EXPLORING MECHANISMS OF T-CELL ACUTE LYMPHOBLASTIC LEUKEMIA PATHOGENESIS THROUGH MOLECULAR CHARACTERIZATION OF NORMAL T-CELL DEVELOPMENT

by

Alana Rachel Wong

B.Sc. (Honours and Co-op), University of Victoria, 2015

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES

(Interdisciplinary Oncology)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

September 2022

The following individuals certify that they have read, and recommend to the Faculty of Graduate and Postdoctoral Studies for acceptance, the dissertation entitled:

Exploring mechanisms of T-cell acute lymphoblastic leukemia pathogenesis through molecular characterization of normal T-cell development

submitted by Alana Rachel Wong in partial fulfillment of the requirements for

the degree of Doctor of Philosophy

in Interdisciplinary Oncology

Examining Committee:

Andrew Weng, Professor, Pathology & Laboratory Medicine, UBC
Supervisor

Gregg Morin, Professor, Medical Genetics, UBC
Supervisory Committee Member

Ninan Abraham, Professor, Department of Microbiology and Immunology, Department of Zoology, UBC
University Examiner

Fumio Takei, Professor, Pathology & Laboratory Medicine, UBC
University Examiner

Additional Supervisory Committee Members:

Christian Steidl, Professor, Pathology & Laboratory Medicine, UBC
Supervisory Committee Member

Abstract

T-cell development and lineage commitment are temporally protracted processes in which the interplay between transcription factors and epigenetic modulators orchestrate the sequential exclusion of alternative fates and acquisition of specialized T-cell functions. Alterations during this process can lead to diseases such as T-cell acute lymphoblastic leukemia (T-ALL). While much work has focused on the transcription factors which drive normal development and T-ALL, the underlying epigenetic constraints remain poorly characterized. The central objective of the work presented in this thesis is to gain a greater understanding of the molecular events during early human T-cell differentiation to allow for dissection of particular genetic events that may occur in T-ALL. This was addressed through a detailed transcriptomic characterization of T-ALL within the framework of normal development, and through genetic perturbation and functional characterization of *in vitro*-differentiating T-cell subsets. The findings revealed; 1) an improved classification of T-ALL which better captures the developmental context in which specific transcription factors operate and which improves upon the identification of clinically relevant disease subgroups, and 2) novel insights into a role of *DNMT3A* in restricting lineage-specific signal responses in T-cell development and T-ALL. The results of functional assays revealed that *DNMT3A* loss increased signaling plasticity in T-lineage-restricted populations, as measured by response to cytokines typically affiliated with the myeloid lineage. This was enriched within a population of lineage-restricted T-cells expressing the G-, M-, and/or GM-CSF receptors, and was associated with proliferation and acquisition of a myeloid-like phenotype. Further investigation of the expression of these receptors revealed their prevalence in human T-ALL cell lines and post-natal thymus, thus indicating the biological relevance of CSF receptor expression in human tissues. Stimulation of the CSF receptor-expressing population by adding cognate ligands revealed a proliferative advantage in T-ALL cells, but only upon *DNMT3A* loss. These results suggest that DNMT3A may act to preclude response to alternative-lineage factors in T-cells, and furthermore point to a mechanism of selective advantage in a subset of T-ALL.

Lay Summary

The development of T-cells normally occurs in a highly sequential manner, in which genes are turned on or off in response to external signaling cues in order to drive normal T-cell development. Failure to appropriately regulate these genes can result in diseases such as leukemia. To gain a better understanding of how aberrations in development may progress to disease, a comprehensive comparison of normal vs leukemic cells was performed, and the effect of altering genes in normal cells was interrogated using a model of normal T-cell development. The results of these findings revealed that alteration of a specific gene in normal cells can reprogram T-cells. In the context of leukemia, this can result in increased cell growth. The results presented here provide insight into normal cell development and novel mechanisms for disease pathogenesis.

Preface

The experiments outlined in this thesis were conceived and designed by Dr. Andrew Weng and me. I designed and performed all experiments, data analysis and generated the figures and tables that are presented in this thesis unless otherwise stated below.

A version of Chapter 2 has been published: **Wong R,** Nguyen A, Wang X, Chong L, Tyshchenko K, Brown SD, Holt RA, Steidl C, Weng AP. Improved resolution of phenotypic subsets in human T-ALL by incorporation of RNA-seq based developmental profiling. Leuk Res. 2021 Nov;110:106712. doi: 10.1016/j.leukres.2021.106712. Epub 2021 Sep 22. PMID: 34583126. I conceived and designed the study together with Dr. Andrew Weng. I performed all bioinformatics analyses and figure generation. RNA-seq alignment was performed by Kateryna Tyshchenko and Lauren Chong under the supervision of Dr. Christian Steidl. Dr. Xuehai Wang provided guidance for data analysis. MiXCR analysis was performed by Andrew Nguyen with assistance from Dr. Scott Brown and under the supervision of Dr. Robert Holt. I wrote the manuscript with Andrew Weng.

Chapters 3 and 4 are being prepared for publication. Andrew Weng and I conceived and designed all experiments with assistance from Dr. Samuel Gusscott. I performed all experiments, data analyses, and figure generation. Samuel Gusscott produced the western blot of DNMT3A protein levels, performed cloning of constructs, designed CRISPR guide RNAs, and assisted with sample collection for RNA-seq and whole genome bisulfite sequencing. Andrew Weng and I interpreted the data. The gemBS pipeline for quantifying CpG methylation levels, as well as data quality control, was performed by Reanne Bowlby and Anaaïck Carles. All cell sorting was performed in the Terry Fox Laboratory Flow Core with assistance from Wenbo Xu and Guillermo Simkin.

The research presented in this thesis was approved by UBC BC Cancer Research Ethics Board (H20-00710). Primary patient tissues were obtained with informed consent from donors.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

