 years (Raetz et al., 2023). In view of these two competing considerations, risk stratification becomes paramount: children with T-ALL should be given the least toxic treatment that is still able to eradicate the disease. However, despite years of research in T-ALL, including genomic and transcriptomic sequencing of large cohorts, there is currently no risk stratification for T-ALL at the point of diagnosis and all patients are treated with the same induction protocol (Teachey and O'Connor, 2020).

It is in this context – the conundrum of T-ALL risk stratification – that my discovery of a refractory *ZBTB16*+ T-ALL phenotype is both unexpected and significant. Most notably, signals of *ZBTB16*+ blasts were more predictive of overall survival than the well-established ETP immunophenotype (Coustan-Smith et al., 2009) or the recently described BMP-like blast (Xu et al., 2024). Defining high-risk and low-risk disease subtypes is the first step towards

optimising the management of children with T-ALL. From this perspective, there is a growing movement today towards treatment de-escalation in children with ALL, considering how the 5-year overall survival rate in ALL has increased dramatically from 10% in the 1960s to 90% today (Hunger et al., 2012). Even for T-ALL, which is considered a high-risk form of ALL, the 5-year overall survival is around 85-90% (Wood et al., 2023). If one is able to identify the responsive *ZBTB16*− T-ALL cases, then de-escalation clinical trials could be conducted on this group of patients.

In this dissertation, I demonstrated the superior sensitivity and specificity of scRNA-seq to detect refractory T-ALL phenotypes in much smaller study cohorts (n < 100), compared to bulk transcriptome and genome sequencing studies with larger cohorts (n > 1000). As the costs of sequencing continue to fall, more research groups are opting to use scRNA-seq to study precious clinical specimens. WGS which began only about 15 years ago, is now routinely used in children with suspected cancer and is able to provide genetic information with diagnostic and prognostic relevance, that supersedes current standard of care genetic testing assays (Hodder et al., 2024). It is not inconceivable that scRNA-seq (or other single cell sequencing modalities) might one day become a routine clinical investigation for cancer diagnosis and risk stratification. However, from my personal experience, analysing data from scRNA-seq requires iterative and subjective processing steps, particular with cell annotation, making it slower and more labour intensive compared to WGS. Moreover, scRNA-seq data quality varies more greatly between samples than for WGS. Part of this stems from the fact that gene expression signals are analogue and on a continuous scale, while genomic mutations are digital signals (a mutation is either present or absent). With advances in sequencing chemistry and data analytics, scRNA-seq may one day evolve into a readily deployable test that is used routinely in the clinic. Currently, however, I believe scRNA-seq will remain as a discovery tool to identify predictive biomarkers that will be assayed via flow cytometry or immunohistochemistry in the clinic.

6.3 Discovery of new therapeutics for T-ALL

It is not sufficient to risk stratify leukaemia unless different treatment options are available for the high-risk subtypes. T-ALL is currently treated with intensive cytotoxic chemotherapy regimens which are associated with adverse side effects (Malard and Mohty, 2020). There is growing interest for targetted therapies such as small-molecule inhibitors, monoclonal antibodies and chimeric antigen receptor (CAR) T cells. This has been fairly successful in the case of B-ALL, with development of monoclonal antibodies (e.g. blinatumomab) (Kantarjian et

al., 2017) and CAR T cell therapies (e.g. tisagenlecleucel) (Maude et al., 2014) against CD19, an antigen that is ubiquitously expressed on the cell surface of B-ALL blasts.

Progress in this area, however, has been slow for T-ALL, partly hindered by the lack of an ideal cell surface antigen target, due to the unique challenges of T-ALL compared to B-ALL. T-ALL targets need to be highly expressed on leukaemia cells, but not on normal T cells, to prevent severe T cell aplasia. T cells play an essential role in both the cellular and humoural arms of adaptive immunity, unlike B cells which are only involved in humoural immunity and their depletion is more tolerable. This is evident in the differences in clinical severity between T cell and B cell primary immunodeficiencies (Shields and Patel, 2017). Another issue with targetting T-ALL antigens that are also expressed on normal T cells is CAR T cell fratricide, where CAR T cells, which both express and target the T cell antigen, kill each other instead of the leukaemia cells, reducing overall treatment efficacy. CD7 has been an attractive CAR T cell target because its near-universal expression on T-ALL blasts. However, it has the problems of both T cell aplasia and CAR T cell fratricide. Strategies have been devised to overcome CAR T cell fratricide in anti-CD7 CAR T cell therapy, including base editing to ablate CD7 expression on CAR T cells (Chiesa et al., 2023) and using a CD7 protein expression blocker which prevents the cell surface expression of CD7 (Oh et al., 2024). However, these methods do not prevent T cell aplasia and anti-CD7 CAR T cell therapy is only used to a bridging therapy to haematopoietic stem cell transplant. My unbiased transcriptome analysis of T-ALL blasts identified potential cell surface targets that are highly expressed on leukaemia cells but absent from normal T cells. Flow cytometry studies are required to demonstrate their expression at the protein level on T-ALL blasts and their absence on normal T cells. Additionally, I found potential targets on refractory *ZBTB16*+ blasts but not normal T cells. As *ZBTB16*+ T-ALL represents a biologically distinct and clinically refractory subtype, identification of cell surface targets on *ZBTB16*+ T-ALL is highly relevant and impactful.

ZBTB16, a lineage-defining transcription factor of unconventional T cells and other innate-like lymphocytes, if subsequently shown to be functional and necessary to maintain the refractory leukaemia cell state, may represent a potential drug target itself. Historically, transcription factors have been considered “undruggable”, due to the lack of deep pockets (such as the active sites of kinases) and their intracellular location (making it difficult for monoclonal antibodies or CAR T cells to reach). However, numerous methods to target transcription factors have proven successful and even led to drugs in clinical trials for various haematological malignancies as well as solid cancers. A classic example is the use of menin inhibitors for acute myeloid leukaemia (AML) and ALL with *KMT2A* (MLL) rearrangements. Following the discovery that interaction between MLL and the transcriptional co-activator

menin is essential in these leukaemias (Yokoyama et al., 2005), small molecule menin inhibitors were designed to block protein-protein interaction between MLL and menin (Borkin et al., 2015; Grembecka et al., 2012). Protein interactors of ZBTB16 may be identified with techniques such as rapid immunoprecipitation mass spectrometry of endogenous protein (RIME) (Mohammed et al., 2016). Targetted protein degradation of transcription factor is an alternative therapeutic strategy. Molecular glue compounds have recently been developed to promote proteasomal degradation of ZBTB16 for the treatment of acute promyelocytic leukaemia driven by the *ZBTB16-RARA* translocation (Matyskiela et al., 2020). These molecules bind to Cereblon E3 ubiquitin ligase and create a surface to interact with ZBTB16, resulting in ubiquitylation and proteasomal degradation of ZBTB16. This is similar in concept to a PROTAC (proteolysis targetting chimera), which is a bifunctional molecule containing a ligand that interacts with an E3 ubiquitin ligase, connected to a second ligand which independently interacts with the target protein (Bondeson et al., 2015; Winter et al., 2015).

6.4 Genotype-to-phenotype inference at the resolution of single cells

Recent years have been marked by a rapidly growing interest to study single cells in complex multicellular tissues, particularly of the human body in homeostasis and disease. This has been driven by two parallel strands of research. On one hand, there is an increasing appreciation of somatic mutations and clonal expansions in non-malignant tissues. For instance, genomic sequencing of individual blood stem cells showed that over our lifetime, these cells accumulate mutations which confer small fitness advantages, resulting in a reduction in blood cell clonal diversity in old age (Mitchell et al., 2022). However, we do not know how these mutations alter gene expression to increase cellular fitness. On the other hand, transcriptome sequencing of single cells has generated gene expression profiles across diverse cell types (Domínguez Conde et al., 2022; Jardine et al., 2021; Park et al., 2020; Suo et al., 2022). Although transitions between cell types can be inferred using pseudotime trajectory methods (Setty et al., 2019; Street et al., 2018; Wolf et al., 2019), which order cells by the similarity of their gene expression profiles, they do not provide definitive evidence of clonal relationships between cells. Thus, there is a growing desire to combine DNA with RNA sequencing to investigate how genomic mutations alter gene expression and behaviour of individual cells.

One method of genotype-to-phenotype analysis is the projection of WGS-derived somatic mutations (such as copy number alterations and SNVs) onto scRNA-seq. This approach, however, has a number of limitations, including dependence on WGS for calling somatic

mutations and limited coverage of the genome by scRNA-seq. Where WGS is not possible due to limited sample availability or cost restrictions, mutations can be directly called on scRNA-seq using SComatic (Muyas et al., 2024), although this is more reliable in cancers with higher mutation rates (to increase the chance that a mutation occurs on transcribed regions) and when there are admixed normal cells (for the removal of germline variants). To increase the sensitivity of detecting mutations in specific transcribed genes, Genotyping of Transcriptomes (GoT) can be used, where primers are added to amplify specific short regions of interest in scRNA-seq libraries (Nam et al., 2019). This may be useful where malignant cells only differ slightly in gene expression profile from normal cells, and malignant cells can only be identified by the presence of specific gene mutations, such as *CALR* mutations in myeloproliferative neoplasms. Furthermore, extending from standard short-read scRNA-seq to long-read sequencing ensures greater coverage of the transcribed genome to permit calling of more classes of mutations including gene fusions (Penter et al., 2024).

When the goal is to reconstruct phylogenetic relationships between cells, rather than detecting the presence or absence of mutations in specific genes, joint sequencing of mitochondrial genome and transcriptome is a powerful approach, as the mitochondrial genome has a higher mutation rate (10-100 vs 1-2 per cell division) and smaller genome size (16.7 kb vs 3.1 Gb), so more phylogenetic markers can be captured for less sequencing coverage (Ludwig et al., 2019; Weng et al., 2024). Ultimately, the “holy grail” for genotype-to-phenotype inference is simultaneous whole genome and whole transcriptome sequencing of single cells. However, error rates of standard sequencing chemistry currently preclude confident variant calling at base-pair resolution, so phylogenies can only be inferred from copy number alterations, which aggregate signals across thousands to millions of sites (Theunis et al., 2025; Wang et al., 2025). Advances in sequencing chemistry, such as duplex sequencing where both strands of DNA are sequenced independently, may enable confident calling of mutations at base-pair resolution on single DNA molecules, including SNVs and insertion/deletion events (Abascal et al., 2021). If high-fidelity genome sequencing can be combined with transcriptome sequencing at single cell level, we can then interrogate within the constant germline genomic background and microenvironment of the same person, how somatic mutations subtly alter gene expression and behaviour of individual cells.

6.5 Closing remarks

This dissertation explored the transcriptome and genomes of T-ALL using scRNA-seq and WGS. With these contemporary sequencing and analytical approaches, I characterised the

cellular composition of T-ALL and elucidated the nature of refractory T-ALL blasts. These findings generate more questions than they answer. Notably, the cell of origin of refractory *ZBTB16*+ blasts and the molecular mechanisms leading to *ZBTB16* expression remain unknown. Solving this modern-day mystery in T cell leukaemia may shed insights into the fundamental processes underlying T cell development, just as Jacques Miller did when he discovered the function of the thymus over 50 years ago.

References

Abascal, F., Harvey, L.M.R., Mitchell, E., Lawson, A.R.J., Lensing, S.V., Ellis, P., Russell, A.J.C., Alcantara, R.E., Baez-Ortega, A., Wang, Y., Kwa, E.J., Lee-Six, H., Cagan, A., Coorens, T.H.H., Chapman, M.S., Olafsson, S., Leonard, S., Jones, D., Machado, H.E., Davies, M., Øbro, N.F., Mahubani, K.T., Allinson, K., Gerstung, M., Saeb-Parsy, K., Kent, D.G., Laurenti, E., Stratton, M.R., Rahbari, R., Campbell, P.J., Osborne, R.J., Martincorena, I., 2021. Somatic mutation landscapes at single-molecule resolution. Nature 593, 405–410. https://doi.org/10.1038/s41586-021-03477-4

Agata, Y., Tamaki, N., Sakamoto, S., Ikawa, T., Masuda, K., Kawamoto, H., Murre, C., 2007. Regulation of T Cell Receptor β Gene Rearrangements and Allelic Exclusion by the Helix-Loop-Helix Protein, E47. Immunity 27, 871–884. https://doi.org/10.1016/j.immuni.2007.11.015

Aifantis, I., Raetz, E., Buonamici, S., 2008. Molecular pathogenesis of T-cell leukaemia and lymphoma. Nat Rev Immunol 8, 380–390. https://doi.org/10.1038/nri2304

Albertí-Servera, L., Demeyer, S., Govaerts, I., Swings, T., De Bie, J., Gielen, O., Brociner, M., Michaux, L., Maertens, J., Uyttebroeck, A., De Keersmaecker, K., Boeckx, N., Segers, H., Cools, J., 2021. Single-cell DNA amplicon sequencing reveals clonal heterogeneity and evolution in T-cell acute lymphoblastic leukemia. Blood 137, 801–811. https://doi.org/10.1182/blood.2020006996

Alexander, T.B., Gu, Z., Iacobucci, I., Dickerson, K., Choi, J.K., Xu, B., Payne-Turner, D., Yoshihara, H., Loh, M.L., Horan, J., Buldini, B., Basso, G., Elitzur, S., De Haas, V., Zwaan, C.M., Yeoh, A., Reinhardt, D., Tomizawa, D., Kiyokawa, N., Lammens, T., De Moerloose, B., Catchpoole, D., Hori, H., Moorman, A., Moore, A.S., Hrusak, O., Meshinchi, S., Orgel, E., Devidas, M., Borowitz, M., Wood, B., Heerema, N.A., Carrol, A., Yang, Y.-L., Smith, M.A., Davidsen, T.M., Hermida, L.C., Gesuwan, P., Marra, M.A., Ma, Y., Mungall, A.J., Moore, R.A., Jones, S.J.M., Valentine, M., Janke, L.J., Rubnitz, J.E., Pui, C.-H., Ding, L., Liu, Y., Zhang, J., Nichols, K.E., Downing, J.R., Cao, X., Shi, L., Pounds, S., Newman, S., Pei, D., Guidry Auvil, J.M., Gerhard, D.S., Hunger, S.P., Inaba, H., Mullighan, C.G., 2018. The genetic basis and cell of origin of mixed phenotype acute leukaemia. Nature 562, 373–379. https://doi.org/10.1038/s41586-018-0436-0

Alonzo, E.S., Sant'Angelo, D.B., 2011. Development of PLZF-expressing innate T cells. Current Opinion in Immunology 23, 220–227. https://doi.org/10.1016/j.coi.2010.12.016

Aplan, P.D., Jones, C.A., Chervinsky, D.S., Zhao, X., Ellsworth, M., Wu, C., McGuire, E.A., Gross, K.W., 1997. An scl gene product lacking the transactivation domain induces bony abnormalities and cooperates with LMO1 to generate T-cell malignancies in transgenic mice. EMBO J 16, 2408–2419. https://doi.org/10.1093/emboj/16.9.2408

Arber, D.A., Orazi, A., Hasserjian, R., Thiele, J., Borowitz, M.J., Le Beau, M.M., Bloomfield, C.D., Cazzola, M., Vardiman, J.W., 2016. The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia. Blood 127, 2391–2405. https://doi.org/10.1182/blood-2016-03-643544

Artis, D., Spits, H., 2015. The biology of innate lymphoid cells. Nature 517, 293–301. https://doi.org/10.1038/nature14189

Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., Harris, M.A., Hill, D.P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J.C., Richardson, J.E., Ringwald, M., Rubin, G.M., Sherlock, G., 2000. Gene Ontology: tool for the unification of biology. Nat Genet 25, 25–29. https://doi.org/10.1038/75556

Ashby, K.M., Hogquist, K.A., 2024. A guide to thymic selection of T cells. Nat Rev Immunol 24, 103–117. https://doi.org/10.1038/s41577-023-00911-8

Asnafi, V., Beldjord, K., Boulanger, E., Landman-Parker, J., Quartier, P., 2003. Analysis of TCR, pTa, and RAG-1 in T-acute lymphoblastic leukemias improves understanding of early human T-lymphoid lineage commitment 101.

Bassing, C.H., Swat, W., Alt, F.W., 2002. The Mechanism and Regulation of Chromosomal V(D)J Recombination. Cell 109, S45–S55. https://doi.org/10.1016/S0092-8674(02)00675-X

Begley, C.G., Aplan, P.D., Davey, M.P., Nakahara, K., Tchorz, K., Kurtzberg, J., Hershfield, M.S., Haynes, B.F., Cohen, D.I., Waldmann, T.A., 1989. Chromosomal translocation in a human leukemic stem-cell line disrupts the T-cell antigen receptor delta-chain diversity region and results in a previously unreported fusion transcript. Proc Natl Acad Sci U S A 86, 2031–2035. https://doi.org/10.1073/pnas.86.6.2031

Behjati, S., Lindsay, S., Teichmann, S.A., Haniffa, M., 2018. Mapping human development at single-cell resolution. Development 145, dev152561. https://doi.org/10.1242/dev.152561

Bene, M.C., Castoldi, G., Knapp, W., Ludwig, W.D., Matutes, E., Orfao, A., van't Veer, M.B., 1995. Proposals for the immunological classification of acute leukemias. European Group for the Immunological Characterization of Leukemias (EGIL). Leukemia 9, 1783–1786.

Bernard, O., Busson-LeConiat, M., Ballerini, P., Mauchauffé, M., Della Valle, V., Monni, R., Khac, F.N., Mercher, T., Penard-Lacronique, V., Pasturaud, P., Gressin, L., Heilig, R., Daniel, M.-T., Lessard, M., Berger, R., 2001. A new recurrent and specific cryptic translocation, t(5;14)(q35;q32), is associated with expression of the Hox11L2 gene in T acute lymphoblastic leukemia. Leukemia 15, 1495–1504. https://doi.org/10.1038/sj.leu.2402249

Bernink, J.H., Peters, C.P., Munneke, M., Te Velde, A.A., Meijer, S.L., Weijer, K., Hreggvidsdottir, H.S., Heinsbroek, S.E., Legrand, N., Buskens, C.J., Bemelman, W.A., Mjösberg, J.M., Spits, H., 2013. Human type 1 innate lymphoid cells accumulate in inflamed mucosal tissues. Nat Immunol 14, 221–229. https://doi.org/10.1038/ni.2534

Bondeson, D.P., Mares, A., Smith, I.E.D., Ko, E., Campos, S., Miah, A.H., Mulholland, K.E., Routly, N., Buckley, D.L., Gustafson, J.L., Zinn, N., Grandi, P., Shimamura, S., Bergamini, G., Faelth-Savitski, M., Bantscheff, M., Cox, C., Gordon, D.A., Willard, R.R., Flanagan, J.J., Casillas, L.N., Votta, B.J., den Besten, W., Famm, K., Kruidenier, L., Carter, P.S., Harling, J.D., Churcher, I., Crews, C.M., 2015. Catalytic in vivo protein knockdown by small-molecule PROTACs. Nature Chemical Biology 11, 611–617. https://doi.org/10.1038/nchembio.1858

Boos, M.D., Yokota, Y., Eberl, G., Kee, B.L., 2007. Mature natural killer cell and lymphoid tissue–inducing cell development requires Id2-mediated suppression of E protein activity. The Journal of Experimental Medicine 204, 1119–1130. https://doi.org/10.1084/jem.20061959

Borkin, D., He, S., Miao, H., Kempinska, K., Pollock, J., Chase, J., Purohit, T., Malik, B., Zhao, T., Wang, J., Wen, B., Zong, H., Jones, M., Danet-Desnoyers, G., Guzman, M.L., Talpaz, M., Bixby, D.L., Sun, D., Hess, J.L., Muntean, A.G., Maillard, I., Cierpicki, T., Grembecka, J., 2015. Pharmacologic Inhibition of the Menin-MLL Interaction Blocks Progression of MLL Leukemia In Vivo. Cancer Cell 27, 589–602. https://doi.org/10.1016/j.ccell.2015.02.016

Brady, S.W., Roberts, K.G., Gu, Z., Shi, L., Pounds, S., Pei, D., Cheng, C., Dai, Y., Devidas, M., Qu, C., Hill, A.N., Payne-Turner, D., Ma, X., Iacobucci, I., Baviskar, P., Wei, L., Arunachalam, S., Hagiwara, K., Liu, Yanling, Flasch, D.A., Liu, Yu, Parker, M., Chen, X., Elsayed, A.H., Pathak, O., Li, Y., Fan, Y., Michael, J.R., Rusch, M., Wilkinson, M.R., Foy, S., Hedges, D.J., Newman, S., Zhou, X., Wang, J., Reilly, C., Sioson, E., Rice, S.V., Pastor Loyola, V., Wu, G., Rampersaud, E., Reshmi, S.C., Gastier-Foster, J., Guidry Auvil, J.M., Gesuwan, P., Smith, M.A., Winick, N., Carroll, A.J., Heerema, N.A., Harvey, R.C., Willman, C.L., Larsen, E., Raetz, E.A., Borowitz, M.J., Wood, B.L., Carroll, W.L., Zweidler-McKay, P.A., Rabin, K.R., Mattano, L.A., Maloney, K.W., Winter, S.S., Burke, M.J., Salzer, W., Dunsmore, K.P., Angiolillo, A.L., Crews, K.R., Downing, J.R., Jeha, S., Pui, C.-H., Evans, W.E., Yang, J.J., Relling, M.V., Gerhard, D.S., Loh, M.L., Hunger, S.P., Zhang, J., Mullighan, C.G., 2022. The genomic landscape of pediatric acute lymphoblastic leukemia. Nat Genet 54, 1376–1389. https://doi.org/10.1038/s41588-022-01159-z

Cameron, D.L., Baber, J., Shale, C., Valle-Inclan, J.E., Besselink, N., Van Hoeck, A., Janssen, R., Cuppen, E., Priestley, P., Papenfuss, A.T., 2021. GRIDSS2: comprehensive characterisation of somatic structural variation using single breakend variants and structural variant phasing. Genome Biol 22, 202. https://doi.org/10.1186/s13059-021-02423-x

Carroll, A.J., Crist, W.M., Link, M.P., Amylon, M.D., Pullen, D.J., Ragab, A.H., Buchanan, G.R., Wimmer, R.S., Vietti, T.J., 1990. The t(l;14)(p34;q11) Is Nonrandom and Restricted to T-Cell Acute Lymphoblastic Leukemia: A Pediatric Oncology Group Study. Blood 76, 1220–1224. https://doi.org/10.1182/blood.V76.6.1220.1220

Cayuela, J.-M., Madani, A., Sanhes, L., Stern, M.-H., Sigaux, F., 1996. Multiple Tumor-Suppressor Gene 1 Inactivation Is the Most Frequent Genetic Alteration in T-Cell Acute Lymphoblastic Leukemia. Blood 87, 2180–2186. https://doi.org/10.1182/blood.V87.6.2180.bloodjournal8762180

Cella, M., Fuchs, A., Vermi, W., Facchetti, F., Otero, K., Lennerz, J.K.M., Doherty, J.M., Mills, J.C., Colonna, M., 2009. A human natural killer cell subset provides an innate source of IL-22 for mucosal immunity. Nature 457, 722–725. https://doi.org/10.1038/nature07537

Chan, A.C., Leeansyah, E., Cochrane, A., D' Udekem D' Acoz, Y., Mittag, D., Harrison, L.C., Godfrey, D.I., Berzins, S.P., 2013. *Ex-vivo* analysis of human Natural Killer T cells demonstrates heterogeneity between tissues and within established CD4+ and CD4− subsets. Clinical and Experimental Immunology 172, 129–137. https://doi.org/10.1111/cei.12045

Chen, B., Jiang, L., Zhong, M.-L., Li, J.-F., Li, B.-S., Peng, L.-J., Dai, Y.-T., Cui, B.-W., Yan, T.-Q., Zhang, W.-N., Weng, X.-Q., Xie, Y.-Y., Lu, J., Ren, R.-B., Chen, S.-N., Hu, J.-D., Wu, D.-P., Chen, Z., Tang, J.-Y., Huang, J.-Y., Mi, J.-Q., Chen, S.-J., 2018. Identification of fusion genes and characterization of transcriptome features in T-cell acute lymphoblastic leukemia. Proc. Natl. Acad. Sci. U.S.A. 115, 373–378. https://doi.org/10.1073/pnas.1717125115

Chen, Changya, Yu, W., Alikarami, F., Qiu, Q., Chen, Chia-hui, Flournoy, J., Gao, P., Uzun, Y., Fang, L., Davenport, J.W., Hu, Y., Zhu, Q., Wang, K., Libbrecht, C., Felmeister, A., Rozich, I., Ding, Y., Hunger, S.P., Felix, C.A., Wu, H., Brown, P.A., Guest, E.M., Barrett, D.M., Bernt, K.M., Tan, K., 2022. Single-cell multiomics reveals increased plasticity, resistant populations, and stem-cell–like blasts in *KMT2A* -rearranged leukemia. Blood 139, 2198–2211. https://doi.org/10.1182/blood.2021013442

Chen, Z., Brand, N.J., Chen, A., Chen, S.J., Tong, J.H., Wang, Z.Y., Waxman, S., Zelent, A., 1993. Fusion between a novel Krüppel-like zinc finger gene and the retinoic acid receptor-alpha locus due to a variant t(11;17) translocation associated with acute promyelocytic leukaemia. The EMBO Journal 12, 1161–1167. https://doi.org/10.1002/j.1460-2075.1993.tb05757.x

Cheroutre, H., Lambolez, F., Mucida, D., 2011. The light and dark sides of intestinal intraepithelial lymphocytes. Nat Rev Immunol 11, 445–456. https://doi.org/10.1038/nri3007

Chervinsky, D.S., Sait, S.N., Nowak, N.J., Shows, T.B., Aplan, P.D., 1995. Complex MLL rearrangement in a patient with T-cell acute lymphoblastic leukemia. Genes Chromosomes Cancer 14, 76–84. https://doi.org/10.1002/gcc.2870140114

Chien, Y., Meyer, C., Bonneville, M., 2014. γδ T Cells: First Line of Defense and Beyond. Annu. Rev. Immunol. 32, 121–155. https://doi.org/10.1146/annurev-immunol-032713-120216

Chiesa, R., Georgiadis, C., Syed, F., Zhan, H., Etuk, A., Gkazi, S.A., Preece, R., Ottaviano, G., Braybrook, T., Chu, J., Kubat, A., Adams, S., Thomas, R., Gilmour, K., O'Connor, D., Vora, A., Qasim, W., 2023. Base-Edited CAR7 T Cells for Relapsed T-Cell Acute Lymphoblastic Leukemia. N Engl J Med 389, 899–910. https://doi.org/10.1056/NEJMoa2300709

Coles, M.C., Raulet, D.H., 2000. NK1.1+ T Cells in the Liver Arise in the Thymus and Are Selected by Interactions with Class I Molecules on CD4+CD8+ Cells. The Journal of Immunology 164, 2412–2418. https://doi.org/10.4049/jimmunol.164.5.2412

Constantinides, M.G., McDonald, B.D., Verhoef, P.A., Bendelac, A., 2014. A committed precursor to innate lymphoid cells. Nature 508, 397–401. https://doi.org/10.1038/nature13047

Conter, V., Valsecchi, M.G., Buldini, B., Parasole, R., Locatelli, F., Colombini, A., Rizzari, C., Putti, M.C., Barisone, E., Nigro, L.L., Santoro, N., Ziino, O., Pession, A., Testi, A.M., Micalizzi, C., Casale, F., Pierani, P., Cesaro, S., Cellini, M., Silvestri, D., Cazzaniga, G., Biondi, A., Basso, G., 2016. Early T-cell precursor acute lymphoblastic leukaemia in children treated in AIEOP centres with AIEOP-BFM protocols: a retrospective analysis. The Lancet Haematology 3, e80–e86. https://doi.org/10.1016/S2352-3026(15)00254-9

Coorens, T.H.H., Oliver, T.R.W., Sanghvi, R., Sovio, U., Cook, E., Vento-Tormo, R., Haniffa, M., Young, M.D., Rahbari, R., Sebire, N., Campbell, P.J., Charnock-Jones, D.S., Smith, G.C.S., Behjati, S., 2021. Inherent mosaicism and extensive mutation

of human placentas. Nature 592, 80–85. https://doi.org/10.1038/s41586-021-03345-1

Coustan-Smith, E., Mullighan, C.G., Onciu, M., Behm, F.G., Raimondi, S.C., Pei, D., Cheng, C., Su, X., Rubnitz, J.E., Basso, G., Biondi, A., Pui, C.-H., Downing, J.R., Campana, D., 2009. Early T-cell precursor leukaemia: a subtype of very high-risk acute lymphoblastic leukaemia. The Lancet Oncology 10, 147–156. https://doi.org/10.1016/S1470-2045(08)70314-0

Dadi, S., Le Noir, S., Payet-Bornet, D., Lhermitte, L., Zacarias-Cabeza, J., Bergeron, J., Villarèse, P., Vachez, E., Dik, W.A., Millien, C., Radford, I., Verhoeyen, E., Cosset, F.-L., Petit, A., Ifrah, N., Dombret, H., Hermine, O., Spicuglia, S., Langerak, A.W., Macintyre, E.A., Nadel, B., Ferrier, P., Asnafi, V., 2012. TLX Homeodomain Oncogenes Mediate T Cell Maturation Arrest in T-ALL via Interaction with ETS1 and Suppression of TCRα Gene Expression. Cancer Cell 21, 563–576. https://doi.org/10.1016/j.ccr.2012.02.013

Danecek, P., Bonfield, J.K., Liddle, J., Marshall, J., Ohan, V., Pollard, M.O., Whitwham, A., Keane, T., McCarthy, S.A., Davies, R.M., Li, H., 2021. Twelve years of SAMtools and BCFtools. GigaScience 10, giab008. https://doi.org/10.1093/gigascience/giab008

De Keersmaecker, K., Atak, Z.K., Li, N., Vicente, C., Patchett, S., Girardi, T., Gianfelici, V., Geerdens, E., Clappier, E., Porcu, M., Lahortiga, I., Lucà, R., Yan, J., Hulselmans, G., Vranckx, H., Vandepoel, R., Sweron, B., Jacobs, K., Mentens, N., Wlodarska, I., Cauwelier, B., Cloos, J., Soulier, J., Uyttebroeck, A., Bagni, C., Hassan, B.A., Vandenberghe, P., Johnson, A.W., Aerts, S., Cools, J., 2013. Exome sequencing identifies mutation in CNOT3 and ribosomal genes RPL5 and RPL10 in T-cell acute lymphoblastic leukemia. Nat Genet 45, 186–190. https://doi.org/10.1038/ng.2508

De Lalla, C., Lepore, M., Piccolo, F.M., Rinaldi, A., Scelfo, A., Garavaglia, C., Mori, L., De Libero, G., Dellabona, P., Casorati, G., 2011. High-frequency and adaptive-like dynamics of human CD1 self-reactive T cells. Eur J Immunol 41, 602–610. https://doi.org/10.1002/eji.201041211

Dentro, S.C., Wedge, D.C., Van Loo, P., 2017. Principles of Reconstructing the Subclonal Architecture of Cancers. Cold Spring Harb Perspect Med 7, a026625. https://doi.org/10.1101/cshperspect.a026625

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., Gingeras, T.R., 2013. STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29, 15–21. https://doi.org/10.1093/bioinformatics/bts635

Domínguez Conde, C., Xu, C., Jarvis, L.B., Rainbow, D.B., Wells, S.B., Gomes, T., Howlett, S.K., Suchanek, O., Polanski, K., King, H.W., Mamanova, L., Huang, N., Szabo, P.A., Richardson, L., Bolt, L., Fasouli, E.S., Mahbubani, K.T., Prete, M., Tuck, L., Richoz, N., Tuong, Z.K., Campos, L., Mousa, H.S., Needham, E.J., Pritchard, S., Li, T., Elmentaite, R., Park, J., Rahmani, E., Chen, D., Menon, D.K., Bayraktar, O.A., James, L.K., Meyer, K.B., Yosef, N., Clatworthy, M.R., Sims, P.A., Farber, D.L., Saeb-Parsy, K., Jones, J.L., Teichmann, S.A., 2022. Cross-tissue immune cell analysis reveals tissue-specific features in humans. Science 376, eabl5197. https://doi.org/10.1126/science.abl5197

Doulatov, S., Notta, F., Rice, K.L., Howell, L., Zelent, A., Licht, J.D., Dick, J.E., 2009. PLZF is a regulator of homeostatic and cytokine-induced myeloid development. Genes Dev. 23, 2076–2087. https://doi.org/10.1101/gad.1788109

Dusseaux, M., Martin, E., Serriari, N., Péguillet, I., Premel, V., Louis, D., Milder, M., Le Bourhis, L., Soudais, C., Treiner, E., Lantz, O., 2011. Human MAIT cells are xenobiotic-resistant, tissue-targeted, CD161hi IL-17–secreting T cells. Blood 117, 1250–1259. https://doi.org/10.1182/blood-2010-08-303339

Eberl, G., Colonna, M., Di Santo, J.P., McKenzie, A.N.J., 2015. Innate lymphoid cells: A new paradigm in immunology. Science 348, aaa6566. https://doi.org/10.1126/science.aaa6566

Eberl, G., Marmon, S., Sunshine, M.-J., Rennert, P.D., Choi, Y., Littman, D.R., 2004. An essential function for the nuclear receptor RORγt in the generation of fetal lymphoid tissue inducer cells. Nat Immunol 5, 64–73. https://doi.org/10.1038/ni1022

Ellisen, L.W., Bird, J., West, D.C., Soreng, A.L., Reynolds, T.C., Smith, S.D., Sklar, J., 1991. TAN-1, the human homolog of the Drosophila Notch gene, is broken by chromosomal translocations in T lymphoblastic neoplasms. Cell 66, 649–661. https://doi.org/10.1016/0092-8674(91)90111-B

Fergusson, J.R., Fleming, V.M., Klenerman, P., 2011. CD161-Expressing Human T Cells. Front. Immun. 2. https://doi.org/10.3389/fimmu.2011.00036

Ferrando, A.A., Neuberg, D.S., Staunton, J., Loh, M.L., Huard, C., Raimondi, S.C., Behm, F.G., Pui, C.-H., Downing, J.R., Gilliland, D.G., Lander, E.S., Golub, T.R., Look, A.T., 2002. Gene expression signatures define novel oncogenic pathways in T cell acute lymphoblastic leukemia. Cancer Cell 1, 75–87. https://doi.org/10.1016/S1535-6108(02)00018-1

Ferreira, A.C.F., Szeto, A.C.H., Heycock, M.W.D., Clark, P.A., Walker, J.A., Crisp, A., Barlow, J.L., Kitching, S., Lim, A., Gogoi, M., Berks, R., Daly, M., Jolin, H.E., McKenzie, A.N.J., 2021. RORα is a critical checkpoint for T cell and ILC2 commitment in the embryonic thymus. Nat Immunol 22, 166–178. https://doi.org/10.1038/s41590-020-00833-w

Foroutan, M., Bhuva, D.D., Lyu, R., Horan, K., Cursons, J., Davis, M.J., 2018. Single sample scoring of molecular phenotypes. BMC Bioinformatics 19, 404. https://doi.org/10.1186/s12859-018-2435-4

Fuchs, A., Vermi, W., Lee, J.S., Lonardi, S., Gilfillan, S., Newberry, R.D., Cella, M., Colonna, M., 2013. Intraepithelial Type 1 Innate Lymphoid Cells Are a Unique Subset of IL-12- and IL-15-Responsive IFN-γ-Producing Cells. Immunity 38, 769–781. https://doi.org/10.1016/j.immuni.2013.02.010

Fugmann, S.D., Lee, A.I., Shockett, P.E., Villey, I.J., Schatz, D.G., 2000. The RAG Proteins and V(D)J Recombination: Complexes, Ends, and Transposition. Annu. Rev. Immunol. 18, 495–527. https://doi.org/10.1146/annurev.immunol.18.1.495

Gachet, S., El-Chaar, T., Avran, D., Genesca, E., Catez, F., Quentin, S., Delord, M., Thérizols, G., Briot, D., Meunier, G., Hernandez, L., Pla, M., Smits, W.K., Buijs-Gladdines, J.G., Van Loocke, W., Menschaert, G., André-Schmutz, I., Taghon, T., Van Vlierberghe, P., Meijerink, J.P., Baruchel, A., Dombret, H., Clappier, E., Diaz, J.-J., Gazin, C., De Thé, H., Sigaux, F., Soulier, J., 2018. Deletion 6q Drives T-cell Leukemia Progression by Ribosome Modulation. Cancer Discovery 8, 1614–1631. https://doi.org/10.1158/2159-8290.CD-17-0831