5caC – 5-carboxylcytosine

5mC – 5-methylcytosine

ADD – ATRX-DNMT3-DNMT3L

AML – Acute myeloid leukemia

BAH – bromo-adjacent homology

BM – bone marrow

CB – cord blood

CCL – chemokine ligand

CCR – chemokine receptor

CD – cluster of differentiation

CGI – CpG island

CMJ – cortico-medullary junction

CRISPR – Clustered regularly interspaced short palindromic repeats

CSF – colony-stimulating factor

CSFR – colony-stimulating factor receptor

cTEC – cortical thymic epithelial cell

CXCR – CXC chemokine receptor

DC – dendritic cell

DE – differentially expressed

DLL – Delta-like

DMAP1 – DNA methyltransferase 1-associated protein 1

DN – double negative

DNMT – DNA methyltransferase

DNMT3L – DNA methyltransferase 3-like

DP – double positive

ETP – early T-cell precursor

ETP-ALL – early T-cell precursor acute lymphoblastic leukemia

FTOC – fetal thymus organoid culture

G-CSF – granulocyte colony stimulating factor

G-CSFR – granulocyte colony stimulating factor receptor

GFP – green fluorescent protein

UHRF1 – Ubiquitin-like, containing PHD and RING finger domains

GM-CSF – Granulocyte-macrophage colony-stimulating factor

GM-CSFR – Granulocyte-macrophage colony-stimulating factor receptor

GSI – gamma secretase inhibitor

GSVA – gene set variation analysis

IL-7R – interleukin-7 receptor

IL7 – interleukin-7

M-CSF – macrophage colony-stimulating factor

M-CSFR – macrophage colony-stimulating factor receptor

MHC – major histocompatibility complex

MRD – minimal residual disease

MTase – methyltransferase

mTEC – medullary thymic epithelial cell

Myeloid cytokines – M-CSF, G-CSF and GM-CSF

NK – natural killer

PC – principal component

PCA – principal component analysis

PCNA – Proliferating cell nuclear antigen

PNT – postnatal thymus

PWWP – proline-tryptophan-tryptophan-proline

RFTS – Replication focus targeting sequence

RNP – ribonucleoprotein

SCF – Stem cell factor

shRNA – short hairpin RNA

SP – single positive

T-ALL – T-cell acute lymphoblastic leukemia

TCR – T-cell receptor

TEC – thymic epithelial cell

TET – ten-eleven translocation

TF – transcription factor

TRD – target recognition domain

TSP – thymus seeding progenitor

UMAP – uniform manifold approximation and projection

VCAM – Vascular cell adhesion protein 1

Acknowledgements

Throughout this PhD, I have been privileged to learn from many scientists and trainees. Their support and enthusiasm have helped me grow both personally and academically, and I am grateful to have been introduced to such a thoughtful group of people. First and foremost, my sincere gratitude and appreciation go towards my supervisor, Dr. Andrew Weng, for teaching me to approach research with a careful and critical eye, and for sharing an ever-abundant enthusiasm towards science. I have grown to appreciate your endless analogies, even if they still leave me confused.

To my committee members, Dr. Christian Steidl and Dr. Gregg Morin, I thank you for providing an abundance of advice and expertise throughout the years. Your insightful perspectives, optimism and encouragement have been instrumental in shaping this project. I would also like to extend my sincere gratitude towards my dissertation examiners – Drs Ninan Abraham, Fumio Takei, Gregg Morin and Gay Crooks, for their thoughtful comments and probing questions.

I would like to thank Weng lab members, past and present – Drs Catherine Jenkins, Michael Nissen, Sam Gusscott, Xuehai Wang, Vincenzo Giambra, and Cath Hoofd, thank you for the insightful discussions, mentorship, and for providing guidance whenever I felt lost in data. Thanks also to Ann Sun, Gabriela Segat, Kate Tyshchenko, Kash Doshi, Prakruti Uday, Dr. Hayedeh Behzad, Dr. Matthew Carr, Dr. Anne-Sophie Hatat and Marissa Foo for the comradery. It has been a pleasure working with all of you, and I will always value the encouragement and thoughtful discussions that we have had over the years.

My appreciation also goes towards the support I have received within the department – Drs Wenbo Xu and Guillermo Simkin, thank you for your patience and assistance in the Flow Core, and for the encouraging discussions over hours spent sorting.

I am grateful to have received financial assistance over the duration of my PhD. The work presented in this thesis would not have been possible without funding from the Canadian Institutes for Health Research, the Terry Fox Research Institute, and the University of British Columbia.

And to my friends and family – thank you for always inspiring me. I am bolstered by your ever-supportive words and love.

Dedication

To Doris.

Chapter 1: Normal and aberrant T-cell development

1.1 Introduction

Leukemogenesis can be understood as an interruption of normal differentiation in which the typical developmental cues that facilitate the controlled development of a cell may be superseded. Establishing a blueprint of normal developmental systems can provide integral context for the study of diseases that arise from them. This thesis investigates normal T-cell development to provide a framework within which to understand the pathogenesis of T-cell leukemia.

In this introductory chapter, I will first provide a current perspective of normal T-cell development. As many aspects of T-cell differentiation are primarily understood from studies that have focused on development in mouse models, many of these sections will refer to T-cell development as it pertains to murine systems. Nonetheless, references to human T-cell development are included where data are available. I will also discuss T-cell leukemia as a consequence of aberrant T-cell development. Focusing on mutations present in T-cell acute lymphoblastic leukemia, I will highlight the role of *DNMT3A* in both normal and aberrant human T-cell development. Lastly, I will provide the objectives and major findings of the work discussed in this thesis summarized per chapter.

1.2 Normal T-cell development

The focus of this section is $\alpha\beta$ T-cell development, which comprises the majority path that T-cells follow, and which differs from $\gamma\delta$ T-cell development. For the purpose of clarity, all phenotypic T-cell populations described in the following chapters refer to T-cell populations in human development, unless otherwise specified. The naming convention used throughout this text to describe human T-cell populations can be found in Figure 1.1 (adapted from[1,2]).