Gapin, L., Matsuda, J.L., Surh, C.D., Kronenberg, M., 2001. NKT cells derive from double-positive thymocytes that are positively selected by CD1d. Nat Immunol 2, 971–978. https://doi.org/10.1038/ni710

García-Ojeda, M.E., Klein Wolterink, R.G.J., Lemaître, F., Richard-Le Goff, O., Hasan, M., Hendriks, R.W., Cumano, A., Di Santo, J.P., 2013. GATA-3 promotes T-cell specification by repressing B-cell potential in pro–T cells in mice. Blood 121, 1749–1759. https://doi.org/10.1182/blood-2012-06-440065

Germar, K., Dose, M., Konstantinou, T., Zhang, J., Wang, H., Lobry, C., Arnett, K.L., Blacklow, S.C., Aifantis, I., Aster, J.C., Gounari, F., 2011. T-cell factor 1 is a gatekeeper for T-cell specification in response to Notch signaling. Proc. Natl. Acad. Sci. U.S.A. 108, 20060–20065. https://doi.org/10.1073/pnas.1110230108

Gerstung, M., Papaemmanuil, E., Campbell, P.J., 2014. Subclonal variant calling with multiple samples and prior knowledge. Bioinformatics 30, 1198–1204. https://doi.org/10.1093/bioinformatics/btt750

Gianni, F., Belver, L., Ferrando, A., 2020. The Genetics and Mechanisms of T-Cell Acute Lymphoblastic Leukemia. Cold Spring Harb Perspect Med 10, a035246. https://doi.org/10.1101/cshperspect.a035246

Ginhoux, F., Yalin, A., Dutertre, C.A., Amit, I., 2022. Single-cell immunology: Past, present, and future. Immunity 55, 393–404. https://doi.org/10.1016/j.immuni.2022.02.006

Godfrey, D.I., MacDonald, H.R., Kronenberg, M., Smyth, M.J., Kaer, L.V., 2004. NKT cells: what's in a name? Nat Rev Immunol 4, 231–237. https://doi.org/10.1038/nri1309

Godfrey, D.I., Uldrich, A.P., McCluskey, J., Rossjohn, J., Moody, D.B., 2015. The burgeoning family of unconventional T cells. Nat Immunol 16, 1114–1123. https://doi.org/10.1038/ni.3298

Goldberg, J.M., Silverman, L.B., Levy, D.E., Dalton, V.K., Gelber, R.D., Lehmann, L., Cohen, H.J., Sallan, S.E., Asselin, B.L., 2003. Childhood T-Cell Acute Lymphoblastic Leukemia: The Dana-Farber Cancer Institute Acute Lymphoblastic Leukemia Consortium Experience. JCO 21, 3616–3622. https://doi.org/10.1200/JCO.2003.10.116

Gordon, S.M., Chaix, J., Rupp, L.J., Wu, J., Madera, S., Sun, J.C., Lindsten, T., Reiner, S.L., 2012. The Transcription Factors T-bet and Eomes Control Key Checkpoints of Natural Killer Cell Maturation. Immunity 36, 55–67. https://doi.org/10.1016/j.immuni.2011.11.016

Grabher, C., Von Boehmer, H., Look, A.T., 2006. Notch 1 activation in the molecular pathogenesis of T-cell acute lymphoblastic leukaemia. Nat Rev Cancer 6, 347–359. https://doi.org/10.1038/nrc1880

Grembecka, J., He, S., Shi, A., Purohit, T., Muntean, A.G., Sorenson, R.J., Showalter, H.D., Murai, M.J., Belcher, A.M., Hartley, T., Hess, J.L., Cierpicki, T., 2012. Menin-MLL inhibitors reverse oncogenic activity of MLL fusion proteins in leukemia. Nature Chemical Biology 8, 277–284. https://doi.org/10.1038/nchembio.773

Gross, L., 1951. Pathogenic Properties, and "Vertical" Transmission of the Mouse Leukemia Agent. Proceedings of the Society for Experimental Biology and Medicine 78, 342–348. https://doi.org/10.3181/00379727-78-19068

Halim, T.Y.F., MacLaren, A., Romanish, M.T., Gold, M.J., McNagny, K.M., Takei, F., 2012. Retinoic-Acid-Receptor-Related Orphan Nuclear Receptor Alpha Is Required for

Natural Helper Cell Development and Allergic Inflammation. Immunity 37, 463–474. https://doi.org/10.1016/j.immuni.2012.06.012

Hammond, K., Cain, W., van Driel, I., Godfrey, D., 1998. Three day neonatal thymectomy selectively depletes NK1.1+ T cells. Int Immunol 10, 1491–1499. https://doi.org/10.1093/intimm/10.10.1491

Haniffa, M., Taylor, D., Linnarsson, S., Aronow, B.J., Bader, G.D., Barker, R.A., Camara, P.G., Camp, J.G., Chédotal, A., Copp, A., Etchevers, H.C., Giacobini, P., Göttgens, B., Guo, G., Hupalowska, A., James, K.R., Kirby, E., Kriegstein, A., Lundeberg, J., Marioni, J.C., Meyer, K.B., Niakan, K.K., Nilsson, M., Olabi, B., Pe'er, D., Regev, A., Rood, J., Rozenblatt-Rosen, O., Satija, R., Teichmann, S.A., Treutlein, B., Vento-Tormo, R., Webb, S., Human Cell Atlas Developmental Biological Network, Barbry, P., Bayraktar, O., Behjati, S., Bosio, A., Canque, B., Chalmel, F., Gitton, Y., Henderson, D., Jorgensen, A., Lisgo, S., Liu, J., Lundberg, E., Maitre, J.-L., Mazaud-Guittot, S., Robertson, E., Rolland, A., Scharfmann, R., Souyri, M., Sundström, E., Zaffran, S., Zilbauer, M., 2021. A roadmap for the Human Developmental Cell Atlas. Nature 597, 196–205. https://doi.org/10.1038/s41586-021-03620-1

Harrison, C.J., Foroni, L., 2002. Cytogenetics and molecular genetics of acute lymphoblastic leukemia. Reviews in Clinical and Experimental Hematology 6, 91–113. https://doi.org/10.1046/j.1468-0734.2002.00069.x

Hatano, M., Roberts, C.W.M., Minden, M., Crist, W.M., Korsmeyer, S.J., 1991. Deregulation of a Homeobox Gene, HOX11, by the t(10;14) in T Cell Leukemia. Science 253, 79–82. https://doi.org/10.1126/science.1676542

Hayday, A.C., Pennington, D.J., 2007. Key factors in the organized chaos of early T cell development. Nat Immunol 8, 137–144. https://doi.org/10.1038/ni1436

He, J., Wu, M., Xiong, L., Gong, Y., Yu, R., Peng, W., Li, Lili, Li, Li, Tian, S., Wang, Y., Tao, Q., Xiang, T., 2020. BTB/POZ zinc finger protein ZBTB16 inhibits breast cancer proliferation and metastasis through upregulating ZBTB28 and antagonizing BCL6/ZBTB27. Clinical Epigenetics 12, 82. https://doi.org/10.1186/s13148-020-00867-9

Hebert, J., Cayuela, J.M., Berkeley, J., Sigaux, F., 1994. Candidate Tumor-Suppressor Genes MTS1 (p16INK4A) and MTS2 (p15INK4B) Display Frequent Homozygous Deletions in Primary Cells From T-But Not From B-Cell Lineage Acute Lymphoblastic Leukemias. Blood 84, 4038–4044. https://doi.org/10.1182/blood.V84.12.4038.bloodjournal84124038

Herranz, D., Ambesi-Impiombato, A., Palomero, T., Schnell, S.A., Belver, L., Wendorff, A.A., Xu, L., Castillo-Martin, M., Llobet-Navás, D., Cordon-Cardo, C., Clappier, E., Soulier, J., Ferrando, A.A., 2014. A NOTCH1-driven MYC enhancer promotes T cell development, transformation and acute lymphoblastic leukemia. Nat Med 20, 1130–1137. https://doi.org/10.1038/nm.3665

Hodder, A., Leiter, S.M., Kennedy, J., Addy, D., Ahmed, M., Ajithkumar, T., Allinson, K., Ancliff, P., Bailey, S., Barnard, G., Burke, G.A.A., Burns, C., Cano-Flanagan, J., Chalker, J., Coleman, N., Cheng, D., Clinch, Y., Dryden, C., Ghorashian, S., Griffin, B., Horan, G., Hubank, M., May, P., McDerra, J., Nagrecha, R., Nicholson, J., O'Connor, D., Pavasovic, V., Quaegebeur, A., Rao, A., Roberts, T., Samarasinghe, S., Stasevich, I., Tadross, J.A., Trayers, C., Trotman, J., Vora, A., Watkins, J., Chitty, L.S., Bowdin, S., Armstrong, R., Murray, M.J., Hook, C.E., Tarpey, P., Vedi, A., Bartram, J., Behjati, S., 2024. Benefits for children with suspected cancer from

routine whole-genome sequencing. Nat Med 30, 1905–1912. https://doi.org/10.1038/s41591-024-03056-w

Hosokawa, H., Romero-Wolf, M., Yui, M.A., Ungerbäck, J., Quiloan, M.L.G., Matsumoto, M., Nakayama, K.I., Tanaka, T., Rothenberg, E.V., 2018. Bcl11b sets pro-T cell fate by site-specific cofactor recruitment and by repressing Id2 and Zbtb16. Nat Immunol 19, 1427–1440. https://doi.org/10.1038/s41590-018-0238-4

Hosokawa, H., Rothenberg, E.V., 2021. How transcription factors drive choice of the T cell fate. Nat Rev Immunol 21, 162–176. https://doi.org/10.1038/s41577-020-00426-6

Hoyler, T., Klose, C.S.N., Souabni, A., Turqueti-Neves, A., Pfeifer, D., Rawlins, E.L., Voehringer, D., Busslinger, M., Diefenbach, A., 2012. The Transcription Factor GATA-3 Controls Cell Fate and Maintenance of Type 2 Innate Lymphoid Cells. Immunity 37, 634–648. https://doi.org/10.1016/j.immuni.2012.06.020

Hozumi, K., Mailhos, C., Negishi, N., Hirano, K., Yahata, T., Ando, K., Zuklys, S., Holländer, G.A., Shima, D.T., Habu, S., 2008. Delta-like 4 is indispensable in thymic environment specific for T cell development. The Journal of Experimental Medicine 205, 2507–2513. https://doi.org/10.1084/jem.20080134

Hu, S., Qian, M., Zhang, H., Guo, Y., Yang, J., Zhao, X., He, H., Lu, J., Pan, J., Chang, M., Du, G., Lin, T.-N., Kham, S.K.-Y., Quah, T.C., Ariffin, H., Tan, A.-M., Cheng, Y., Li, C., Yeoh, A.E.-J., Pui, C.-H., Skanderup, A.J., Yang, J.J., 2017. Whole-genome noncoding sequence analysis in T-cell acute lymphoblastic leukemia identifies oncogene enhancer mutations. Blood 129, 3264–3268. https://doi.org/10.1182/blood-2017-03-771162

Huang, X., Li, Y., Zhang, J., Yan, L., Zhao, H., Ding, L., Bhatara, S., Yang, X., Yoshimura, S., Yang, W., Karol, S.E., Inaba, H., Mullighan, C., Litzow, M., Zhu, X., Zhang, Y., Stock, W., Jain, N., Jabbour, E., Kornblau, S.M., Konopleva, M., Pui, C.-H., Paietta, E., Evans, W., Yu, J., Yang, J.J., 2024. Single-cell systems pharmacology identifies development-driven drug response and combination therapy in B cell acute lymphoblastic leukemia. Cancer Cell 42, 552-567.e6. https://doi.org/10.1016/j.ccell.2024.03.003

Hunger, S.P., Lu, X., Devidas, M., Camitta, B.M., Gaynon, P.S., Winick, N.J., Reaman, G.H., Carroll, W.L., 2012. Improved Survival for Children and Adolescents With Acute Lymphoblastic Leukemia Between 1990 and 2005: A Report From the Children's Oncology Group. JCO 30, 1663–1669. https://doi.org/10.1200/JCO.2011.37.8018

Hunger, S.P., Mullighan, C.G., 2015. Acute Lymphoblastic Leukemia in Children. N Engl J Med 373, 1541–1552. https://doi.org/10.1056/NEJMra1400972

Iacobucci, I., Kimura, S., Mullighan, C.G., 2021. Biologic and Therapeutic Implications of Genomic Alterations in Acute Lymphoblastic Leukemia. JCM 10, 3792. https://doi.org/10.3390/jcm10173792

Ikawa, T., Hirose, S., Masuda, K., Kakugawa, K., Satoh, R., Shibano-Satoh, A., Kominami, R., Katsura, Y., Kawamoto, H., 2010. An Essential Developmental Checkpoint for Production of the T Cell Lineage. Science 329, 93–96. https://doi.org/10.1126/science.1188995

Ikawa, T., Kawamoto, H., Goldrath, A.W., Murre, C., 2006. E proteins and Notch signaling cooperate to promote T cell lineage specification and commitment. The Journal

of Experimental Medicine 203, 1329–1342. https://doi.org/10.1084/jem.20060268

Ivanov, I.I., McKenzie, B.S., Zhou, L., Tadokoro, C.E., Lepelley, A., Lafaille, J.J., Cua, D.J., Littman, D.R., 2006. The Orphan Nuclear Receptor RORγt Directs the Differentiation Program of Proinflammatory IL-17+ T Helper Cells. Cell 126, 1121–1133. https://doi.org/10.1016/j.cell.2006.07.035

Jaeger, N., Antonova, A.U., Kreisel, D., Roan, F., Lantelme, E., Ziegler, S.F., Cella, M., Colonna, M., 2024. Diversity of group 1 innate lymphoid cells in human tissues. Nat Immunol 25, 1460–1473. https://doi.org/10.1038/s41590-024-01885-y

Janssen, J.W., Ludwig, W.D., Sterry, W., Bartram, C.R., 1993. SIL-TAL1 deletion in T-cell acute lymphoblastic leukemia. Leukemia 7, 1204–1210.

Jardine, L., Webb, S., Goh, I., Quiroga Londoño, M., Reynolds, G., Mather, M., Olabi, B., Stephenson, E., Botting, R.A., Horsfall, D., Engelbert, J., Maunder, D., Mende, N., Murnane, C., Dann, E., McGrath, J., King, H., Kucinski, I., Queen, R., Carey, C.D., Shrubsole, C., Poyner, E., Acres, M., Jones, C., Ness, T., Coulthard, R., Elliott, N., O'Byrne, S., Haltalli, M.L.R., Lawrence, J.E., Lisgo, S., Balogh, P., Meyer, K.B., Prigmore, E., Ambridge, K., Jain, M.S., Efremova, M., Pickard, K., Creasey, T., Bacardit, J., Henderson, D., Coxhead, J., Filby, A., Hussain, R., Dixon, D., McDonald, D., Popescu, D.-M., Kowalczyk, M.S., Li, B., Ashenberg, O., Tabaka, M., Dionne, D., Tickle, T.L., Slyper, M., Rozenblatt-Rosen, O., Regev, A., Behjati, S., Laurenti, E., Wilson, N.K., Roy, A., Göttgens, B., Roberts, I., Teichmann, S.A., Haniffa, M., 2021. Blood and immune development in human fetal bone marrow and Down syndrome. Nature 598, 327–331. https://doi.org/10.1038/s41586-021-03929-x

Jenkinson, S., Koo, K., Mansour, M.R., Goulden, N., Vora, A., Mitchell, C., Wade, R., Richards, S., Hancock, J., Moorman, A.V., Linch, D.C., Gale, R.E., 2013. Impact of NOTCH1/FBXW7 mutations on outcome in pediatric T-cell acute lymphoblastic leukemia patients treated on the MRC UKALL 2003 trial. Leukemia 27, 41–47. https://doi.org/10.1038/leu.2012.176

Jeon, Y.K., Go, H., Nam, S.J., Keam, B., Kim, T.M., Jung, K.C., Kang, H.J., Lee, D.S., Huh, J.R., Park, S.H., 2012. Expression of the promyelocytic leukemia zinc-finger in T-lymphoblastic lymphoma and leukemia has strong implications for their cellular origin and greater association with initial bone marrow involvement. Modern Pathology 25, 1236–1245. https://doi.org/10.1038/modpathol.2012.82

Jones, D., Raine, K.M., Davies, H., Tarpey, P.S., Butler, A.P., Teague, J.W., Nik-Zainal, S., Campbell, P.J., 2016. cgpCaVEManWrapper: Simple Execution of CaVEMan in Order to Detect Somatic Single Nucleotide Variants in NGS Data. CP in Bioinformatics 56. https://doi.org/10.1002/cpbi.20

Jordan, M.S., Boesteanu, A., Reed, A.J., Petrone, A.L., Holenbeck, A.E., Lerman, M.A., Naji, A., Caton, A.J., 2001. Thymic selection of CD4+CD25+ regulatory T cells induced by an agonist self-peptide. Nat Immunol 2, 301–306. https://doi.org/10.1038/86302

Kampen, K.R., Sulima, S.O., Verbelen, B., Girardi, T., Vereecke, S., Rinaldi, G., Verbeeck, J., Op de Beeck, J., Uyttebroeck, A., Meijerink, J.P.P., Moorman, A.V., Harrison, C.J., Spincemaille, P., Cools, J., Cassiman, D., Fendt, S.-M., Vermeersch, P., De Keersmaecker, K., 2019. The ribosomal RPL10 R98S mutation drives IRES-

dependent BCL-2 translation in T-ALL. Leukemia 33, 319–332. https://doi.org/10.1038/s41375-018-0176-z

Kantarjian, H., Stein, A., Gökbuget, N., Fielding, A.K., Schuh, A.C., Ribera, J.-M., Wei, A., Dombret, H., Foà, R., Bassan, R., Arslan, Ö., Sanz, M.A., Bergeron, J., Demirkan, F., Lech-Maranda, E., Rambaldi, A., Thomas, X., Horst, H.-A., Brüggemann, M., Klapper, W., Wood, B.L., Fleishman, A., Nagorsen, D., Holland, C., Zimmerman, Z., Topp, M.S., 2017. Blinatumomab versus Chemotherapy for Advanced Acute Lymphoblastic Leukemia. N Engl J Med 376, 836–847. https://doi.org/10.1056/NEJMoa1609783