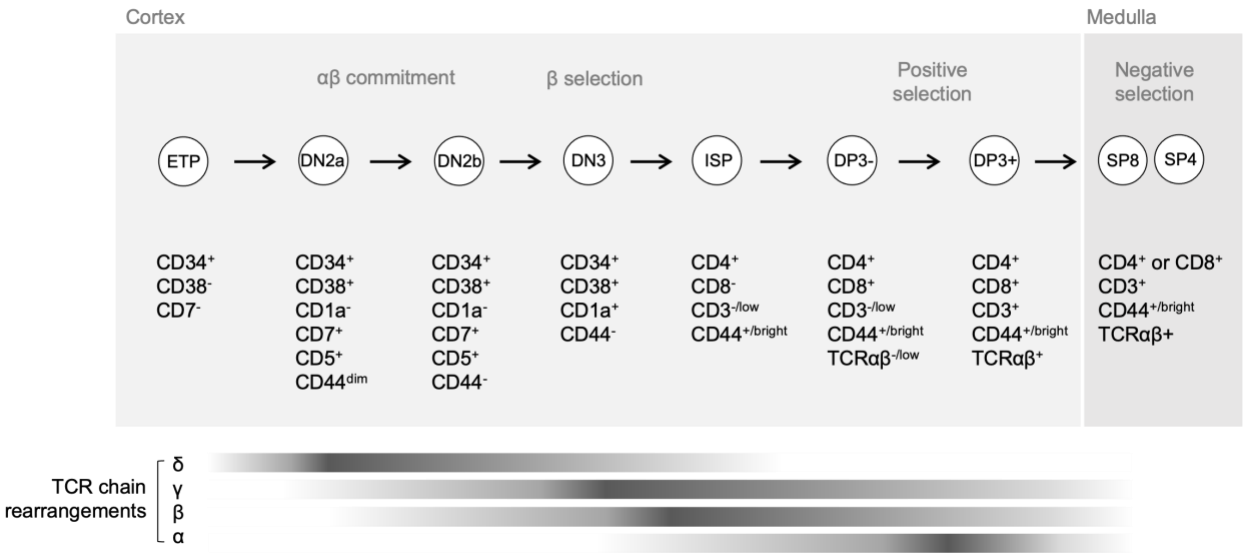


Figure 1.1 Overview of human T-cell development.
Human αβ T-cell development in the cortex and medulla. Early thymic progenitors (ETP) develop from hematopoietic progenitors which seed the thymus. Exposure to Notch ligand results in T-cell specification and lineage commitment within the double negative (DN) compartment. Expression of an in-frame TCRβ chain together with an invariant pre-TCRα results in β-selection and differentiation into immature single positive (ISP) T-cells. Further differentiation results in upregulation of CD8, and together with CD4 expression these cells are classified as double positive (DP) T-cells. Positive selection through interaction with cells within the cortex results in progression to single positive (SP) CD4 or CD8 T-cells. These cells undergo negative selection in the medulla prior to exiting the thymus as mature T-cells. The approximate stages in which rearrangement of the TCR loci (i.e., *TRD, TRG, TRB* and *TRA*) occur are shown below.

1.2.1 Overview of T-cell development in the thymus

Although the thymus is the primary lymphoid organ responsible for generating mature T-cells, it must be continuously seeded by hematopoietic progenitors which emigrate from the bone marrow (BM) (reviewed in[3]). The mechanisms that underly which progenitors will exit the BM and mobilize throughout blood to enter the thymus, or the factors within the niche that recruit progenitor migration, are not fully understood[4–12]. In the context of postnatal thymus in mice, thymic homing of progenitors may be partly controlled through low-affinity interactions between P-selectins expressed on the endothelium of thymic blood vessels and the P-selectin ligand PSGL-1, which is expressed on the surface of a limited subset of BM progenitors[13]. Also identified to be important from murine studies are the chemokine receptors CCR9, CCR7, and possibly CXCR4[14–16]. The expression of cognate ligands, CCL25, CCL19, CCL21, and CXC12, by the thymic epithelium may guide those progenitors expressing the appropriate receptors to settle into

and throughout the thymic niche[17,18]. The action of integrins, including the α4 integrin ligand VCAM1, and other adhesion molecules, such as CD44, are also implicated in thymus homing, migration, and maintaining continuous T-cell—stromal interaction[19,20]. The importance of these molecules in migration may be observed by enrichment of ligands near sites of thymic entry[19,20], and in the case of VCAM1 due to lymphocyte migration defects upon receptor-ligand blockade[21].

Relatively few BM progenitors are thought to journey into the thymus per day, but the few progenitors that do enter through vasculature near the cortico-medullary junction[22,23], a region within each thymus lobule that lies at the interface of two major cellular zones referred to as the cortex and medulla. After thymic entry by extravasation through blood vessels near the cortico-medullary junction, the progenitor cells, typically referred to as thymus seeding progenitors (TSPs), migrate deeper into the cortex and in doing so begin to progressively differentiate into early T-cell precursors (ETPs) through continuous interaction with stromal cells present in the thymus[22,23] (reviewed in[24]). Both the TSP and ETP populations represent very primitive stages of T-cell development in which multilineage potential is thought to be accessible (reviewed in[25,26]).

Early progenitor exposure to Notch ligands in the thymus, such as DLL1, DLL4, JAGGED1, and JAGGED2 expressed by thymic stroma, initiates and maintains these primitive T-cells along a T-cell developmental path[27–33], and this can be measured by the eventual co-expression of CD4 and CD8[31]. Within the thymus, the $CD4^+CD8^+$ double positive (DP) population can be found near the outermost region of the cortex, the subcapsular region, which is marked by an abundance of mitotically active cells[34,35]. Further maturation and eventual thymus egress occurs once cells migrate to and differentiate throughout the medulla, where they eventually downregulate either CD4 or CD8 expression and enter into the periphery as mature $CD4^+$ or $CD8^+$ single positive (SP) T-cells (reviewed in[36]).

Historically, the differentiation of T-cells has been assessed based on the expression of various CD markers which were initially identified in murine subsets; downstream of the TSP and ETP stages, the major stages of T-cell development could be primarily defined according to the expression (or lack thereof) of CD4 and CD8. Immature T-cells, which lack both CD4 and CD8 expression and are termed double negative (DN) T-cells, eventually differentiate into DP T-cells ($CD4^+CD8^+$), and finally $CD4^+$ or $CD8^+$ SP T-cells. Although the expression of these markers can provide rough landmarks to estimate the stage of

differentiation, further resolution of the DN population may be achieved with the inclusion of additional markers. In murine subsets, this occurs upon the basis of CD44 and CD25[37]. These markers are not analogous in human T-cell populations, however, and instead subdivision may be achieved through the temporally coordinated expression of markers such as CD1a, CD44, CD34, CD38, CD7, and CD5, which are sequentially gained or lost throughout human T-cell differentiation (reviewed in[25]). A current understanding of how T-cells passage throughout the thymus, the CD markers that approximate each stage of human T-cell development, and additional features of T-cell development which will be covered in the ensuing text are summarized in Figure 1.1 and Figure 1.2. While the naming conventions to describe these early stages of human T-cell development often vary among different research groups, the convention used in this thesis to define the $CD4^{-}CD8^{-}$ DN T-cell subsets is indicated in Figure 1.1 (adapted from[1,2]).

Importantly, although the majority of the thymic compartment comprises of migrating and developing thymocytes, they are not the only cells which constitute the thymic cellularity; also resident within the thymus are thymic epithelial cells (TECs), as well as mesenchymal cells, endothelial cells, dendritic cells, and B-cell populations, which altogether form the bedrock upon which T-cells develop[38]. Collectively referred to as the thymic stroma, these cells provide indispensable signals to support T-cell specification, maturation, migration, and selection, and are thus fundamental in T-cell ontogeny (reviewed in[39]). The focus of the next section will highlight how these cells, and in particular the thymic epithelial compartment, support the ever-evolving developmental requirements of distinct T-cell stages.

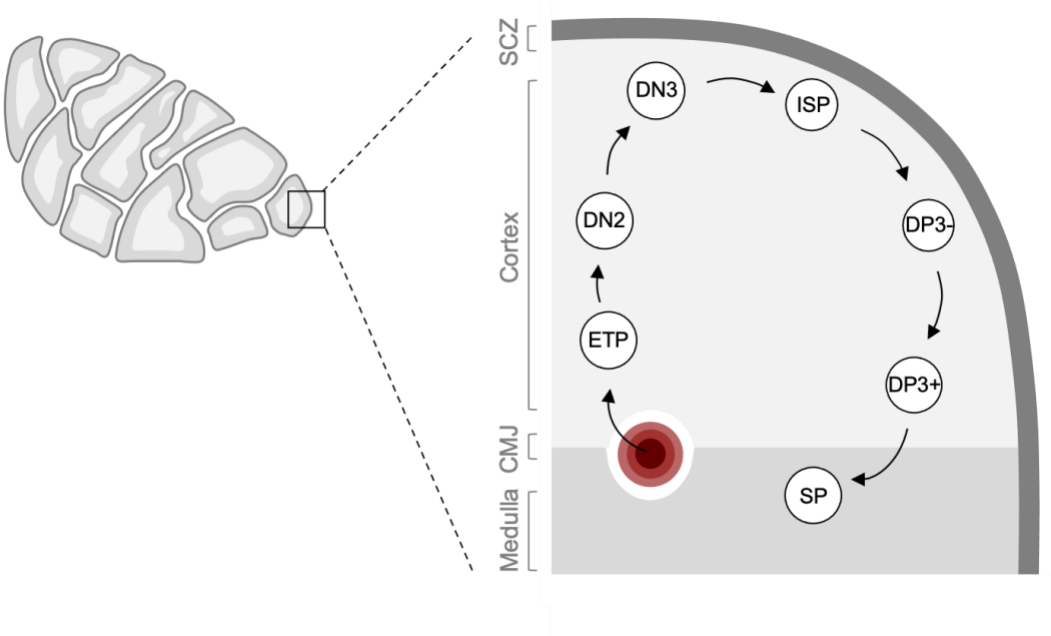


Figure 1.2 Overview of intrathymic αβ T-cell migration.
Thymic seeding precursors enter through blood vessels (indicated by the red circle) near the cortico-medullary junction (CMJ) and circulate throughout the cortex. Entry into the thymus results in progression from early thymic precursors (ETPs) to various stages of $CD4^{-}CD8^{-}$ double negative (DN) subsets. Further differentiation leads to acquisition of CD4 positivity in the immature single positive (ISP) subset. Cells then progress into double positive (DP) cells, which are characterized by expression of both CD4 and CD8. After reaching the subcapsular zone (SCZ), cells undergo positive selection through interaction with cortical thymic epithelial cells and this results in the generation of CD8 and CD4 single positive (SP) cells. Further selection of the SP subset occurs in the medulla through interactions with medullary epithelial cells and dendritic cells before mature thymocytes exit the thymus.

1.2.2 The role of thymic epithelial cells in T-cell development

In early T-cell progenitors, activation of the T-cell program occurs by engagement of the Notch receptor expressed on T-cells by Notch ligands expressed by the thymic stroma[30,31,33,40]. Canonical activation of Notch signaling is thought to occur through four receptors (NOTCH1-4) and five ligands (DLL1,3,4 and JAGGED1,2) (reviewed in)[41]. Activation of the Notch receptor results in a series of proteolytic cleavages which release the intracellular portion of the receptor from the plasma membrane, leading to translocation to the nucleus and activation of downstream target genes (e.g., *Hes-1, p21, cMyc*) through interaction with additional proteins (reviewed in[42]). Continuous activation of Notch signaling through ongoing interaction with the stromal layer is thought to constrain cells along a T-lineage trajectory by

potently excluding the development of alternative lineages (e.g., natural killer (NK) cells, dendritic cells (DCs), and macrophages)[9,43,44] (reviewed in[45]). Depending on the stage of T-cell development, Notch signaling performs various roles in cell proliferation, survival, and enforcement of T-lineage specification[27,46,47], and is thus necessary for proper T-cell development and function. Nonetheless, T-cell differentiation likely requires additional stage-specific factors that extend beyond Notch signaling[48,49]. In fact, various stages of progressive T-cell differentiation can be traced to particular regions of the thymic architecture throughout both the cortex and medulla[22,23]. This observation possibly suggests that disparate regions of the thymus provide unique conditions that trigger or support distinct stages of development (reviewed in[24]).

The two main regions within the thymus can be broadly organized into the cortex and the medulla (Figure 1.2). These regions are not separated by a formal anatomical structure but rather can be distinguished by functional and spatial heterogeneity[35]. Functionally, differences between the cortex and medulla are driven by variation in cellularity. The cortex and medulla are also spatially disparate, with the cortex forming the outermost region of each thymus lobule and the medulla comprising the central core (reviewed in[50,51]). These differences can be visually discerned by histological staining[35]; the darker-staining cortex is densely packed and resident primarily to developing thymocytes, but also cortical TECs (cTECs), and to a lesser extent mesenchymal cells and endothelial cells (reviewed in[50,51]). The lighter-staining medulla is relatively less cellular, with a smaller population of maturing T-cells and a stromal compartment populated mostly by medullary TECs (mTECs), though dendritic cells may also be found (reviewed in[52]). Although the thymic epithelial compartment is divided according to the spatial localization within the thymus (i.e., cortical or medullary), differences among TECs also extend beyond thymic location. These differences, which include variations in secreted cytokine and chemokine milieu, transcription factor expression, and antigen processing machinery, ultimately reflect a shift in the developmental requirements of a T-cell as it migrates throughout the cortex to the medulla, and in doing so begins to lose features of multipotency and gain features that are unique to T-cell function and survival (reviewed in[53]).

The cortical epithelium are among the first cells that thymus seeding progenitors encounter[22,23]. During early development, the role of cTECs is primarily implicated in T-lineage specification/early activation of the T-cell program and expansion of the thymic immigrant pool (reviewed in[18]). It is estimated

that newly arrived thymic immigrants expand ~1000-fold, perhaps through the action of cytokines secreted by the cTECs and other stromal cells, which at least in mice may include stem cell factor (SCF)[54], interleukin-7 (IL-7)[55], and Wnt ligand[56], although data on the precise cellular source, concentrations, and compendia of these signaling factors is limited (reviewed in[24]). It is likely that the composition and concentration of signaling factors secreted by cTECs varies even within the cortex; many of these cytokines have pleiotropic effects on T-cell development, and their function in supporting the DN compartment must too evolve with the developmental requirements of migrating T-cells (reviewed in[57]). Some of these pleiotropic, stage-specific effects include provision of proliferation and survival signals during early development, and contribution to cell adhesion and specialized T-cell function during subsequent developmental stages (reviewed in[24]). For example, IL-7 secreted by TECs supports early T-cell development by promoting cell expansion and survival[49]; in the context of immature T-cells, engagement of the IL-7 receptor has been shown to activate STAT5, resulting in upregulation of the pro-survival protein Bcl-2[58–61]. At later stages of development, however, IL-7 may instead act to increase accessibility of the T-cell receptor (TCR) loci[62]. Increased locus availability, coupled with upregulation of *RAG* recombinase, initiates the irreversible and T-cell-specific process of TCR rearrangement[63].

cTECs also play important roles in positive selection of DP thymocytes post-TCR rearrangement, primarily through diversification and positive selection of the TCR repertoire (reviewed in[53]). This occurs through the expression of a unique subset of peptides, thought to be generated through the action of a unique set of proteolytic enzymes presented by major histocompatibility complexes (MHC)[64–66] (reviewed in[18]). Engagement of peptide-MHC complexes by the TCR results in survival signals among the T-cells expressing receptors that bind with the appropriate affinity, and death by neglect in cells that do not (reviewed in[67]). At this stage, cells are thought to be programmed for cell death, perhaps due to a decrease in IL-7 signaling as a consequence of IL-7R downregulation at the DP stage, and can only be rescued from apoptosis through the process of positive selection[68].