Khabirova, E., Jardine, L., Coorens, T.H.H., Webb, S., Treger, T.D., Engelbert, J., Porter, T., Prigmore, E., Collord, G., Piapi, A., Teichmann, S.A., Inglott, S., Williams, O., Heidenreich, O., Young, M.D., Straathof, K., Bomken, S., Bartram, J., Haniffa, M., Behjati, S., 2022. Single-cell transcriptomics reveals a distinct developmental state of KMT2A-rearranged infant B-cell acute lymphoblastic leukemia. Nat Med 28, 743–751. https://doi.org/10.1038/s41591-022-01720-7

Kiessling, R., Klein, E., Wigzell, H., 1975. “Natural” killer cells in the mouse. I. Cytotoxic cells with specificity for mouse Moloney leukemia cells. Specificity and distribution according to genotype. Eur J Immunol 5, 112–117. https://doi.org/10.1002/eji.1830050208

Kildisiute, G., Kalyva, M., Elmentaite, R., Van Dongen, S., Thevanesan, C., Piapi, A., Ambridge, K., Prigmore, E., Haniffa, M., Teichmann, S.A., Straathof, K., Cortés-Ciriano, I., Behjati, S., Young, M.D., 2024. Transcriptional signals of transformation in human cancer. Genome Med 16, 8. https://doi.org/10.1186/s13073-023-01279-z

Kildisiute, G., Kholosy, W.M., Young, M.D., Roberts, K., Elmentaite, R., van Hooff, S.R., Pacyna, C.N., Khabirova, E., Piapi, A., Thevanesan, C., Bugallo-Blanco, E., Burke, C., Mamanova, L., Keller, K.M., Langenberg-Ververgaert, K.P.S., Lijnzaad, P., Margaritis, T., Holstege, F.C.P., Tas, M.L., Wijnen, M.H.W.A., van Noesel, M.M., del Valle, I., Barone, G., van der Linden, R., Duncan, C., Anderson, J., Achermann, J.C., Haniffa, M., Teichmann, S.A., Rampling, D., Sebire, N.J., He, X., de Krijger, R.R., Barker, R.A., Meyer, K.B., Bayraktar, O., Straathof, K., Molenaar, J.J., Behjati, S., 2021. Tumor to normal single-cell mRNA comparisons reveal a pan-neuroblastoma cancer cell. Science Advances 7, eabd3311. https://doi.org/10.1126/sciadv.abd3311

Kimura, S., Park, C.S., Montefiori, L.E., Iacobucci, I., Pölönen, P., Gao, Q., Arnold, E.D., Attarbaschi, A., Brown, A., Buldini, B., Caldwell, K.J., Chang, Y., Chen, C., Cheng, C., Cheng, Z., Choi, J., Conter, V., Crews, K.R., De Groot-Kruseman, H.A., Deguchi, T., Eguchi, M., Muhle, H.E., Elitzur, S., Escherich, G., Freeman, B.B., Gu, Z., Han, K., Horibe, K., Imamura, T., Jeha, S., Kato, M., Chiew, K.H., Khan, T., Kicinski, M., Köhrer, S., Kornblau, S.M., Kotecha, R.S., Li, C.-K., Liu, Y.-C., Locatelli, F., Luger, S.M., Paietta, E.M., Manabe, A., Marquart, H.V., Masetti, R., Maybury, M., Mazilier, P., Meijerink, J.P.P., Mitchell, S., Miyamura, T., Moore, A.S., Oshima, K., Pawinska-Wasikowska, K., Pieters, R., Prater, M.S., Pruett-Miller, S.M., Pui, C.-H., Qu, C., Reiterova, M., Reyes, N., Roberts, K.G., Rowe, J.M., Sato, A., Schmiegelow, K., Schrappe, M., Shen, S., Skoczeń, S., Spinelli, O., Stary, J., Svaton, M., Takagi, M., Takita, J., Tang, Y., Teachey, D.T., Thomas, P.G., Tomizawa, D., Trka, J., Varotto, E., Vincent, T.L., Yang, J.J., Yeoh, A.E.J., Zhou, Y.,

Zimmermann, M., Inaba, H., Mullighan, C.G., 2024. Biologic and Clinical Analysis of Childhood Gamma Delta T-ALL Identifies *LMO2/STAG2* Rearrangements as Extremely High Risk. Cancer Discovery OF1–OF22. https://doi.org/10.1158/2159-8290.CD-23-1452

Kjer-Nielsen, L., Patel, O., Corbett, A.J., Le Nours, J., Meehan, B., Liu, L., Bhati, M., Chen, Z., Kostenko, L., Reantragoon, R., Williamson, N.A., Purcell, A.W., Dudek, N.L., McConville, M.J., O'Hair, R.A.J., Khairallah, G.N., Godfrey, D.I., Fairlie, D.P., Rossjohn, J., McCluskey, J., 2012. MR1 presents microbial vitamin B metabolites to MAIT cells. Nature 491, 717–723. https://doi.org/10.1038/nature11605

Klein Wolterink, R.G.J., Serafini, N., Van Nimwegen, M., Vosshenrich, C.A.J., De Bruijn, M.J.W., Fonseca Pereira, D., Veiga Fernandes, H., Hendriks, R.W., Di Santo, J.P., 2013. Essential, dose-dependent role for the transcription factor *Gata3* in the development of IL-5 $^{+}$ and IL-13 $^{+}$ type 2 innate lymphoid cells. Proc. Natl. Acad. Sci. U.S.A. 110, 10240–10245. https://doi.org/10.1073/pnas.1217158110

Kloetgen, A., Thandapani, P., Ntziachristos, P., Ghebrechristos, Y., Nomikou, S., Lazaris, C., Chen, X., Hu, H., Bakogianni, S., Wang, J., Fu, Y., Boccalatte, F., Zhong, H., Paietta, E., Trimarchi, T., Zhu, Y., Van Vlierberghe, P., Inghirami, G.G., Lionnet, T., Aifantis, I., Tsirigos, A., 2020. Three-dimensional chromatin landscapes in T cell acute lymphoblastic leukemia. Nat Genet 52, 388–400. https://doi.org/10.1038/s41588-020-0602-9

Klose, C.S.N., Flach, M., Möhle, L., Rogell, L., Hoyler, T., Ebert, K., Fabiunke, C., Pfeifer, D., Sexl, V., Fonseca-Pereira, D., Domingues, R.G., Veiga-Fernandes, H., Arnold, S.J., Busslinger, M., Dunay, I.R., Tanriver, Y., Diefenbach, A., 2014. Differentiation of Type 1 ILCs from a Common Progenitor to All Helper-like Innate Lymphoid Cell Lineages. Cell 157, 340–356. https://doi.org/10.1016/j.cell.2014.03.030

Koay, H.-F., Gherardin, N.A., Enders, A., Loh, L., Mackay, L.K., Almeida, C.F., Russ, B.E., Nold-Petry, C.A., Nold, M.F., Bedoui, S., Chen, Z., Corbett, A.J., Eckle, S.B.G., Meehan, B., d'Udekem, Y., Konstantinov, I.E., Lappas, M., Liu, L., Goodnow, C.C., Fairlie, D.P., Rossjohn, J., Chong, M.M., Kedzierska, K., Berzins, S.P., Belz, G.T., McCluskey, J., Uldrich, A.P., Godfrey, D.I., Pellicci, D.G., 2016. A three-stage intrathymic development pathway for the mucosal-associated invariant T cell lineage. Nat Immunol 17, 1300–1311. https://doi.org/10.1038/ni.3565

Komori, T., Okada, A., Stewart, V., Alt, F.W., 1993. Lack of N Regions in Antigen Receptor Variable Region Genes of TdT-Deficient Lymphocytes. Science 261, 1171–1175. https://doi.org/10.1126/science.8356451

Konno, A., Okada, K., Mizuno, K., Nishida, M., Nagaoki, S., Toma, T., Uehara, T., Ohta, K., Kasahara, Y., Seki, H., Yachie, A., Koizumi, S., 2002. CD8αα memory effector T cells descend directly from clonally expanded CD8α+βhigh TCRαβ T cells in vivo. Blood 100, 4090–4097. https://doi.org/10.1182/blood-2002-04-1136

Kovalovsky, D., Alonzo, E.S., Uche, O.U., Eidson, M., Nichols, K.E., Sant'Angelo, D.B., 2010. PLZF Induces the Spontaneous Acquisition of Memory/Effector Functions in T Cells Independently of NKT Cell-Related Signals. The Journal of Immunology 184, 6746–6755. https://doi.org/10.4049/jimmunol.1000776

Kovalovsky, D., Uche, O.U., Eladad, S., Hobbs, R.M., Yi, W., Alonzo, E., Chua, K., Eidson, M., Kim, H.-J., Im, J.S., Pandolfi, P.P., Sant'Angelo, D.B., 2008. The BTB–zinc finger transcriptional regulator PLZF controls the development of invariant natural killer

T cell effector functions. Nat Immunol 9, 1055–1064. https://doi.org/10.1038/ni.1641

Krangel, M.S., 2009. Mechanics of T cell receptor gene rearrangement. Current Opinion in Immunology 21, 133–139. https://doi.org/10.1016/j.coi.2009.03.009

Larson, R.C., Lavenir, I., Larson, T.A., Baer, R., Warren, A.J., Wadman, I., Nottage, K., Rabbitts, T.H., 1996. Protein dimerization between Lmo2 (Rbtn2) and Tal1 alters thymocyte development and potentiates T cell tumorigenesis in transgenic mice. EMBO J 15, 1021–1027.

Le Bourhis, L., Martin, E., Péguillet, I., Guihot, A., Froux, N., Coré, M., Lévy, E., Dusseaux, M., Meyssonnier, V., Premel, V., Ngo, C., Riteau, B., Duban, L., Robert, D., Huang, S., Rottman, M., Soudais, C., Lantz, O., 2010. Antimicrobial activity of mucosal-associated invariant T cells. Nat Immunol 11, 701–708. https://doi.org/10.1038/ni.1890

Lee, S.H.R., Yang, Wenjian, Gocho, Y., John, A., Rowland, L., Smart, B., Williams, H., Maxwell, D., Hunt, J., Yang, Wentao, Crews, K.R., Roberts, K.G., Jeha, S., Cheng, C., Karol, S.E., Relling, M.V., Rosner, G.L., Inaba, H., Mullighan, C.G., Pui, C.-H., Evans, W.E., Yang, J.J., 2023. Pharmacotypes across the genomic landscape of pediatric acute lymphoblastic leukemia and impact on treatment response. Nat Med 29, 170–179. https://doi.org/10.1038/s41591-022-02112-7

Lefranc, M.-P., Giudicelli, V., Ginestoux, C., Bodmer, J., Müller, W., Bontrop, R., Lemaitre, M., Malik, A., Barbié, V., Chaume, D., 1999. IMGT, the international ImMunoGeneTics database. Nucleic Acids Research 27, 209–212. https://doi.org/10.1093/nar/27.1.209

Legoux, F., Salou, M., Lantz, O., 2017. Unconventional or Preset αβ T Cells: Evolutionarily Conserved Tissue-Resident T Cells Recognizing Nonpeptidic Ligands. Annu. Rev. Cell Dev. Biol. 33, 511–535. https://doi.org/10.1146/annurev-cellbio-100616-060725

Li, H., Durbin, R., 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics 25, 1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Li, H., Van Der Leun, A.M., Yofe, I., Lubling, Y., Gelbard-Solodkin, D., Van Akkooi, A.C.J., Van Den Braber, M., Rozeman, E.A., Haanen, J.B.A.G., Blank, C.U., Horlings, H.M., David, E., Baran, Y., Bercovich, A., Lifshitz, A., Schumacher, T.N., Tanay, A., Amit, I., 2019. Dysfunctional CD8 T Cells Form a Proliferative, Dynamically Regulated Compartment within Human Melanoma. Cell 176, 775-789.e18. https://doi.org/10.1016/j.cell.2018.11.043

Li, L., Leid, M., Rothenberg, E.V., 2010. An Early T Cell Lineage Commitment Checkpoint Dependent on the Transcription Factor *Bcl11b*. Science 329, 89–93. https://doi.org/10.1126/science.1188989

Li, P., Burke, S., Wang, J., Chen, X., Ortiz, M., Lee, S.-C., Lu, D., Campos, L., Goulding, D., Ng, B.L., Dougan, G., Huntly, B., Gottgens, B., Jenkins, N.A., Copeland, N.G., Colucci, F., Liu, P., 2010. Reprogramming of T Cells to Natural Killer–Like Cells upon *Bcl11b* Deletion. Science 329, 85–89. https://doi.org/10.1126/science.1188063

Liao, Y., Smyth, G.K., Shi, W., 2014. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics 30, 923–930. https://doi.org/10.1093/bioinformatics/btt656

Lim, A.I., Li, Y., Lopez-Lastra, S., Stadhouders, R., Paul, F., Casrouge, A., Serafini, N., Puel, A., Bustamante, J., Surace, L., Masse-Ranson, G., David, E., Strick-Marchand, H., Le Bourhis, L., Cocchi, R., Topazio, D., Graziano, P., Muscarella, L.A., Rogge, L., Norel, X., Sallenave, J.-M., Allez, M., Graf, T., Hendriks, R.W., Casanova, J.-L., Amit, I., Yssel, H., Di Santo, J.P., 2017. Systemic Human ILC Precursors Provide a Substrate for Tissue ILC Differentiation. Cell 168, 1086-1100.e10. https://doi.org/10.1016/j.cell.2017.02.021

Liu, Y., Easton, J., Shao, Y., Maciaszek, J., Wang, Z., Wilkinson, M.R., McCastlain, K., Edmonson, M., Pounds, S.B., Shi, L., Zhou, X., Ma, X., Sioson, E., Li, Y., Rusch, M., Gupta, P., Pei, D., Cheng, C., Smith, M.A., Auvil, J.G., Gerhard, D.S., Relling, M.V., Winick, N.J., Carroll, A.J., Heerema, N.A., Raetz, E., Devidas, M., Willman, C.L., Harvey, R.C., Carroll, W.L., Dunsmore, K.P., Winter, S.S., Wood, B.L., Sorrentino, B.P., Downing, J.R., Loh, M.L., Hunger, S.P., Zhang, J., Mullighan, C.G., 2017. The genomic landscape of pediatric and young adult T-lineage acute lymphoblastic leukemia. Nat Genet 49, 1211–1218. https://doi.org/10.1038/ng.3909

Lu, J., Song, G., Tang, Q., Zou, C., Han, F., Zhao, Z., Yong, B., Yin, J., Xu, H., Xie, X., Kang, T., Lam, Y., Yang, H., Shen, J., Wang, J., 2015. IRX1 hypomethylation promotes osteosarcoma metastasis via induction of CXCL14/NF-κB signaling. J Clin Invest 125, 1839–1856. https://doi.org/10.1172/JCI78437

Luci, C., Reynders, A., Ivanov, I.I., Cognet, C., Chiche, L., Chasson, L., Hardwigsen, J., Anguiano, E., Banchereau, J., Chaussabel, D., Dalod, M., Littman, D.R., Vivier, E., Tomasello, E., 2009. Influence of the transcription factor RORγt on the development of NKp46+ cell populations in gut and skin. Nat Immunol 10, 75–82. https://doi.org/10.1038/ni.1681

Ludwig, L.S., Lareau, C.A., Ulirsch, J.C., Christian, E., Muus, C., Li, L.H., Pelka, K., Ge, W., Oren, Y., Brack, A., Law, T., Rodman, C., Chen, J.H., Boland, G.M., Hacohen, N., Rozenblatt-Rosen, O., Aryee, M.J., Buenrostro, J.D., Regev, A., Sankaran, V.G., 2019. Lineage Tracing in Humans Enabled by Mitochondrial Mutations and Single-Cell Genomics. Cell 176, 1325-1339.e22. https://doi.org/10.1016/j.cell.2019.01.022

Machado, H.E., Mitchell, E., Øbro, N.F., Kübler, K., Davies, M., Leongamornlert, D., Cull, A., Maura, F., Sanders, M.A., Cagan, A.T.J., McDonald, C., Belmonte, M., Shepherd, M.S., Vieira Braga, F.A., Osborne, R.J., Mahbubani, K., Martincorena, I., Laurenti, E., Green, A.R., Getz, G., Polak, P., Saeb-Parsy, K., Hodson, D.J., Kent, D.G., Campbell, P.J., 2022. Diverse mutational landscapes in human lymphocytes. Nature 608, 724–732. https://doi.org/10.1038/s41586-022-05072-7

Maciocia, P.M., Wawrzyniecka, P.A., Maciocia, N.C., Burley, A., Karpanasamy, T., Devereaux, S., Hoekx, M., O'Connor, D., Leon, T., Rapoz-D'Silva, T., Pocock, R., Rahman, S., Gritti, G., Yánez, D.C., Ross, S., Crompton, T., Williams, O., Lee, L., Pule, M.A., Mansour, M.R., 2022. Anti-CCR9 chimeric antigen receptor T cells for T-cell acute lymphoblastic leukemia. Blood 140, 25–37. https://doi.org/10.1182/blood.2021013648

Makino, Y., Kanno, R., Ito, T., Higashino, K., Taniguchi, M., 1995. Predominant expression of invariant Vα14+ TCR α chain in NK1.1+ T cell populations. International Immunology 7, 1157–1161. https://doi.org/10.1093/intimm/7.7.1157

Malard, F., Mohty, M., 2020. Acute lymphoblastic leukaemia. The Lancet 395, 1146–1162. https://doi.org/10.1016/S0140-6736(19)33018-1

Mansour, M.R., Abraham, B.J., Anders, L., Berezovskaya, A., Gutierrez, A., Durbin, A.D., Etchin, J., Lawton, L., Sallan, S.E., Silverman, L.B., Loh, M.L., Hunger, S.P., Sanda, T., Young, R.A., Look, A.T., 2014. An oncogenic super-enhancer formed through somatic mutation of a noncoding intergenic element. Science 346, 1373–1377. https://doi.org/10.1126/science.1259037

Mansour, M.R., Duke, V., Foroni, L., Patel, B., Allen, C.G., Ancliff, P.J., Gale, R.E., Linch, D.C., 2007. Notch-1 Mutations Are Secondary Events in Some Patients with T-Cell Acute Lymphoblastic Leukemia. Clinical Cancer Research 13, 6964–6969. https://doi.org/10.1158/1078-0432.CCR-07-1474

Marks, B.R., Nowyhed, H.N., Choi, J.-Y., Poholek, A.C., Odegard, J.M., Flavell, R.A., Craft, J., 2009. Thymic self-reactivity selects natural interleukin 17–producing T cells that can regulate peripheral inflammation. Nat Immunol 10, 1125–1132. https://doi.org/10.1038/ni.1783

Martin, E., Treiner, E., Duban, L., Guerri, L., Laude, H., Toly, C., Premel, V., Devys, A., Moura, I.C., Tilloy, F., Cherif, S., Vera, G., Latour, S., Soudais, C., Lantz, O., 2009. Stepwise Development of MAIT Cells in Mouse and Human. PLoS Biol 7, e1000054. https://doi.org/10.1371/journal.pbio.1000054

Matyskiela, M.E., Zhu, J., Baughman, J.M., Clayton, T., Slade, M., Wong, H.K., Danga, K., Zheng, X., Labow, M., LeBrun, L., Lu, G., Chamberlain, P.P., Thompson, J.W., 2020. Cereblon Modulators Target ZBTB16 and Its Oncogenic Fusion Partners for Degradation via Distinct Structural Degrons. ACS Chem. Biol. 15, 3149–3158. https://doi.org/10.1021/acschembio.0c00674

Maude, S.L., Frey, N., Shaw, P.A., Aplenc, R., Barrett, D.M., Bunin, N.J., Chew, A., Gonzalez, V.E., Zheng, Z., Lacey, S.F., Mahnke, Y.D., Melenhorst, J.J., Rheingold, S.R., Shen, A., Teachey, D.T., Levine, B.L., June, C.H., Porter, D.L., Grupp, S.A., 2014. Chimeric Antigen Receptor T Cells for Sustained Remissions in Leukemia. N Engl J Med 371, 1507–1517. https://doi.org/10.1056/NEJMoa1407222

McGregor, S., Shah, A., Raca, G., Mirza, M.K., Smith, S.M., Anastasi, J., Vardiman, J.W., Hyjek, E., Gurbuxani, S., 2014. PLZF staining identifies peripheral T-cell lymphomas derived from innate-like T-cells with TRAV1-2-TRAJ33 TCR-α rearrangement. Blood 123, 2742–2743. https://doi.org/10.1182/blood-2014-02-555482

McGuire, E.A., Hockett, R.D., Pollock, K.M., Bartholdi, M.F., O'Brien, S.J., Korsmeyer, S.J., 1989. The t(11;14)(p15;q11) in a T-cell acute lymphoblastic leukemia cell line activates multiple transcripts, including Ttg-1, a gene encoding a potential zinc finger protein. Mol Cell Biol 9, 2124–2132. https://doi.org/10.1128/mcb.9.5.2124-2132.1989

McVay, L.D., Carding, S.R., 1996. Extrathymic origin of human gamma delta T cells during fetal development. J Immunol 157, 2873–2882.

McVay, L.D., Jaswal, S.S., Kennedy, C., Hayday, A., Carding, S.R., 1998. The generation of human gammadelta T cell repertoires during fetal development. J Immunol 160, 5851–5860.

Mebius, R.E., Rennert, P., Weissman, I.L., 1997. Developing Lymph Nodes Collect CD4+CD3− LTβ+ Cells That Can Differentiate to APC, NK Cells, and Follicular

Cells but Not T or B Cells. Immunity 7, 493–504. https://doi.org/10.1016/S1074-7613(00)80371-4
Mellentin, J.D., Smith, S.D., Cleary, M.L., 1989. lyl-1, a novel gene altered by chromosomal translocation in T cell leukemia, codes for a protein with a helix-loop-helix DNA binding motif. Cell 58, 77–83. https://doi.org/10.1016/0092-8674(89)90404-2
Miller, J.F., Mitchell, G.F., 1968. Cell to cell interaction in the immune response. I. Hemolysin-forming cells in neonatally thymectomized mice reconstituted with thymus or thoracic duct lymphocytes. J Exp Med 128, 801–820. https://doi.org/10.1084/jem.128.4.801
Miller, J.F.A.P., 1961a. Analysis of the Thymus Influence in Leukaemogenesis. Nature 191, 248–249. https://doi.org/10.1038/191248a0
Miller, J.F.A.P., 1961b. Immunological Function of the Thymus. The Lancet 278, 748–749. https://doi.org/10.1016/S0140-6736(61)90693-6
Miller, J.F.A.P., 1959a. Role of the Thymus in Murine Leukaemia. Nature 183, 1069–1069. https://doi.org/10.1038/1831069a0
Miller, J.F.A.P., 1959b. Fate of Subcutaneous Thymus Grafts in Thymectomized Mice inoculated with Leukaemic Filtrate. Nature 184, 1809–1810. https://doi.org/10.1038/1841809a0
Miller, J.F.A.P., Haddow, A., 1962. Effect of neonatal thymectomy on the immunological responsiveness of the mouse. Proceedings of the Royal Society of London. Series B. Biological Sciences 156, 415–428. https://doi.org/10.1098/rspb.1962.0048
Mitchell, E., Spencer Chapman, M., Williams, N., Dawson, K.J., Mende, N., Calderbank, E.F., Jung, H., Mitchell, T., Coorens, T.H.H., Spencer, D.H., Machado, H., Lee-Six, H., Davies, M., Hayler, D., Fabre, M.A., Mahbubani, K., Abascal, F., Cagan, A., Vassiliou, G.S., Baxter, J., Martincorena, I., Stratton, M.R., Kent, D.G., Chatterjee, K., Parsy, K.S., Green, A.R., Nangalia, J., Laurenti, E., Campbell, P.J., 2022. Clonal dynamics of haematopoiesis across the human lifespan. Nature 606, 343–350. https://doi.org/10.1038/s41586-022-04786-y
Mitchell, G.F., Miller, J.F., 1968a. Immunological activity of thymus and thoracic-duct lymphocytes. Proc Natl Acad Sci U S A 59, 296–303. https://doi.org/10.1073/pnas.59.1.296
Mitchell, G.F., Miller, J.F., 1968b. Cell to cell interaction in the immune response. II. The source of hemolysin-forming cells in irradiated mice given bone marrow and thymus or thoracic duct lymphocytes. J Exp Med 128, 821–837. https://doi.org/10.1084/jem.128.4.821
Miyazaki, K., Miyazaki, M., Murre, C., 2014. The establishment of B versus T cell identity. Trends in Immunology 35, 205–210. https://doi.org/10.1016/j.it.2014.02.009
Mjösberg, J., Bernink, J., Golebski, K., Karrich, J.J., Peters, C.P., Blom, B., te Velde, A.A., Fokkens, W.J., van Drunen, C.M., Spits, H., 2012. The Transcription Factor GATA3 Is Essential for the Function of Human Type 2 Innate Lymphoid Cells. Immunity 37, 649–659. https://doi.org/10.1016/j.immuni.2012.08.015
Mohammed, H., Taylor, C., Brown, G.D., Papachristou, E.K., Carroll, J.S., D'Santos, C.S., 2016. Rapid immunoprecipitation mass spectrometry of endogenous proteins (RIME) for analysis of chromatin complexes. Nature Protocols 11, 316–326. https://doi.org/10.1038/nprot.2016.020

Montefiori, L.E., Bendig, S., Gu, Z., Chen, X., Pölönen, P., Ma, X., Murison, A., Zeng, A., Garcia-Prat, L., Dickerson, K., Iacobucci, I., Abdelhamed, S., Hiltenbrand, R., Mead, P.E., Mehr, C.M., Xu, B., Cheng, Z., Chang, T.-C., Westover, T., Ma, J., Stengel, A., Kimura, S., Qu, C., Valentine, M.B., Rashkovan, M., Luger, S., Litzow, M.R., Rowe, J.M., Den Boer, M.L., Wang, V., Yin, J., Kornblau, S.M., Hunger, S.P., Loh, M.L., Pui, C.-H., Yang, W., Crews, K.R., Roberts, K.G., Yang, J.J., Relling, M.V., Evans, W.E., Stock, W., Paietta, E.M., Ferrando, A.A., Zhang, J., Kern, W., Haferlach, T., Wu, G., Dick, J.E., Klco, J.M., Haferlach, C., Mullighan, C.G., 2021. Enhancer Hijacking Drives Oncogenic *BCL11B* Expression in Lineage-Ambiguous Stem Cell Leukemia. Cancer Discovery 11, 2846–2867. https://doi.org/10.1158/2159-8290.CD-21-0145

Monzó, C., Liu, T., Conesa, A., 2025. Transcriptomics in the era of long-read sequencing. Nat Rev Genet. https://doi.org/10.1038/s41576-025-00828-z

Moro, K., Yamada, T., Tanabe, M., Takeuchi, T., Ikawa, T., Kawamoto, H., Furusawa, J., Ohtani, M., Fujii, H., Koyasu, S., 2010. Innate production of TH2 cytokines by adipose tissue-associated c-Kit+Sca-1+ lymphoid cells. Nature 463, 540–544. https://doi.org/10.1038/nature08636

Muyas, F., Sauer, C.M., Valle-Inclán, J.E., Li, R., Rahbari, R., Mitchell, T.J., Hormoz, S., Cortés-Ciriano, I., 2024. De novo detection of somatic mutations in high-throughput single-cell profiling data sets. Nat Biotechnol 42, 758–767. https://doi.org/10.1038/s41587-023-01863-z

Nam, A.S., Kim, K.-T., Chaligne, R., Izzo, F., Ang, C., Taylor, J., Myers, R.M., Abu-Zeinah, G., Brand, R., Omans, N.D., Alonso, A., Sheridan, C., Mariani, M., Dai, X., Harrington, E., Pastore, A., Cubillos-Ruiz, J.R., Tam, W., Hoffman, R., Rabadan, R., Scandura, J.M., Abdel-Wahab, O., Smibert, P., Landau, D.A., 2019. Somatic mutations and cell identity linked by Genotyping of Transcriptomes. Nature 571, 355–360. https://doi.org/10.1038/s41586-019-1367-0

Neill, D.R., Wong, S.H., Bellosi, A., Flynn, R.J., Daly, M., Langford, T.K.A., Bucks, C., Kane, C.M., Fallon, P.G., Pannell, R., Jolin, H.E., McKenzie, A.N.J., 2010. Nuocytes represent a new innate effector leukocyte that mediates type-2 immunity. Nature 464, 1367–1370. https://doi.org/10.1038/nature08900

Nik-Zainal, S., Davies, H., Staaf, J., Ramakrishna, M., Glodzik, D., Zou, X., Martincorena, I., Alexandrov, L.B., Martin, S., Wedge, D.C., Van Loo, P., Ju, Y.S., Smid, M., Brinkman, A.B., Morganella, S., Aure, M.R., Lingjærde, O.C., Langerød, A., Ringnér, M., Ahn, S.-M., Boyault, S., Brock, J.E., Broeks, A., Butler, A., Desmedt, C., Dirix, L., Dronov, S., Fatima, A., Foekens, J.A., Gerstung, M., Hooijer, G.K.J., Jang, S.J., Jones, D.R., Kim, H.-Y., King, T.A., Krishnamurthy, S., Lee, H.J., Lee, J.-Y., Li, Y., McLaren, S., Menzies, A., Mustonen, V., O'Meara, S., Pauporté, I., Pivot, X., Purdie, C.A., Raine, K., Ramakrishnan, K., Rodríguez-González, F.G., Romieu, G., Sieuwerts, A.M., Simpson, P.T., Shepherd, R., Stebbings, L., Stefansson, O.A., Teague, J., Tommasi, S., Treilleux, I., Van Den Eynden, G.G., Vermeulen, P., Vincent-Salomon, A., Yates, L., Caldas, C., Veer, L.V., Tutt, A., Knappskog, S., Tan, B.K.T., Jonkers, J., Borg, Å., Ueno, N.T., Sotiriou, C., Viari, A., Futreal, P.A., Campbell, P.J., Span, P.N., Van Laere, S., Lakhani, S.R., Eyfjord, J.E., Thompson, A.M., Birney, E., Stunnenberg, H.G., Van De Vijver, M.J., Martens, J.W.M., Børresen-Dale, A.-L., Richardson, A.L., Kong, G., Thomas, G., Stratton, M.R.,

2016. Landscape of somatic mutations in 560 breast cancer whole-genome sequences. Nature 534, 47–54. https://doi.org/10.1038/nature17676

Nik-Zainal, S., Van Loo, P., Wedge, D.C., Alexandrov, L.B., Greenman, C.D., Lau, K.W., Raine, K., Jones, D., Marshall, J., Ramakrishna, M., Shlien, A., Cooke, S.L., Hinton, J., Menzies, A., Stebbings, L.A., Leroy, C., Jia, M., Rance, R., Mudie, L.J., Gamble, S.J., Stephens, P.J., McLaren, S., Tarpey, P.S., Papaemmanuil, E., Davies, H.R., Varela, I., McBride, D.J., Bignell, G.R., Leung, K., Butler, A.P., Teague, J.W., Martin, S., Jönsson, G., Mariani, O., Boyault, S., Miron, P., Fatima, A., Langerød, A., Aparicio, S.A.J.R., Tutt, A., Sieuwerts, A.M., Borg, Å., Thomas, G., Salomon, A.V., Richardson, A.L., Børresen-Dale, A.-L., Futreal, P.A., Stratton, M.R., Campbell, P.J., 2012. The Life History of 21 Breast Cancers. Cell 149, 994–1007. https://doi.org/10.1016/j.cell.2012.04.023

O'Connor, D., Demeulemeester, J., Conde, L., Kirkwood, A., Fung, K., Papaleonidopoulou, F., Bloye, G., Farah, N., Rahman, S., Hancock, J., Bateman, C., Inglott, S., Mee, J., Herrero, J., Van Loo, P., Moorman, A.V., Vora, A., Mansour, M.R., 2023. The Clinicogenomic Landscape of Induction Failure in Childhood and Young Adult T-Cell Acute Lymphoblastic Leukemia. JCO 41, 3545–3556. https://doi.org/10.1200/JCO.22.02734

O'Connor, D., Moorman, A.V., Wade, R., Hancock, J., Tan, R.M.R., Bartram, J., Moppett, J., Schwab, C., Patrick, K., Harrison, C.J., Hough, R., Goulden, N., Vora, A., Samarasinghe, S., 2017. Use of Minimal Residual Disease Assessment to Redefine Induction Failure in Pediatric Acute Lymphoblastic Leukemia. JCO 35, 660–667. https://doi.org/10.1200/JCO.2016.69.6278

O'Connor, D., Valle-Inclán, J.E., Conde, L., Bloye, G., Rahman, S., Costa, J.R., Bartram, J., Adams, S., Wright, G., Elrick, H., Wall, K., Dyer, S., Howell, C., Jigoulina, G., Herrero, J., Cortes-Ciriano, I., Moorman, A.V., Mansour, M.R., 2024. Noncoding mutations drive persistence of a founder preleukemic clone which initiates late relapse in T-ALL. Blood 143, 933–937. https://doi.org/10.1182/blood.2023021906

Oeffinger, K.C., Kawashima, T., Friedman, D.L., Kadan-Lottick, N.S., Robison, L.L., 2006. Chronic Health Conditions in Adult Survivors of Childhood Cancer. n engl j med.

Oh, B.L.Z., Shimasaki, N., Coustan-Smith, E., Chan, E., Poon, L., Lee, S.H.R., Yeap, F., Tan, L.K., Chai, L.Y.A., Le Bert, N., Tan, N., Bertoletti, A., Chen, S.P., Del Bufalo, F., Becilli, M., Locatelli, F., Yeoh, A.E.J., Campana, D., 2024. Fratricide-resistant CD7-CAR T cells in T-ALL. Nat Med 30, 3687–3696. https://doi.org/10.1038/s41591-024-03228-8

Omura-Minamisawa, M., Diccianni, M.B., Batova, A., Chang, R.C., Bridgeman, L.J., Yu, J., Pullen, J., Bowman, W.P., Yu, A.L., 2000. Universal inactivation of both p16 and p15 but not downstream components is an essential event in the pathogenesis of T-cell acute lymphoblastic leukemia. Clin Cancer Res 6, 1219–1228.