By the time thymocytes have begun positive TCR selection, they are phenotypically distinct from when their migration first initiated, and now express e.g., CD4 and CD8. At this point in development, cells have reached the subcapsular zone of the cortex, located at the outermost region of each thymic lobule, and have begun migrating to the medulla for further TCR selection. It is perhaps paradoxical that the same

stromal tissue that supported the differentiation of these once-DN cells migrating outwards to the subcapsular region must also support the DP and $CD4^+/CD8^+$ SP cells that now traverse back throughout the cortex for further selection and maturation in the medulla (reviewed in[36]). Although the stromal matrix remains itself unchanged, the developmental requirements of the T-cells now differ; though a main feature of early T-cell development was proliferation induced by e.g. SCF and Notch signaling, DP cells now migrating towards the medulla show a reduction in both proliferation[34] and reliance on Notch signaling[69] (reviewed in[24]). The ability of T-cells to shift signaling programs in response to the same environmental conditions may reflect developmental context-specific signaling and the pleiotropic effects of signaling factors (reviewed in[24]). For the cells to continue their development, however, they must eventually encounter an altogether new environment.

In the cortex, migration is thought to be influenced by the presence of CXCL12-producing stromal cells, which act on CXCR4-expressing T-cells[17]. The paucity of CXCL12-producing cells in the medulla, however, suggests that there are alternative factors that influence migration[17]. Indeed, the stromal compartment expresses an overlapping and distinct complement of cytokines that influences cell migration at various stages of development (reviewed in[53]). Migration to the medulla, which is thought to be required for further TCR selection and maturation, may occur through upregulation of CCR7 at the surface of T-cells after they undergo positive selection in the cortex and through secretion of the CCR7 ligands, CCL19 and CCL21, by resident mTECs[70].

mTECs, through interplay with dendritic cells and other antigen-presenting cells, act to remove any cells bearing TCRs that bind to self-antigen with too high affinity in a process referred to as negative selection (reviewed in[67]). This is achieved through the unique capacity of mTECs to express nearly all coding genes present in the genome through the activity of the autoimmune regulator (AIRE)[71,72]. *AIRE* facilitates ectopic expression of tissue-restricted proteins, thus generating a plethora of self-peptides which are presented by MHC. Ultimately, interaction with mTECs and antigen presenting cells in the medulla results in screening against T-cells bearing TCRs that will be autoreactive upon exit from the thymus (reviewed in[73]).

Overall, it is estimated that only 1-3% of T-cells survive the stringent selection pressures presented by the thymus[34,74]. Yet, it is estimated that approximately 1-2 million post-selected cells leave the thymus

per day in young adult mice[75]. Thus, while a prominent role of the stromal compartment is to provide a series of unique microenvironments that select for T-cells which express the appropriate TCRs, it must also function to efficiently support the proliferative demands of developing T-cells and their ever-shifting requirements as they enter the thymus, establish the T-cell program, and undergo the highly rigorous process of positive and negative selection, which together serves to prepare cells for their eventual exit from the thymus.

1.2.3 Egress from the thymus

Equally important to how cells enter into and develop within the thymus is the process by which cells exit into the periphery, where further maturation and acquisition of specialized T-cell functions occurs. Thymic egress necessitates a shift in signaling requirements and/or signaling source. For example, reliance on thymic signaling factors for survival and proliferation, such as Notch signaling, may be lost or contracted upon thymic egress; where the activation of Notch signaling through continuous interaction between ligand-expressing stromal cells and early developing T-cells is thought to constrain cells to the T-lineage and provide pro-survival cues, the degree/duration of signaling may be reduced at the DP stage, after a cell is T-cell committed[24,30,69,76] (reviewed in[77]). This may be reflected by the level of Notch ligand expression throughout the thymus, which is highest among the cortical epithelial compartment[54]. Although Notch signaling remains important once cells leave the periphery, its function in this context is thought to shift towards specialized roles in immune response including cytokine production, proliferation, and T-cell activation, and away from roles in T-cell proliferation, survival, and lineage enforcement[77,78] (reviewed in[77]).

Nevertheless, reliance on some signaling factors (e.g., IL-7) is maintained even after cells exit the thymus, and may be necessary for mature T-cell survival[79,80] (reviewed in[81]). The source of IL-7 stimulation necessarily shifts from stromal cells in the medulla to cells in the periphery. Although the precise extrathymic cellular source of IL-7 is difficult to determine, studies using IL-7 reporters in mice indicate the presence of IL-7 in peripheral tissues including the lymph nodes, liver, and intestine[79,80].

1.2.4 T-cell receptor rearrangement

Expression of a functional TCR is central to the role of T-cells and is the outcome of successful rearrangement of the variable region of the *TRA, TRB, TRG*, and *TRD* genes. Each TCR chain consists of a variable (V), diversity (D), joining (J) and constant (C) gene segment, and the unique recombination of the VDJ segments (in the case of TCRβ and TCRδ) or the VJ segments (in the case of TCRα and TCRγ) generates a unique receptor (Figure 1.3, reviewed in[82]). Productive rearrangement of either *TRD* and *TRG*, or *TRB*, results in the surface expression of either TCRγδ or TCRαβ, respectively.

The majority of cells (>90%) adopt a TCRαβ fate, in which an in-frame *TRB* rearrangement results in expression of the pre-TCR composed of a TCRβ chain, an invariant pre-TCRα chain, and CD3 signaling molecules (Figure 1.3, reviewed in[45]). Cells with an in-frame TCRβ rearrangement are selected in a process known as β-selection; signaling through a functional pre-TCR complex rescues T-cells from apoptosis and induces proliferation, cessation of further rearrangement of the TCRβ locus, and maturation by initiating TCRα gene rearrangement and CD4/CD8 co-expression (reviewed in[45]). The β-selection checkpoint is reserved for cells that express a pre-TCR; the minority of cells that express a γδ TCR undergo a distinct maturation and selection path which involves signaling through the γδ TCR and does not typically result in the generation of DP T-cells (reviewed in[83]). The mechanisms controlling αβ and γδ lineage bifurcation are incompletely understood but may be related to TCR signaling strength[84–86].