O'Neil, J., Grim, J., Strack, P., Rao, S., Tibbitts, D., Winter, C., Hardwick, J., Welcker, M., Meijerink, J.P., Pieters, R., Draetta, G., Sears, R., Clurman, B.E., Look, A.T., 2007. FBW7 mutations in leukemic cells mediate NOTCH pathway activation and resistance to γ-secretase inhibitors. Journal of Experimental Medicine 204, 1813–1824. https://doi.org/10.1084/jem.20070876

Oshima, K., Khiabanian, H., Da Silva-Almeida, A.C., Tzoneva, G., Abate, F., Ambesi-Impiombato, A., Sanchez-Martin, M., Carpenter, Z., Penson, A., Perez-Garcia, A., Eckert, C., Nicolas, C., Balbin, M., Sulis, M.L., Kato, M., Koh, K., Paganin, M.,

Basso, G., Gastier-Foster, J.M., Devidas, M., Loh, M.L., Kirschner-Schwabe, R., Palomero, T., Rabadan, R., Ferrando, A.A., 2016. Mutational landscape, clonal evolution patterns, and role of RAS mutations in relapsed acute lymphoblastic leukemia. Proc. Natl. Acad. Sci. U.S.A. 113, 11306–11311. https://doi.org/10.1073/pnas.1608420113

Pagliaro, L., Chen, S.-J., Herranz, D., Mecucci, C., Harrison, C.J., Mullighan, C.G., Zhang, M., Chen, Z., Boissel, N., Winter, S.S., Roti, G., 2024. Acute lymphoblastic leukaemia. Nat Rev Dis Primers 10, 41. https://doi.org/10.1038/s41572-024-00525-x

Palomero, T., Lim, W.K., Odom, D.T., Sulis, M.L., Real, P.J., Margolin, A., Barnes, K.C., O'Neil, J., Neuberg, D., Weng, A.P., Aster, J.C., Sigaux, F., Soulier, J., Look, A.T., Young, R.A., Califano, A., Ferrando, A.A., 2006. NOTCH1 directly regulates *c-MYC* and activates a feed-forward-loop transcriptional network promoting leukemic cell growth. Proc. Natl. Acad. Sci. U.S.A. 103, 18261–18266. https://doi.org/10.1073/pnas.0606108103

Pan, J., Tan, Y., Shan, L., Seery, S., Deng, B., Ling, Z., Xu, J., Duan, J., Wang, Z., Wang, K., Yu, X., Zheng, Q., Xu, X., Hu, G., Tan, T., Yuan, Y., Tian, Z., Yan, F., Han, Y., Zhang, J., Feng, X., 2025. Allogeneic CD5-specific CAR-T therapy for relapsed/refractory T-ALL: a phase 1 trial. Nat Med 31, 126–136. https://doi.org/10.1038/s41591-024-03282-2

Papadopoulou, M., Tieppo, P., McGovern, N., Gosselin, F., Chan, J.K.Y., Goetgeluk, G., Dauby, N., Cogan, A., Donner, C., Ginhoux, F., Vandekerckhove, B., Vermijlen, D., 2019. TCR Sequencing Reveals the Distinct Development of Fetal and Adult Human Vγ9Vδ2 T Cells. The Journal of Immunology 203, 1468–1479. https://doi.org/10.4049/jimmunol.1900592

Papaemmanuil, E., Rapado, I., Li, Y., Potter, N.E., Wedge, D.C., Tubio, J., Alexandrov, L.B., Van Loo, P., Cooke, S.L., Marshall, J., Martincorena, I., Hinton, J., Gundem, G., Van Delft, F.W., Nik-Zainal, S., Jones, D.R., Ramakrishna, M., Titley, I., Stebbings, L., Leroy, C., Menzies, A., Gamble, J., Robinson, B., Mudie, L., Raine, K., O'Meara, S., Teague, J.W., Butler, A.P., Cazzaniga, G., Biondi, A., Zuna, J., Kempski, H., Muschen, M., Ford, A.M., Stratton, M.R., Greaves, M., Campbell, P.J., 2014. RAG-mediated recombination is the predominant driver of oncogenic rearrangement in ETV6-RUNX1 acute lymphoblastic leukemia. Nat Genet 46, 116–125. https://doi.org/10.1038/ng.2874

Park, J.-E., Botting, R.A., Domínguez Conde, C., Popescu, D.-M., Lavaert, M., Kunz, D.J., Goh, I., Stephenson, E., Ragazzini, R., Tuck, E., Wilbrey-Clark, A., Roberts, K., Kedlian, V.R., Ferdinand, J.R., He, X., Webb, S., Maunder, D., Vandamme, N., Mahbubani, K.T., Polanski, K., Mamanova, L., Bolt, L., Crossland, D., de Rita, F., Fuller, A., Filby, A., Reynolds, G., Dixon, D., Saeb-Parsy, K., Lisgo, S., Henderson, D., Vento-Tormo, R., Bayraktar, O.A., Barker, R.A., Meyer, K.B., Saeys, Y., Bonfanti, P., Behjati, S., Clatworthy, M.R., Taghon, T., Haniffa, M., Teichmann, S.A., 2020. A cell atlas of human thymic development defines T cell repertoire formation. Science 367, eaay3224. https://doi.org/10.1126/science.aay3224

Patrick, K., Wade, R., Goulden, N., Mitchell, C., Moorman, A.V., Rowntree, C., Jenkinson, S., Hough, R., Vora, A., 2014. Outcome for children and young people with E arly T -cell precursor acute lymphoblastic leukaemia treated on a contemporary

protocol, UKALL 2003. Br J Haematol 166, 421–424. https://doi.org/10.1111/bjh.12882

Pellicci, D.G., Koay, H.-F., Berzins, S.P., 2020. Thymic development of unconventional T cells: how NKT cells, MAIT cells and γδ T cells emerge. Nat Rev Immunol 20, 756–770. https://doi.org/10.1038/s41577-020-0345-y

Penter, L., Borji, M., Nagler, A., Lyu, H., Lu, W.S., Cieri, N., Maurer, K., Oliveira, G., Al'Khafaji, A.M., Garimella, K.V., Li, S., Neuberg, D.S., Ritz, J., Soiffer, R.J., Garcia, J.S., Livak, K.J., Wu, C.J., 2024. Integrative genotyping of cancer and immune phenotypes by long-read sequencing. Nat Commun 15, 32. https://doi.org/10.1038/s41467-023-44137-7

Perriman, L., Tavakolinia, N., Jalali, S., Li, S., Hickey, P.F., Amann-Zalcenstein, D., Ho, W.W.H., Baldwin, T.M., Piers, A.T., Konstantinov, I.E., Anderson, J., Stanley, E.G., Licciardi, P.V., Kannourakis, G., Naik, S.H., Koay, H.-F., Mackay, L.K., Berzins, S.P., Pellicci, D.G., 2023. A three-stage developmental pathway for human Vγ9Vδ2 T cells within the postnatal thymus. Sci. Immunol. 8, eabo4365. https://doi.org/10.1126/sciimmunol.abo4365

Petit, A., Trinquand, A., Chevret, S., Ballerini, P., Cayuela, J.-M., Grardel, N., Touzart, A., Brethon, B., Lapillonne, H., Schmitt, C., Thouvenin, S., Michel, G., Preudhomme, C., Soulier, J., Landman-Parker, J., Leverger, G., Macintyre, E., Baruchel, A., Asnafi, V., 2018. Oncogenetic mutations combined with MRD improve outcome prediction in pediatric T-cell acute lymphoblastic leukemia. Blood 131, 289–300. https://doi.org/10.1182/blood-2017-04-778829

Phipson, B., Sim, C.B., Porrello, E.R., Hewitt, A.W., Powell, J., Oshlack, A., 2022. propeller: testing for differences in cell type proportions in single cell data. Bioinformatics 38, 4720–4726. https://doi.org/10.1093/bioinformatics/btac582

Pobezinsky, L.A., Etzensperger, R., Jeurling, S., Alag, A., Kadakia, T., McCaughtry, T.M., Kimura, M.Y., Sharrow, S.O., Guinter, T.I., Feigenbaum, L., Singer, A., 2015. Let-7 microRNAs target the lineage-specific transcription factor PLZF to regulate terminal NKT cell differentiation and effector function. Nat Immunol 16, 517–524. https://doi.org/10.1038/ni.3146

Pölönen, P., Di Giacomo, D., Seffernick, A.E., Elsayed, A., Kimura, S., Benini, F., Montefiori, L.E., Wood, B.L., Xu, J., Chen, C., Cheng, Z., Newman, H., Myers, J., Iacobucci, I., Li, E., Sussman, J., Hedges, D., Hui, Y., Diorio, C., Uppuluri, L., Frank, D., Fan, Y., Chang, Y., Meshinchi, S., Ries, R., Shraim, R., Li, A., Bernt, K.M., Devidas, M., Winter, S.S., Dunsmore, K.P., Inaba, H., Carroll, W.L., Ramirez, N.C., Phillips, A.H., Kriwacki, R.W., Yang, J.J., Vincent, T.L., Zhao, Y., Ghate, P.S., Wang, J., Reilly, C., Zhou, X., Sanders, M.A., Takita, J., Kato, M., Takasugi, N., Chang, B.H., Press, R.D., Loh, M., Rampersaud, E., Raetz, E., Hunger, S.P., Tan, K., Chang, T.-C., Wu, G., Pounds, S.B., Mullighan, C.G., Teachey, D.T., 2024. The genomic basis of childhood T-lineage acute lymphoblastic leukaemia. Nature 632, 1082–1091. https://doi.org/10.1038/s41586-024-07807-0

Price, A.E., Liang, H.-E., Sullivan, B.M., Reinhardt, R.L., Eisley, C.J., Erle, D.J., Locksley, R.M., 2010. Systemically dispersed innate IL-13–expressing cells in type 2 immunity. Proc. Natl. Acad. Sci. U.S.A. 107, 11489–11494. https://doi.org/10.1073/pnas.1003988107

Priestley, P., Baber, J., Lolkema, M.P., Steeghs, N., De Bruijn, E., Shale, C., Duyvesteyn, K., Haidari, S., Van Hoeck, A., Onstenk, W., Roepman, P., Voda, M., Bloemendal,

H.J., Tjan-Heijnen, V.C.G., Van Herpen, C.M.L., Labots, M., Witteveen, P.O., Smit, E.F., Sleijfer, S., Voest, E.E., Cuppen, E., 2019. Pan-cancer whole-genome analyses of metastatic solid tumours. Nature 575, 210–216. https://doi.org/10.1038/s41586-019-1689-y

Rabilloud, T., Potier, D., Pankaew, S., Nozais, M., Loosveld, M., Payet-Bornet, D., 2021. Single-cell profiling identifies pre-existing CD19-negative subclones in a B-ALL patient with CD19-negative relapse after CAR-T therapy. Nat Commun 12, 865. https://doi.org/10.1038/s41467-021-21168-6

Radtke, F., Wilson, A., Stark, G., Bauer, M., Van Meerwijk, J., MacDonald, H.R., Aguet, M., 1999. Deficient T Cell Fate Specification in Mice with an Induced Inactivation of Notch1. Immunity 10, 547–558. https://doi.org/10.1016/S1074-7613(00)80054-0

Raetz, E.A., Rebora, P., Conter, V., Schrappe, M., Devidas, M., Escherich, G., Imai, C., De Moerloose, B., Schmiegelow, K., Burns, M.A., Elitzur, S., Pieters, R., Attarbaschi, A., Yeoh, A., Pui, C.-H., Stary, J., Cario, G., Bodmer, N., Moorman, A.V., Buldini, B., Vora, A., Valsecchi, M.G., 2023. Outcome for Children and Young Adults With T-Cell ALL and Induction Failure in Contemporary Trials. Journal of Clinical Oncology 41, 5025–5034. https://doi.org/10.1200/JCO.23.00088

Rebuffet, L., Melsen, J.E., Escalière, B., Basurto-Lozada, D., Bhandoola, A., Björkström, N.K., Bryceson, Y.T., Castriconi, R., Cichocki, F., Colonna, M., Davis, D.M., Diefenbach, A., Ding, Y., Haniffa, M., Horowitz, A., Lanier, L.L., Malmberg, K.-J., Miller, J.S., Moretta, L., Narni-Mancinelli, E., O'Neill, L.A.J., Romagnani, C., Ryan, D.G., Sivori, S., Sun, D., Vagne, C., Vivier, E., 2024. High-dimensional single-cell analysis of human natural killer cell heterogeneity. Nat Immunol 25, 1474–1488. https://doi.org/10.1038/s41590-024-01883-0

Richter-Pechańska, P., Kunz, J.B., Rausch, T., Erarslan-Uysal, B., Bornhauser, B., Frismantas, V., Assenov, Y., Zimmermann, M., Happich, M., Von Knebel-Doeberitz, C., Von Neuhoff, N., Köhler, R., Stanulla, M., Schrappe, M., Cario, G., Escherich, G., Kirschner-Schwabe, R., Eckert, C., Avigad, S., Pfister, S.M., Muckenthaler, M.U., Bourquin, J.-P., Korbel, J.O., Kulozik, A.E., 2022. Pediatric T-ALL type-1 and type-2 relapses develop along distinct pathways of clonal evolution. Leukemia 36, 1759–1768. https://doi.org/10.1038/s41375-022-01587-0

Roberts, K.G., Mullighan, C.G., 2020. The Biology of B-Progenitor Acute Lymphoblastic Leukemia. Cold Spring Harb Perspect Med 10, a034835. https://doi.org/10.1101/cshperspect.a034835

Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., Mesirov, J.P., 2011. Integrative genomics viewer. Nature Biotechnology 29, 24–26. https://doi.org/10.1038/nbt.1754

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, 139–140. https://doi.org/10.1093/bioinformatics/btp616

Royer-Pokora, B., Loos, U., Ludwig, W.D., 1991. TTG-2, a new gene encoding a cysteine-rich protein with the LIM motif, is overexpressed in acute T-cell leukaemia with the t(11;14)(p13;q11). Oncogene 6, 1887–1893.

Rubnitz, J.E., Behm, F.G., Curcio-Brint, A.M., Pinheiro, V.R.P., Carroll, A.J., Raimondi, S.C., Shurtleff, S.A., Downing, J.R., 1996. Molecular Analysis of t(11;19)

Breakpoints in Childhood Acute Leukemias. Blood 87, 4804–4808. https://doi.org/10.1182/blood.V87.11.4804.bloodjournal87114804

Ruscher, R., Kummer, R.L., Lee, Y.J., Jameson, S.C., Hogquist, K.A., 2017. CD8αα intraepithelial lymphocytes arise from two main thymic precursors. Nat Immunol 18, 771–779. https://doi.org/10.1038/ni.3751

Sánchez-Martínez, D., Baroni, M.L., Gutierrez-Agüera, F., Roca-Ho, H., Blanch-Lombarte, O., González-García, S., Torrebadell, M., Junca, J., Ramírez-Orellana, M., Velasco-Hernández, T., Bueno, C., Fuster, J.L., Prado, J.G., Calvo, J., Uzan, B., Cools, J., Camos, M., Pflumio, F., Toribio, M.L., Menéndez, P., 2019. Fratricide-resistant CD1a-specific CAR T cells for the treatment of cortical T-cell acute lymphoblastic leukemia. Blood 133, 2291–2304. https://doi.org/10.1182/blood-2018-10-882944

Satoh-Takayama, N., Vosshenrich, C.A.J., Lesjean-Pottier, S., Sawa, S., Lochner, M., Rattis, F., Mention, J.-J., Thiam, K., Cerf-Bensussan, N., Mandelboim, O., Eberl, G., Di Santo, J.P., 2008. Microbial Flora Drives Interleukin 22 Production in Intestinal NKp46+ Cells that Provide Innate Mucosal Immune Defense. Immunity 29, 958–970. https://doi.org/10.1016/j.immuni.2008.11.001

Savage, A.K., Constantinides, M.G., Han, J., Picard, D., Martin, E., Li, B., Lantz, O., Bendelac, A., 2008. The Transcription Factor PLZF Directs the Effector Program of the NKT Cell Lineage. Immunity 29, 391–403. https://doi.org/10.1016/j.immuni.2008.07.011

Schmitt, T.M., Zúñiga-Pflücker, J.C., 2002. Induction of T Cell Development from Hematopoietic Progenitor Cells by Delta-like-1 In Vitro. Immunity 17, 749–756. https://doi.org/10.1016/S1074-7613(02)00474-0

Seach, N., Guerri, L., Le Bourhis, L., Mburu, Y., Cui, Y., Bessoles, S., Soudais, C., Lantz, O., 2013. Double Positive Thymocytes Select Mucosal-Associated Invariant T Cells. The Journal of Immunology 191, 6002–6009. https://doi.org/10.4049/jimmunol.1301212

Seki, M., Kimura, S., Isobe, T., Yoshida, K., Ueno, H., Nakajima-Takagi, Y., Wang, C., Lin, L., Kon, A., Suzuki, H., Shiozawa, Y., Kataoka, K., Fujii, Y., Shiraishi, Y., Chiba, K., Tanaka, H., Shimamura, T., Masuda, K., Kawamoto, H., Ohki, K., Kato, M., Arakawa, Y., Koh, K., Hanada, R., Moritake, H., Akiyama, M., Kobayashi, R., Deguchi, T., Hashii, Y., Imamura, T., Sato, A., Kiyokawa, N., Oka, A., Hayashi, Y., Takagi, M., Manabe, A., Ohara, A., Horibe, K., Sanada, M., Iwama, A., Mano, H., Miyano, S., Ogawa, S., Takita, J., 2017. Recurrent SPI1 (PU.1) fusions in high-risk pediatric T cell acute lymphoblastic leukemia. Nat Genet 49, 1274–1281. https://doi.org/10.1038/ng.3900

Setty, M., Kiseliovas, V., Levine, J., Gayoso, A., Mazutis, L., Pe'er, D., 2019. Characterization of cell fate probabilities in single-cell data with Palantir. Nat Biotechnol 37, 451–460. https://doi.org/10.1038/s41587-019-0068-4

Shah, N., Sukumar, S., 2010. The Hox genes and their roles in oncogenesis. Nat Rev Cancer 10, 361–371. https://doi.org/10.1038/nrc2826

Shields, A.M., Patel, S.Y., 2017. The primary immunodeficiency disorders. Medicine 45, 597–604. https://doi.org/10.1016/j.mpmed.2017.07.011

Song, L., Cohen, D., Ouyang, Z., Cao, Y., Hu, X., Liu, X.S., 2021. TRUST4: immune repertoire reconstruction from bulk and single-cell RNA-seq data. Nat Methods 18, 627–630. https://doi.org/10.1038/s41592-021-01142-2

Soulier, J., Clappier, E., Cayuela, J.-M., Regnault, A., García-Peydró, M., Dombret, H., Baruchel, A., Toribio, M.-L., Sigaux, F., 2005. HOXA genes are included in genetic and biologic networks defining human acute T-cell leukemia (T-ALL). Blood 106, 274–286. https://doi.org/10.1182/blood-2004-10-3900

Speleman, F., Cauwelier, B., Dastugue, N., Cools, J., Verhasselt, B., Poppe, B., Van Roy, N., Vandesompele, J., Graux, C., Uyttebroeck, A., Boogaerts, M., De Moerloose, B., Benoit, Y., Selleslag, D., Billiet, J., Robert, A., Huguet, F., Vandenberghe, P., De Paepe, A., Marynen, P., Hagemeijer, A., 2005. A new recurrent inversion, inv(7)(p15q34), leads to transcriptional activation of HOXA10 and HOXA11 in a subset of T-cell acute lymphoblastic leukemias. Leukemia 19, 358–366. https://doi.org/10.1038/sj.leu.2403657

Spits, H., Artis, D., Colonna, M., Diefenbach, A., Di Santo, J.P., Eberl, G., Koyasu, S., Locksley, R.M., McKenzie, A.N.J., Mebius, R.E., Powrie, F., Vivier, E., 2013. Innate lymphoid cells — a proposal for uniform nomenclature. Nat Rev Immunol 13, 145–149. https://doi.org/10.1038/nri3365

Street, K., Risso, D., Fletcher, R.B., Das, D., Ngai, J., Yosef, N., Purdom, E., Dudoit, S., 2018. Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. BMC Genomics 19, 477. https://doi.org/10.1186/s12864-018-4772-0

Stritesky, G.L., Jameson, S.C., Hogquist, K.A., 2012. Selection of Self-Reactive T Cells in the Thymus. Annu. Rev. Immunol. 30, 95–114. https://doi.org/10.1146/annurev-immunol-020711-075035

Sun, Z., Unutmaz, D., Zou, Y.-R., Sunshine, M.J., Pierani, A., Brenner-Morton, S., Mebius, R.E., Littman, D.R., 2000. Requirement for RORγ in Thymocyte Survival and Lymphoid Organ Development. Science 288, 2369–2373. https://doi.org/10.1126/science.288.5475.2369