In a manner reflective of the stepwise differentiation of thymocytes, rearrangement of TCR loci occurs in a sequential fashion, with the order of rearrangement correlating with development. Studies from human T-ALL and normal T-cell development suggest that rearrangement occurs in as follows: *TRD, TRG, TRB, and TRA*, as indicated in Figure 1.1, with the first instance of *TRD* rearrangements found in the $CD34^+CD38^-CD1a^-$ subset, and detection of complete in-frame *TRB* rearrangements beginning in the CD4 ISP stage[87–89]. This sequential rearrangement of the TCR loci may occur due to the stage-specific expression of recombination-activating genes, *RAG1* and *RAG2* recombinase, which induce double-strand breaks at recombination sequences within the V(D)J region to facilitate TCR rearrangement, but also by stage-specific accessibility of the TCR loci themselves[90].

At the RNA level, *RAG1/2* expression initiates in the earliest T-cell progenitors, reaches maximal expression at the DP $CD3^-$ stage, and is eventually extinguished in SP $CD4^+$ or SP $CD8^+$ cells[87]. Reporter studies in mice suggest that *Rag1/2* levels may be characterized by two expression peaks; one coincident with the initiation of *TRB* rearrangement, and the other occurring immediately prior to *TRA* rearrangement[91,92]. Because the order of chain rearrangement is also influenced by the ability of RAG enzymes to access TCR gene segments, unilateral opening of the TCR loci can also control the order in which rearrangement is initiated (reviewed in[93]). Sequential accessibility of the TCR loci may depend on the action of cytokines such as IL-7; transcription factors such as STAT5, E2A, and HEB; and epigenetic factors such as the chromatin remodeling SWI-SNF complex[62,94], although many additional factors are likely involved.

TCR rearrangement is integral to proper T-cell function, but it also creates an indelible mark of a cell's differentiation history; the permanence of rearrangement, which involves irreversible breaking and re-joining of DNA fragments, can thus indicate a certain degree of T-cell development[95,96]. Under permissive conditions, however, TCR-rearranged cells may yield non-T-lineage cells, such as NK, DC, granulocytic, and monocytic cells[97,98]. This suggests that T-lineage decisions may be altered even after immutable, lineage-specific DNA rearrangements are initiated.

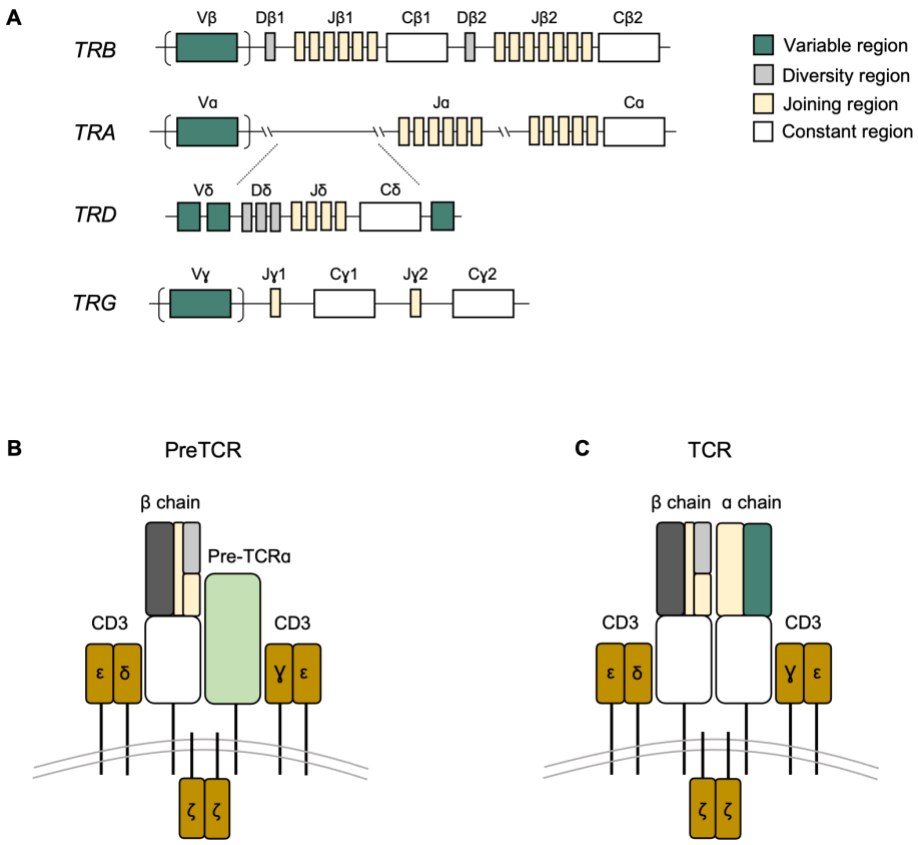


Figure 1.3 Schematic of TCRα and TCRβ rearrangement and assembly of pre-TCR and TCRαβ complexes.

A Organization of the TCR loci. Depicted is the genomic organization of the *TRB*, *TRA*, *TRD* and *TRG* loci containing variable (V), diversity (D), joining (J) and constant (C) regions. Recombination of the V(D)J regions results in a diverse TCR repertoire which is further diversified by addition or deletion of nucleotides at junction sites. The order of processing is listed vertically (from top to bottom) for each gene.

B Diagram of the assembled pre-TCR complex at the cell surface containing a rearranged TCRβ chain and a surrogate pre-TCRα chain complexed with CD3 signaling molecules (ε, δ, γ, and ξ). The pre-TCR complex is important for regulating T-cell development prior to rearrangement of the α chain.

C Diagram of the final assembled TCR complex at the cell surface containing a rearranged TCRα and TCRβ chain in complex with CD3 signaling molecules.

1.2.5 T-lineage commitment

Although the precise cells capable of seeding the thymus are a subject of continued investigation in both murine and human T-cell development[3,6,99,100] (reviewed in[99]), the cells that do migrate to the thymus have multi-lineage potential, including B, T, NK, myeloid, DC, and erythroid lineages, and thus are not yet

bound to the T-cell program – at least when cultured in the appropriate *in vitro* conditions[9,25,43,95,101,102] (reviewed in[25]). The selective elimination of alternative (non-T) lineage potential is referred to as "commitment", and is influenced by cell-intrinsic factors (e.g., transcription factors and chromatin accessibility of gene targets), cell-extrinsic factors (e.g., environmental contexts supporting lineage development), or the interplay of both (e.g., a cell's ability to migrate to new environments and respond to key signals)[103] (reviewed in[104,105]). The process of T-cell commitment is not likely to occur instantaneously; the elimination of alternative lineage potentials is a temporally protracted series of events which occur in an ordered fashion and over multiple rounds of cell division (reviewed in[106]).

Commitment can be measured by probing a cell's potentials under permissive conditions. In the case of T-lineage commitment, this may be done by extracting thymic subpopulations and placing the cells in conditions that support the lineage in question. Loss of potential may be determined by the inability of a cell to respond to lineage-specific conditions. The complete loss of all non-T lineage potentials indicates full commitment to the T-lineage (reviewed in[104,106]). When combined with immunophenotypic and transcriptomic profiling, examination of thymic subsets pre- and post-commitment can reveal the approximate developmental stages in which alternative lineage potentials are lost, as well as the transcriptional or epigenetic changes associated with commitment.

Initial studies in human thymocyte subsets suggested that committed cells reside within the $CD34^+CD1a^+$ compartment[43], a population which also demonstrates increased levels of TCRβ rearrangement[87], and that multipotent cells could be found among $CD34^+CD1a^-$ cells. The $CD34^+CD1a^-$ compartment was later revealed to be functionally heterogeneous, as subsequent studies indicated that the B, NK, myeloid and erythroid lineage potentials varied with CD7 expression[101]. Even so, further enrichment of the T-cell committed population could be achieved by selecting the $CD7^+CD5^+CD45^{dim}CD44^-$ population, with commitment and the onset of TCRβ rearrangements correlating with the loss of $CD44^{dim}$ expression[95]. This was recently confirmed on a more granular level; in a study by Le *et al*. (2020)[107], single-cell RNA sequencing was paired with *in vitro* lineage potential assays, resulting in further division of the uncommitted $CD44^{dim}$ subset into $CD2^+$ and $CD2^-$ fractions. Gain of CD2 expression coincided with a loss of B, NK, and myeloid potential, with a more prominent loss of B-lineage potential. The authors concluded that B-lineage potential is lost prior to NK and myeloid potential during human thymocyte development[107]. Collectively,

studies such as those highlighted here provide a framework for understanding the phenotypic contexts in which alternative lineage potentials remain accessible under permissive conditions *in vitro* in human T-cells, and thus allow for further dissection of the transcriptional and epigenetic factors that guide eventual commitment.

Commitment to the T-cell lineage can be understood as a series of repressive mechanisms; the choice to become a T-cell is only made once all other options have been excluded (reviewed in[106]). This is particularly salient from the aforementioned experiments which have demonstrated that alternative (non-T) lineage potentials remain upon removal of cells from the thymus[10,32,43,95,97,101,107,108], and might suggest that T-cell restriction is necessarily maintained by the external signaling cues provided by the highly specialized thymic microenvironment (e.g., Notch signaling[109]). Yet, T-cells which naturally egress from the thymus remain T-lineage restricted, suggesting that while signaling must be sustained through to commitment, it may not be needed thereafter. This set of observations warrants further investigation of the repressive mechanisms which might constrict non-T lineage potential within the thymus.

In eukaryotic cells, DNA methylation is among the most well-characterized repressive mechanisms that are stably maintained throughout cell division (reviewed in[110]). The role of DNA methylation to repress pluripotent genes has been previously identified in the context of embryonic stem cell differentiation[111,112] (reviewed in[113]). The role of DNA methylation in T-lineage specification and repression of stem-associated genes, however, remains unclear. The next section focuses on DNA methylation and highlights a current understanding of its role throughout T-cell development.

1.3 DNA methylation in normal T-cell development

1.3.1 Overview of DNA methylation by DNMT3A, DNMT3B, and DNMT1

Establishment of DNA methylation patterns is critical for normal tissue differentiation and involves the addition of a methyl group to the 5' carbon of a cytosine base, referred to as 5-methylcytosine (5mC), which typically occurs in the context of CpG dinucleotides in mammals. Methylation is established predominantly through the activity of the *de novo* methyltransferases, DNA methyltransferase 3A (DNMT3A) and DNMT3B, and must be faithfully maintained throughout cellular replication by DNMT1[114–116]. Classification of DNMTs into initiation versus maintenance functions generally relates to substrate

specificity; DNMT1 displays higher catalytic activity directed towards hemimethylated CpGs[115] (reviewed in[117]), whereas DNMT3A and DNMT3B typically show preferential activity toward unmethylated CpGs[118]. However, this division of labor is not always clear; in some studies, DNMT3A and DNMT3B showed no preference for unmethylated or hemimethylated DNA[119], and in other studies, DNMT1 displayed *de novo* methylation activity at certain CpG sites[120,121]. Nonetheless, the coordination of methylation initiation and maintenance creates a system where DNA methylation patterns, once established, are faithfully propagated to downstream progeny.

The distribution of CpG dinucleotides throughout the genome can lend clues to the functional role of DNA methylation; although mammalian genomes are globally depleted of CpGs, the CpGs that are present are distributed asymmetrically throughout the genome[122–124]. Regions of high CpG density tend to concentrate near the transcriptional start sites of genes, especially highly expressed housekeeping genes and developmentally important genes[123–127]. These regions, termed CpG islands (CGIs) are generally hypomethylated relative to other CpG sites throughout the genome (reviewed in[128]). There are exceptions, however, and these include imprinted genes and X-chromosome genes, which are not typically expressed at high levels[129] (reviewed in[130]). Generally, there is an inverse correlation between CpG methylation and gene transcription[131–138], and this becomes stronger with increasing CpG density (reviewed in[128]). Collectively, these studies contribute to a growing body of evidence that implicates CpG methylation in gene silencing.

Both *in vivo* and *in vitro* experiments generally support the role of DNA methylation in transcriptional silencing. Although the mechanism of gene repression is not fully resolved as of yet, two major models have emerged[132,135,138] (reviewed in[139]): First, the addition of a 5-methyl group can prevent some, but not all, transcriptional activators from binding to DNA, perhaps through steric hindrance or by alterations to the local DNA structure[140,141]. In some instances, however, CpG methylation may recruit transcription factor (TF) binding, or may not have an effect on TF binding at all[140,142], suggesting that other regulatory mechanisms may be involved. Second, CpG methylation may inhibit gene transcription by creating binding sites for proteins with a methyl CpG binding domain[143], which can lead to recruitment of chromatin remodeling complexes and transcriptional co-repressors that silence transcription[139,143–145]. In other cases,

DNA methylation machinery may be instead influenced by pre-existing histone marks; active histone marks, such as H3K4 methylation, have been shown to preclude CpG methylation[146,147].

The interplay between CpG methylation by DNMTs, epigenetic regulators, and chromatin structure adds additional layers of transcriptional control which can ultimately result in stable gene repression via heterochromatin formation (reviewed in[148]). This raises the question of how DNMTs are consistently directed to certain genomic regions, such as imprinted genes, X chromosome genes, and repeated DNA sequences like those found in centromeres, to facilitate methylation. The answer may lie in the partners with which DNMTs interact; importantly, DNMTs participate in larger protein complexes through various interaction and binding domains, and do not alone exert CpG methylation (reviewed in[149]).