Suo, C., Dann, E., Goh, I., Jardine, L., Kleshchevnikov, V., Park, J.-E., Botting, R.A., Stephenson, E., Engelbert, J., Tuong, Z.K., Polanski, K., Yayon, N., Xu, C., Suchanek, O., Elmentaite, R., Domínguez Conde, C., He, P., Pritchard, S., Miah, M., Moldovan, C., Steemers, A.S., Mazin, P., Prete, M., Horsfall, D., Marioni, J.C., Clatworthy, M.R., Haniffa, M., Teichmann, S.A., 2022. Mapping the developing human immune system across organs. Science 376, eabo0510. https://doi.org/10.1126/science.abo0510

Suo, C., Polanski, K., Dann, E., Lindeboom, R.G.H., Vilarrasa-Blasi, R., Vento-Tormo, R., Haniffa, M., Meyer, K.B., Dratva, L.M., Tuong, Z.K., Clatworthy, M.R., Teichmann, S.A., 2024. Dandelion uses the single-cell adaptive immune receptor repertoire to explore lymphocyte developmental origins. Nat Biotechnol 42, 40–51. https://doi.org/10.1038/s41587-023-01734-7

Szabo, S.J., Kim, S.T., Costa, G.L., Zhang, X., Fathman, C.G., Glimcher, L.H., 2000. A Novel Transcription Factor, T-bet, Directs Th1 Lineage Commitment. Cell 100, 655–669. https://doi.org/10.1016/S0092-8674(00)80702-3

Szczepański, T., van der Velden, V.H.J., Waanders, E., Kuiper, R.P., Van Vlierberghe, P., Gruhn, B., Eckert, C., Panzer-Grümayer, R., Basso, G., Cavé, H., Stadt, U. zur, Campana, D., Schrauder, A., Sutton, R., van Wering, E., Meijerink, J.P.P., van Dongen, J.J.M., 2011. Late Recurrence of Childhood T-Cell Acute Lymphoblastic Leukemia Frequently Represents a Second Leukemia Rather Than a Relapse: First

Evidence for Genetic Predisposition. JCO 29, 1643–1649. https://doi.org/10.1200/JCO.2010.30.2877

Taghon, T., Thys, K., De Smedt, M., Weerkamp, F., Staal, F.J.T., Plum, J., Leclercq, G., 2003. Homeobox gene expression profile in human hematopoietic multipotent stem cells and T-cell progenitors: implications for human T-cell development. Leukemia 17, 1157–1163. https://doi.org/10.1038/sj.leu.2402947

Taj, M.M., Moorman, A.V., Hamadeh, L., Petit, A., Asnafi, V., Alby-Laurent, F., Vora, A., Mansour, M.R., Gale, R., Chevret, S., Moppett, J., Baruchel, A., Macintyre, E., 2022. Prognostic value of Oncogenetic mutations in pediatric T Acute Lymphoblastic Leukemia: a comparison of UKALL2003 and FRALLE2000T protocols. Leukemia 36, 263–266. https://doi.org/10.1038/s41375-021-01334-x

Tarabichi, M., Salcedo, A., Deshwar, A.G., Ni Leathlobhair, M., Wintersinger, J., Wedge, D.C., Van Loo, P., Morris, Q.D., Boutros, P.C., 2021. A practical guide to cancer subclonal reconstruction from DNA sequencing. Nat Methods 18, 144–155. https://doi.org/10.1038/s41592-020-01013-2

Teachey, D.T., O'Connor, D., 2020. How I treat newly diagnosed T-cell acute lymphoblastic leukemia and T-cell lymphoblastic lymphoma in children. Blood 135, 159–166. https://doi.org/10.1182/blood.2019001557

The Gene Ontology Consortium, Aleksander, S.A., Balhoff, J., Carbon, S., Cherry, J.M., Drabkin, H.J., Ebert, D., Feuermann, M., Gaudet, P., Harris, N.L., Hill, D.P., Lee, R., Mi, H., Moxon, S., Mungall, C.J., Muruganugan, A., Mushayahama, T., Sternberg, P.W., Thomas, P.D., Van Auken, K., Ramsey, J., Siegele, D.A., Chisholm, R.L., Fey, P., Aspromonte, M.C., Nugnes, M.V., Quaglia, F., Tosatto, S., Giglio, M., Nadendla, S., Antonazzo, G., Attrill, H., Dos Santos, G., Marygold, S., Strelets, V., Tabone, C.J., Thurmond, J., Zhou, P., Ahmed, S.H., Asanitthong, P., Luna Buitrago, D., Erdol, M.N., Gage, M.C., Ali Kadhum, M., Li, K.Y.C., Long, M., Michalak, A., Pesala, A., Pritazahra, A., Saverimuttu, S.C.C., Su, R., Thurlow, K.E., Lovering, R.C., Logie, C., Oliferenko, S., Blake, J., Christie, K., Corbani, L., Dolan, M.E., Drabkin, H.J., Hill, D.P., Ni, L., Sitnikov, D., Smith, C., Cuzick, A., Seager, J., Cooper, L., Elser, J., Jaiswal, P., Gupta, P., Jaiswal, P., Naithani, S., Lera-Ramirez, M., Rutherford, K., Wood, V., De Pons, J.L., Dwinell, M.R., Hayman, G.T., Kaldunski, M.L., Kwitek, A.E., Laulederkind, S.J.F., Tutaj, M.A., Vedi, M., Wang, S.-J., D'Eustachio, P., Aimo, L., Axelsen, K., Bridge, A., Hyka-Nouspikel, N., Morgat, A., Aleksander, S.A., Cherry, J.M., Engel, S.R., Karra, K., Miyasato, S.R., Nash, R.S., Skrzypek, M.S., Weng, S., Wong, E.D., Bakker, E., Berardini, T.Z., Reiser, L., Auchincloss, A., Axelsen, K., Argoud-Puy, G., Blatter, M.-C., Boutet, E., Breuza, L., Bridge, A., Casals-Casas, C., Coudert, E., Estreicher, A., Livia Famiglietti, M., Feuermann, M., Gos, A., Gruaz-Gumowski, N., Hulo, C., Hyka-Nouspikel, N., Jungo, F., Le Mercier, P., Lieberherr, D., Masson, P., Morgat, A., Pedruzzi, I., Pourcel, L., Poux, S., Rivoire, C., Sundaram, S., Bateman, A., Bowler-Barnett, E., Bye-A-Jee, H., Denny, P., Ignatchenko, A., Ishtiaq, R., Lock, A., Lussi, Y., Magrane, M., Martin, M.J., Orchard, S., Raposo, P., Speretta, E., Tyagi, N., Warner, K., Zaru, R., Diehl, A.D., Lee, R., Chan, J., Diamantakis, S., Raciti, D., Zarowiecki, M., Fisher, M., James-Zorn, C., Ponferrada, V., Zorn, A., Ramachandran, S., Ruzicka, L., Westerfield, M., Aleksander, S.A., Balhoff, J., Carbon, S., Cherry, J.M., Drabkin, H.J., Ebert, D., Feuermann, M., Gaudet, P., Harris, N.L., Hill, D.P., Lee, R., Mi, H., Moxon, S., Mungall, C.J., Muruganugan, A., Mushayahama, T., Sternberg, P.W.,

Thomas, P.D., Van Auken, K., Ramsey, J., Siegele, D.A., Chisholm, R.L., Fey, P., Aspromonte, M.C., Nugnes, M.V., Quaglia, F., Tosatto, S., Giglio, M., Nadendla, S., Antonazzo, G., Attrill, H., Dos Santos, G., Marygold, S., Strelets, V., Tabone, C.J., Thurmond, J., Zhou, P., Ahmed, S.H., Asanitthong, P., Luna Buitrago, D., Erdol, M.N., Gage, M.C., Ali Kadhum, M., Li, K.Y.C., Long, M., Michalak, A., Pesala, A., Pritazahra, A., Saverimuttu, S.C.C., Su, R., Thurlow, K.E., Lovering, R.C., Logie, C., Oliferenko, S., Blake, J., Christie, K., Corbani, L., Dolan, M.E., Drabkin, H.J., Hill, D.P., Ni, L., Sitnikov, D., Smith, C., Cuzick, A., Seager, J., Cooper, L., Elser, J., Jaiswal, P., Gupta, P., Jaiswal, P., Naithani, S., Lera-Ramirez, M., Rutherford, K., Wood, V., De Pons, J.L., Dwinell, M.R., Hayman, G.T., Kaldunski, M.L., Kwitek, A.E., Laulederkind, S.J.F., Tutaj, M.A., Vedi, M., Wang, S.-J., D'Eustachio, P., Aimo, L., Axelsen, K., Bridge, A., Hyka-Nouspikel, N., Morgat, A., Aleksander, S.A., Cherry, J.M., Engel, S.R., Karra, K., Miyasato, S.R., Nash, R.S., Skrzypek, M.S., Weng, S., Wong, E.D., Bakker, E., Berardini, T.Z., Reiser, L., Auchincloss, A., Axelsen, K., Argoud-Puy, G., Blatter, M.-C., Boutet, E., Breuza, L., Bridge, A., Casals-Casas, C., Coudert, E., Estreicher, A., Livia Famiglietti, M., Feuermann, M., Gos, A., Gruaz-Gumowski, N., Hulo, C., Hyka-Nouspikel, N., Jungo, F., Le Mercier, P., Lieberherr, D., Masson, P., Morgat, A., Pedruzzi, I., Pourcel, L., Poux, S., Rivoire, C., Sundaram, S., Bateman, A., Bowler-Barnett, E., Bye-A-Jee, H., Denny, P., Ignatchenko, A., Ishtiaq, R., Lock, A., Lussi, Y., Magrane, M., Martin, M.J., Orchard, S., Raposo, P., Speretta, E., Tyagi, N., Warner, K., Zaru, R., Diehl, A.D., Lee, R., Chan, J., Diamantakis, S., Raciti, D., Zarowiecki, M., Fisher, M., James-Zorn, C., Ponferrada, V., Zorn, A., Ramachandran, S., Ruzicka, L., Westerfield, M., 2023. The Gene Ontology knowledgebase in 2023. GENETICS 224, iyad031. https://doi.org/10.1093/genetics/iyad031

Therneau, T.M., 2024. A Package for Survival Analysis in R.

Theunis, K., Vanuytven, S., Claes, I., Geurts, J., Rambow, F., Brown, D., Van Der Haegen, M., Marin-Bejar, O., Rogiers, A., Van Raemdonck, N., Leucci, E., Demeulemeester, J., Sifrim, A., Marine, J.-C., Voet, T., 2025. Single-cell genome and transcriptome sequencing without upfront whole-genome amplification reveals cell state plasticity of melanoma subclones. Nucleic Acids Research 53, gkaf173. https://doi.org/10.1093/nar/gkaf173

Thompson, B.J., Buonamici, S., Sulis, M.L., Palomero, T., Vilimas, T., Basso, G., Ferrando, A., Aifantis, I., 2007. The SCFFBW7 ubiquitin ligase complex as a tumor suppressor in T cell leukemia. Journal of Experimental Medicine 204, 1825–1835. https://doi.org/10.1084/jem.20070872

Thomson, Z., He, Z., Swanson, E., Henderson, K., Phalen, C., Zaim, S.R., Pebworth, M.-P., Okada, L.Y., Heubeck, A.T., Roll, C.R., Hernandez, V., Weiss, M., Genge, P.C., Reading, J., Giles, J.R., Manne, S., Dougherty, J., Jasen, C.J., Greenplate, A.R., Becker, L.A., Graybuck, L.T., Vasaikar, S.V., Szeto, G.L., Savage, A.K., Speake, C., Buckner, J.H., Li, X., Bumol, T.F., Wherry, E.J., Torgerson, T.R., Vella, L.A., Henrickson, S.E., Skene, P.J., Gustafson, C.E., 2023. Trimodal single-cell profiling reveals a novel pediatric CD8αα+ T cell subset and broad age-related molecular reprogramming across the T cell compartment. Nat Immunol 24, 1947–1959. https://doi.org/10.1038/s41590-023-01641-8

Tilloy, F., Treiner, E., Park, S.H., Garcia, C., Lemonnier, F., de la Salle, H., Bendelac, A., Bonneville, M., Lantz, O., 1999. An invariant T cell receptor alpha chain defines a

novel TAP-independent major histocompatibility complex class Ib-restricted alpha/beta T cell subpopulation in mammals. J Exp Med 189, 1907–1921. https://doi.org/10.1084/jem.189.12.1907

Tirosh, I., Izar, B., Prakadan, S.M., Wadsworth, M.H., Treacy, D., Trombetta, J.J., Rotem, A., Rodman, C., Lian, C., Murphy, G., Fallahi-Sichani, M., Dutton-Regester, K., Lin, J.-R., Cohen, O., Shah, P., Lu, D., Genshaft, A.S., Hughes, T.K., Ziegler, C.G.K., Kazer, S.W., Gaillard, A., Kolb, K.E., Villani, A.-C., Johannessen, C.M., Andreev, A.Y., Van Allen, E.M., Bertagnolli, M., Sorger, P.K., Sullivan, R.J., Flaherty, K.T., Frederick, D.T., Jané-Valbuena, J., Yoon, C.H., Rozenblatt-Rosen, O., Shalek, A.K., Regev, A., Garraway, L.A., 2016. Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. Science 352, 189–196. https://doi.org/10.1126/science.aad0501

Touzart, A., Mayakonda, A., Smith, C., Hey, J., Toth, R., Cieslak, A., Andrieu, G.P., Tran Quang, C., Latiri, M., Ghysdael, J., Spicuglia, S., Dombret, H., Ifrah, N., Macintyre, E., Lutsik, P., Boissel, N., Plass, C., Asnafi, V., 2021. Epigenetic analysis of patients with T-ALL identifies poor outcomes and a hypomethylating agent-responsive subgroup. Sci. Transl. Med. 13, eabc4834. https://doi.org/10.1126/scitranslmed.abc4834

Treiner, E., Duban, L., Bahram, S., Radosavljevic, M., Wanner, V., Tilloy, F., Affaticati, P., Gilfillan, S., Lantz, O., 2003. Selection of evolutionarily conserved mucosal-associated invariant T cells by MR1. Nature 422, 164–169. https://doi.org/10.1038/nature01433

Trinh, M.K., Pacyna, C.N., Kildisiute, G., Thevanesan, C., Piapi, A., Ambridge, K., Anderson, N.D., Khabirova, E., Prigmore, E., Straathof, K., Behjati, S., Young, M.D., 2022. Precise identification of cancer cells from allelic imbalances in single cell transcriptomes. Commun Biol 5, 884. https://doi.org/10.1038/s42003-022-03808-9

Tzoneva, G., Perez-Garcia, A., Carpenter, Z., Khiabanian, H., Tosello, V., Allegretta, M., Paietta, E., Racevskis, J., Rowe, J.M., Tallman, M.S., Paganin, M., Basso, G., Hof, J., Kirschner-Schwabe, R., Palomero, T., Rabadan, R., Ferrando, A., 2013. Activating mutations in the NT5C2 nucleotidase gene drive chemotherapy resistance in relapsed ALL. Nat Med 19, 368–371. https://doi.org/10.1038/nm.3078

Ueno, H., Yoshida, K., Shiozawa, Y., Nannya, Y., Iijima-Yamashita, Y., Kiyokawa, N., Shiraishi, Y., Chiba, K., Tanaka, H., Isobe, T., Seki, M., Kimura, S., Makishima, H., Nakagawa, M.M., Kakiuchi, N., Kataoka, K., Yoshizato, T., Nishijima, D., Deguchi, T., Ohki, K., Sato, A., Takahashi, H., Hashii, Y., Tokimasa, S., Hara, J., Kosaka, Y., Kato, K., Inukai, T., Takita, J., Imamura, T., Miyano, S., Manabe, A., Horibe, K., Ogawa, S., Sanada, M., 2020. Landscape of driver mutations and their clinical impacts in pediatric B-cell precursor acute lymphoblastic leukemia. Blood Advances 4, 5165–5173. https://doi.org/10.1182/bloodadvances.2019001307

Uldrich, A.P., Le Nours, J., Pellicci, D.G., Gherardin, N.A., McPherson, K.G., Lim, R.T., Patel, O., Beddoe, T., Gras, S., Rossjohn, J., Godfrey, D.I., 2013. CD1d-lipid antigen recognition by the γδ TCR. Nat Immunol 14, 1137–1145. https://doi.org/10.1038/ni.2713

Ussher, J.E., Klenerman, P., Willberg, C.B., 2014. Mucosal-Associated Invariant T-Cells: New Players in Anti-Bacterial Immunity. Front. Immunol. 5. https://doi.org/10.3389/fimmu.2014.00450

Van Loo, P., Nordgard, S.H., Lingjærde, O.C., Russnes, H.G., Rye, I.H., Sun, W., Weigman, V.J., Marynen, P., Zetterberg, A., Naume, B., Perou, C.M., Børresen-Dale, A.-L., Kristensen, V.N., 2010. Allele-specific copy number analysis of tumors. Proc. Natl. Acad. Sci. U.S.A. 107, 16910–16915. https://doi.org/10.1073/pnas.1009843107

Van Vlierberghe, P., Palomero, T., Khiabanian, H., Van Der Meulen, J., Castillo, M., Van Roy, N., De Moerloose, B., Philippé, J., González-García, S., Toribio, M.L., Taghon, T., Zuurbier, L., Cauwelier, B., Harrison, C.J., Schwab, C., Pisecker, M., Strehl, S., Langerak, A.W., Gecz, J., Sonneveld, E., Pieters, R., Paietta, E., Rowe, J.M., Wiernik, P.H., Benoit, Y., Soulier, J., Poppe, B., Yao, X., Cordon-Cardo, C., Meijerink, J., Rabadan, R., Speleman, F., Ferrando, A., 2010. PHF6 mutations in T-cell acute lymphoblastic leukemia. Nat Genet 42, 338–342. https://doi.org/10.1038/ng.542

Van Vlierberghe, P., van Grotel, M., Beverloo, H.B., Lee, C., Helgason, T., Buijs-Gladdines, J., Passier, M., van Wering, E.R., Veerman, A.J.P., Kamps, W.A., Meijerink, J.P.P., Pieters, R., 2006. The cryptic chromosomal deletion del(11)(p12p13) as a new activation mechanism of LMO2 in pediatric T-cell acute lymphoblastic leukemia. Blood 108, 3520–3529. https://doi.org/10.1182/blood-2006-04-019927

Vavassori, S., Kumar, A., Wan, G.S., Ramanjaneyulu, G.S., Cavallari, M., El Daker, S., Beddoe, T., Theodossis, A., Williams, N.K., Gostick, E., Price, D.A., Soudamini, D.U., Voon, K.K., Olivo, M., Rossjohn, J., Mori, L., De Libero, G., 2013. Butyrophilin 3A1 binds phosphorylated antigens and stimulates human γδ T cells. Nat Immunol 14, 908–916. https://doi.org/10.1038/ni.2665

Veldhoen, M., Hocking, R.J., Atkins, C.J., Locksley, R.M., Stockinger, B., 2006. TGFβ in the Context of an Inflammatory Cytokine Milieu Supports De Novo Differentiation of IL-17-Producing T Cells. Immunity 24, 179–189. https://doi.org/10.1016/j.immuni.2006.01.001

Verstichel, G., Vermijlen, D., Martens, L., Goetgeluk, G., Brouwer, M., Thiault, N., Caeneghem, Y.V., Munter, S.D., Weening, K., Bonte, S., Leclercq, G., Taghon, T., Kerre, T., Saeys, Y., Dorpe, J.V., Cheroutre, H., Vandekerckhove, B., 2017. The checkpoint for agonist selection precedes conventional selection in human thymus. SCIENCE IMMUNOLOGY.

Villani, A.-C., Satija, R., Reynolds, G., Sarkizova, S., Shekhar, K., Fletcher, J., Griesbeck, M., Butler, A., Zheng, S., Lazo, S., Jardine, L., Dixon, D., Stephenson, E., Nilsson, E., Grundberg, I., McDonald, D., Filby, A., Li, W., De Jager, P.L., Rozenblatt-Rosen, O., Lane, A.A., Haniffa, M., Regev, A., Hacohen, N., 2017. Single-cell RNA-seq reveals new types of human blood dendritic cells, monocytes, and progenitors. Science 356, eaah4573. https://doi.org/10.1126/science.aah4573

Vivier, E., Artis, D., Colonna, M., Diefenbach, A., Di Santo, J.P., Eberl, G., Koyasu, S., Locksley, R.M., McKenzie, A.N.J., Mebius, R.E., Powrie, F., Spits, H., 2018. Innate Lymphoid Cells: 10 Years On. Cell 174, 1054–1066. https://doi.org/10.1016/j.cell.2018.07.017

Waanders, E., Gu, Z., Dobson, S.M., Antić, Ž., Crawford, J.C., Ma, X., Edmonson, M.N., Payne-Turner, D., van de Vorst, M., Jongmans, M.C.J., McGuire, I., Zhou, X., Wang, J., Shi, L., Pounds, S., Pei, D., Cheng, C., Song, G., Fan, Y., Shao, Y., Rusch, M., McCastlain, K., Yu, J., van Boxtel, R., Blokzijl, F., Iacobucci, I., Roberts, K.G., Wen, J., Wu, G., Ma, J., Easton, J., Neale, G., Olsen, S.R., Nichols, K.E., Pui, C.-H., Zhang, J., Evans, W.E., Relling, M.V., Yang, J.J., Thomas, P.G., Dick, J.E., Kuiper, R.P., Mullighan, C.G., 2020. Mutational Landscape and Patterns of Clonal Evolution in Relapsed Pediatric Acute Lymphoblastic Leukemia. Blood Cancer Discovery 1, 96–111. https://doi.org/10.1158/0008-5472.BCD-19-0041

Wang, J., Jani-Sait, S.N., Escalon, E.A., Carroll, A.J., de Jong, P.J., Kirsch, I.R., Aplan, P.D., 2000. The t(14;21)(q11.2;q22) chromosomal translocation associated with T-cell acute lymphoblastic leukemia activates the BHLHB1 gene. Proc Natl Acad Sci U S A 97, 3497–3502. https://doi.org/10.1073/pnas.97.7.3497

Wang, K., Guo, D., Yan, T., Sun, S., Wang, Y., Zheng, H., Wang, G., Du, J., 2024. ZBTB16 inhibits DNA replication and induces cell cycle arrest by targeting WDHD1 transcription in lung adenocarcinoma. Oncogene 43, 1796–1810. https://doi.org/10.1038/s41388-024-03041-0

Wang, K., Ye, R., Bai, S., Xiao, Z., Yang, L., Li, J., Tang, C., Sei, E., Peng, J., Casasent, A.K., Lin, S.H., Nagi, C., Thompson, A.M., Krishnamurthy, S., Navin, N.E., 2025. Coalescing single-cell genomes and transcriptomes to decode breast cancer progression. Cell. https://doi.org/10.1016/j.cell.2025.08.012

Wang, X., Chen, Y., Li, Z., Huang, B., Xu, L., Lai, J., Lu, Y., Zha, X., Liu, B., Lan, Y., Li, Y., 2021. Single-Cell RNA-Seq of T Cells in B-ALL Patients Reveals an Exhausted Subset with Remarkable Heterogeneity. Advanced Science 8, 2101447. https://doi.org/10.1002/advs.202101447

Warner, N., Szenberg, A., Burnet, F., 1962. The Immunological Role of Different Lymphoid Organs in the Chicken. Australian Journal of Experimental Biology and Medical Science 40, 373–388. https://doi.org/10.1038/icb.1962.42

Wasim, M., Carlet, M., Mansha, M., Greil, R., Ploner, C., Trockenbacher, A., Rainer, J., Kofler, R., 2010. PLZF/ZBTB16, a glucocorticoid response gene in acute lymphoblastic leukemia, interferes with glucocorticoid-induced apoptosis. The Journal of Steroid Biochemistry and Molecular Biology 120, 218–227. https://doi.org/10.1016/j.jsbmb.2010.04.019

Webb, S., Haniffa, M., 2023. Large-scale single-cell RNA sequencing atlases of human immune cells across lifespan: Possibilities and challenges. Eur J Immunol 53, 2250222. https://doi.org/10.1002/eji.202250222

Weber, B.N., Chi, A.W.-S., Chavez, A., Yashiro-Ohtani, Y., Yang, Q., Shestova, O., Bhandoola, A., 2011. A critical role for TCF-1 in T-lineage specification and differentiation. Nature 476, 63–68. https://doi.org/10.1038/nature10279

Weizman, O.-E., Adams, N.M., Schuster, I.S., Krishna, C., Pritykin, Y., Lau, C., Degli-Esposti, M.A., Leslie, C.S., Sun, J.C., O'Sullivan, T.E., 2017. ILC1 Confer Early Host Protection at Initial Sites of Viral Infection. Cell 171, 795-808.e12. https://doi.org/10.1016/j.cell.2017.09.052

Weng, A.P., Ferrando, A.A., Lee, W., Morris, J.P., Silverman, L.B., Sanchez-Irizarry, C., Blacklow, S.C., Look, A.T., Aster, J.C., 2004. Activating Mutations of *NOTCH1* in Human T Cell Acute Lymphoblastic Leukemia. Science 306, 269–271. https://doi.org/10.1126/science.1102160

Weng, A.P., Millholland, J.M., Yashiro-Ohtani, Y., Arcangeli, M.L., Lau, A., Wai, C., Del Bianco, C., Rodriguez, C.G., Sai, H., Tobias, J., Li, Y., Wolfe, M.S., Shachaf, C., Felsher, D., Blacklow, S.C., Pear, W.S., Aster, J.C., 2006. *c-Myc* is an important direct target of Notch1 in T-cell acute lymphoblastic leukemia/lymphoma. Genes Dev. 20, 2096–2109. https://doi.org/10.1101/gad.1450406

Weng, C., Yu, F., Yang, D., Poeschla, M., Liggett, L.A., Jones, M.G., Qiu, X., Wahlster, L., Caulier, A., Hussmann, J.A., Schnell, A., Yost, K.E., Koblan, L.W., Martin-Rufino, J.D., Min, J., Hammond, A., Ssozi, D., Bueno, R., Mallidi, H., Kreso, A., Escabi, J., Rideout, W.M., Jacks, T., Hormoz, S., van Galen, P., Weissman, J.S., Sankaran, V.G., 2024. Deciphering cell states and genealogies of human haematopoiesis. Nature 627, 389–398. https://doi.org/10.1038/s41586-024-07066-z

Winter, G.E., Buckley, D.L., Paulk, J., Roberts, J.M., Souza, A., Dhe-Paganon, S., Bradner, J.E., 2015. Phthalimide conjugation as a strategy for in vivo target protein degradation. Science 348, 1376–1381. https://doi.org/10.1126/science.aab1433

Witkowski, M.T., Dolgalev, I., Evensen, N.A., Ma, C., Chambers, T., Roberts, K.G., Sreeram, S., Dai, Y., Tikhonova, A.N., Lasry, A., Qu, C., Pei, D., Cheng, C., Robbins, G.A., Pierro, J., Selvaraj, S., Mezzano, V., Daves, M., Lupo, P.J., Scheurer, M.E., Loomis, C.A., Mullighan, C.G., Chen, W., Rabin, K.R., Tsirigos, A., Carroll, W.L., Aifantis, I., 2020. Extensive Remodeling of the Immune Microenvironment in B Cell Acute Lymphoblastic Leukemia. Cancer Cell 37, 867-882.e12. https://doi.org/10.1016/j.ccell.2020.04.015

Wolf, F.A., Angerer, P., Theis, F.J., 2018. SCANPY: large-scale single-cell gene expression data analysis. Genome Biol 19, 15. https://doi.org/10.1186/s13059-017-1382-0

Wolf, F.A., Hamey, F.K., Plass, M., Solana, J., Dahlin, J.S., Göttgens, B., Rajewsky, N., Simon, L., Theis, F.J., 2019. PAGA: graph abstraction reconciles clustering with trajectory inference through a topology preserving map of single cells. Genome Biol 20, 59. https://doi.org/10.1186/s13059-019-1663-x

Wolf, N.K., Kissiov, D.U., Raulet, D.H., 2023. Roles of natural killer cells in immunity to cancer, and applications to immunotherapy. Nat Rev Immunol 23, 90–105. https://doi.org/10.1038/s41577-022-00732-1

Wolock, S.L., Lopez, R., Klein, A.M., 2019. Scrublet: Computational Identification of Cell Doublets in Single-Cell Transcriptomic Data. Cell Systems 8, 281-291.e9. https://doi.org/10.1016/j.cels.2018.11.005

Wong, S.H., Walker, J.A., Jolin, H.E., Drynan, L.F., Hams, E., Camelo, A., Barlow, J.L., Neill, D.R., Panova, V., Koch, U., Radtke, F., Hardman, C.S., Hwang, Y.Y., Fallon, P.G., McKenzie, A.N.J., 2012. Transcription factor RORα is critical for nuocyte development. Nat Immunol 13, 229–236. https://doi.org/10.1038/ni.2208

Wood, B.L., Devidas, M., Summers, R.J., Chen, Z., Asselin, B., Rabin, K.R., Zweidler-McKay, P.A., Winick, N.J., Borowitz, M.J., Carroll, W.L., Raetz, E.A., Loh, M.L., Hunger, S.P., Dunsmore, K.P., Teachey, D.T., Winter, S.S., 2023. Prognostic significance of ETP phenotype and minimal residual disease in T-ALL: a Children's Oncology Group study. Blood 142, 2069–2078. https://doi.org/10.1182/blood.2023020678

Xia, Y., Brown, L., Yang, C.Y., Tsan, J.T., Siciliano, M.J., Espinosa, R. 3rd, Le Beau, M.M., Baer, R.J., 1991. TAL2, a helix-loop-helix gene activated by the (7;9)(q34;q32) translocation in human T-cell leukemia. Proc Natl Acad Sci U S A 88, 11416–11420. https://doi.org/10.1073/pnas.88.24.11416

Xu, J., Chen, Changya, Sussman, J.H., Yoshimura, S., Vincent, T., Pölönen, P., Hu, J., Bandyopadhyay, S., Elghawy, O., Yu, W., Tumulty, J., Chen, Chia-hui, Li, E.Y., Diorio, C., Shraim, R., Newman, H., Uppuluri, L., Li, A., Chen, G.M., Wu, D.W., Ding, Y., Xu, J.A., Karanfilovski, D., Lim, T., Hsu, M., Thadi, A., Ahn, K.J., Wu, C.-Y., Peng, J., Sun, Y., Wang, A., Mehta, R., Frank, D., Meyer, L., Loh, M.L., Raetz, E.A., Chen, Z., Wood, B.L., Devidas, M., Dunsmore, K.P., Winter, S.S., Chang, T.-C., Wu, G., Pounds, S.B., Zhang, N.R., Carroll, W., Hunger, S.P., Bernt, K., Yang, J.J., Mullighan, C.G., Tan, K., Teachey, D.T., 2024. A multiomic atlas identifies a treatment-resistant, bone marrow progenitor-like cell population in T cell acute lymphoblastic leukemia. Nat Cancer 6, 102–122. https://doi.org/10.1038/s43018-024-00863-5

Yagi, R., Zhong, C., Northrup, D.L., Yu, F., Bouladoux, N., Spencer, S., Hu, G., Barron, L., Sharma, S., Nakayama, T., Belkaid, Y., Zhao, K., Zhu, J., 2014. The Transcription Factor GATA3 Is Critical for the Development of All IL-7Rα-Expressing Innate Lymphoid Cells. Immunity 40, 378–388. https://doi.org/10.1016/j.immuni.2014.01.012

Ye, K., Schulz, M.H., Long, Q., Apweiler, R., Ning, Z., 2009. Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. Bioinformatics 25, 2865–2871. https://doi.org/10.1093/bioinformatics/btp394

Yokota, Y., Mansouri, A., Mori, S., Sugawara, S., Adachi, S., Nishikawa, S.-I., Gruss, P., 1999. Development of peripheral lymphoid organs and natural killer cells depends on the helix–loop–helix inhibitor Id2. Nature 397, 702–706. https://doi.org/10.1038/17812

Yokoyama, A., Somervaille, T.C.P., Smith, K.S., Rozenblatt-Rosen, O., Meyerson, M., Cleary, M.L., 2005. The Menin Tumor Suppressor Protein Is an Essential Oncogenic Cofactor for MLL-Associated Leukemogenesis. Cell 123, 207–218. https://doi.org/10.1016/j.cell.2005.09.025

Young, M.D., Behjati, S., 2020. SoupX removes ambient RNA contamination from droplet-based single-cell RNA sequencing data. GigaScience 9, giaa151. https://doi.org/10.1093/gigascience/giaa151

Young, M.D., Mitchell, T.J., Vieira Braga, F.A., Tran, M.G.B., Stewart, B.J., Ferdinand, J.R., Collord, G., Botting, R.A., Popescu, D.-M., Loudon, K.W., Vento-Tormo, R., Stephenson, E., Cagan, A., Farndon, S.J., Del Castillo Velasco-Herrera, M., Guzzo, C., Richoz, N., Mamanova, L., Aho, T., Armitage, J.N., Riddick, A.C.P., Mushtaq, I., Farrell, S., Rampling, D., Nicholson, J., Filby, A., Burge, J., Lisgo, S., Maxwell, P.H., Lindsay, S., Warren, A.Y., Stewart, G.D., Sebire, N., Coleman, N., Haniffa, M., Teichmann, S.A., Clatworthy, M., Behjati, S., 2018. Single-cell transcriptomes from human kidneys reveal the cellular identity of renal tumors. Science 361, 594–599. https://doi.org/10.1126/science.aat1699

Zamora, A.E., Crawford, J.C., Allen, E.K., Guo, X.J., Bakke, J., Carter, R.A., Abdelsamed, H.A., Moustaki, A., Li, Y., Chang, T.-C., Awad, W., Dallas, M.H., Mullighan, C.G., Downing, J.R., Geiger, T.L., Chen, T., Green, D.R., Youngblood, B.A., Zhang, J., Thomas, P.G., 2019. Pediatric patients with acute lymphoblastic leukemia generate abundant and functional neoantigen-specific CD8 $^{+}$ T cell responses. Sci. Transl. Med. 11, eaat8549. https://doi.org/10.1126/scitranslmed.aat8549

Zhang, J., Ding, L., Holmfeldt, L., Wu, G., Heatley, S.L., Payne-Turner, D., Easton, J., Chen, X., Wang, J., Rusch, M., Lu, C., Chen, S.-C., Wei, L., Collins-Underwood, J.R., Ma, J., Roberts, K.G., Pounds, S.B., Ulyanov, A., Becksfort, J., Gupta, P., Huether, R., Kriwacki, R.W., Parker, M., McGoldrick, D.J., Zhao, D., Alford, D., Espy, S., Bobba, K.C., Song, G., Pei, D., Cheng, C., Roberts, S., Barbato, M.I., Campana, D., Coustan-Smith, E., Shurtleff, S.A., Raimondi, S.C., Kleppe, M., Cools, J., Shimano, K.A., Hermiston, M.L., Doulatov, S., Eppert, K., Laurenti, E., Notta, F., Dick, J.E., Basso, G., Hunger, S.P., Loh, M.L., Devidas, M., Wood, B., Winter, S., Dunsmore, K.P., Fulton, R.S., Fulton, L.L., Hong, X., Harris, C.C., Dooling, D.J., Ochoa, K., Johnson, K.J., Obenauer, J.C., Evans, W.E., Pui, C.-H., Naeve, C.W., Ley, T.J., Mardis, E.R., Wilson, R.K., Downing, J.R., Mullighan, C.G., 2012. The genetic basis of early T-cell precursor acute lymphoblastic leukaemia. Nature 481, 157–163. https://doi.org/10.1038/nature10725

Zheng, G.X.Y., Terry, J.M., Belgrader, P., Ryvkin, P., Bent, Z.W., Wilson, R., Ziraldo, S.B., Wheeler, T.D., McDermott, G.P., Zhu, J., Gregory, M.T., Shuga, J., Montesclaros, L., Underwood, J.G., Masquelier, D.A., Nishimura, S.Y., Schnall-Levin, M., Wyatt, P.W., Hindson, C.M., Bharadwaj, R., Wong, A., Ness, K.D., Beppu, L.W., Deeg, H.J., McFarland, C., Loeb, K.R., Valente, W.J., Ericson, N.G., Stevens, E.A., Radich, J.P., Mikkelsen, T.S., Hindson, B.J., Bielas, J.H., 2017. Massively parallel digital transcriptional profiling of single cells. Nat Commun 8, 14049. https://doi.org/10.1038/ncomms14049

Zheng, W., Flavell, R.A., 1997. The Transcription Factor GATA-3 Is Necessary and Sufficient for Th2 Cytokine Gene Expression in CD4 T Cells. Cell 89, 587–596. https://doi.org/10.1016/S0092-8674(00)80240-8