1.3.1.1 DNA methyltransferase protein structure

Recruitment to, and methylation of, specific regions of DNA may be facilitated by the various protein domains of DNA methyltransferases. DNMT proteins generally consist of a diverse amino terminus, which differs between DNMT1 and DNMT3A/B, and a relatively conserved carboxyl terminus. The N-terminus has roles in nuclear localization, substrate specificity, activation of enzymatic activity, and protein-protein interactions. The C-terminus is important for cytosine methyltransferase activity and contains the catalytic methyltransferase (MTase) domain which is common to DNMT proteins (reviewed in[150]) (Figure 1.4). The domain architecture of the DNMT1 and DNMT3 enzymes are discussed below.

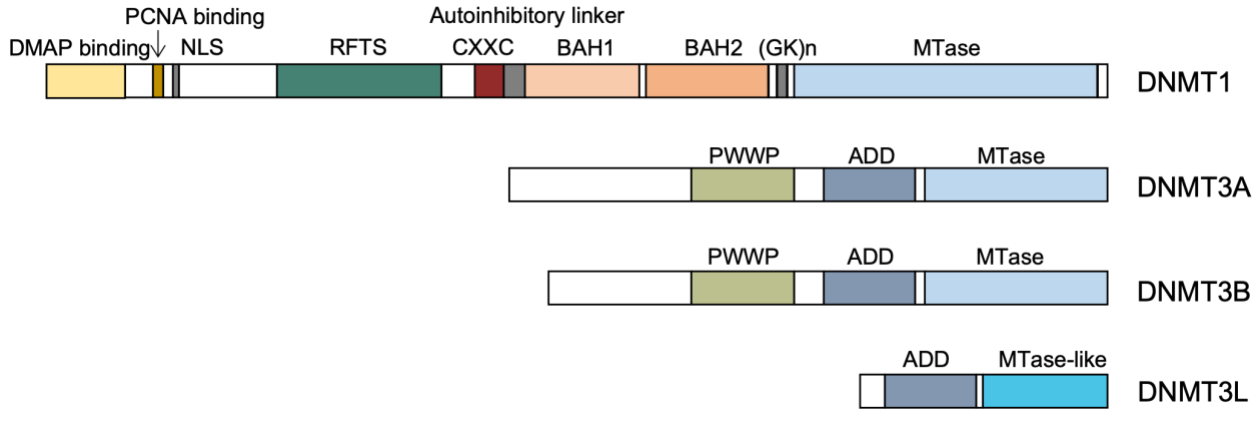


Figure 1.4 Domain architecture of the DNA methyltransferase proteins.
Domains of human DNMT1, DNMT3A, DNMT3B and DNMT3L.

DNMT1

Although DNMT1 contains the conserved C-terminal catalytic domain present in all cytosine DNMTs, the N-terminus of DNMT1 is unique, which may relate to its function in maintaining DNA methylation during replication. The N-terminal region of DNMT1 contains several domains which facilitate DNA or protein binding. This includes the DNA methyltransferase-associated protein 1 (DMAP1) domain, which interacts with the transcriptional repressor DMAP1[151]; the proliferating cell nuclear antigen (PCNA)-interacting domain, which recruits DNMT1 to DNA repair sites[152] and to the replication fork during S-phase[153]; the replication-foci targeting sequence (RFTS) domain, which localizes DNMT1 to replication foci via binding to ubiquitylated histone H3[154] and to constitutive heterochromatin throughout S and G2/M phase and participates in active site regulation[116,155–157]; a CXXC zinc finger domain that binds specifically to DNA containing unmethylated CpG dinucleotides[158]; and two bromo-adjacent-homology (BAH) domains which also have roles in targeting DNMT1 to replication foci during S-phase[159].

To explain the observation that DNMT1 preferentially methylates hemimethylated DNA, a model of allosteric regulation of DNMT1 has emerged in which multiple domains at the N-terminus form a regulatory region that inhibits *de novo* methylation by occluding the active site of DNMT1 unless bound by hemimethylated DNA[158,160,161] (reviewed in[157]). This is achieved through the CXXC domain, RFTS domain and the CXXC-BAH1 linker which joins the CXXC and BAH1 domains (referred to as the autoinhibitory linker[158]). Binding of unmethylated CpGs by the CXXC domain positions the autoinhibitory linker across the catalytic cleft of DNMT1, thereby excluding DNA from the catalytic site and preventing *de novo* methylation[158]. When not bound to DNA, the RFTS domain inserts into the DNA-binding pocket of the catalytic site and consequently blocks access to the active site[156,157]. The catalytic DNA-binding pocket of DNMT1 becomes accessible upon binding of other proteins, such as UHRF1 (ubiquitin-like, containing PHD and RING finger domain protein 1), which interacts with the RFTS domain and results in its extrusion from the catalytic site[162,163].

DNMT3 family

DNMT3A and DNMT3B

DNMT3A and DNMT3B belong to the DNMT3 family, which consists of three paralogues: DNMT3A, DNMT3B, and the catalytically inactive paralogue, DNMT3-like (DNMT3L). DNMT3A and DNMT3B have a similar domain structure; the N-terminal region, which is involved in chromatin targeting and binding, contains an ATRX-DNMT3-DNMT3L (ADD) reader domain and a Pro-Trp-Trp-Pro (PWWP), and the C-terminal region contains an MTase domain. DNMT3L lacks the majority of the N-terminal region, containing only the ADD domain at the N-terminus, and at the C-terminus has an MTase-like domain which lacks motifs integral for catalytic function (reviewed in[164]).

The ADD domain functions in protein targeting by binding to unmodified lysine 4 of H3 tails (H3K4me0)[146] and in enzyme autoinhibition. Structural analysis of DNMT3A indicates that the ADD domain binds to two distinct sites, inducing either allostery or autoinhibition depending on interaction, or lack of, with H3K4me0 tails; when not bound to H3 peptide, DNMT3A adopts an autoinhibitory conformation in which the ADD domain occludes the DNA binding cleft at the catalytic site[165–167], thus linking DNMT3A with specific H3 chromatin marks. This method of enzyme regulation is also likely to operate in DNMT3B due to conservation of residues that facilitate the autoinhibitory mechanism[168].

The PWWP domain of DNMT3A and DNMT3B, named for its proline-tryptophan-tryptophan-proline motif, is important for localization to pericentromeric chromatin. This can occur through recognition of, and subsequent binding to, H3K36me3[169,170]. The PWWP domain is also implicated in targeting enzyme activity to actively transcribed gene bodies[171].

The N-terminus of DNMT3A and DNMT3B also contains a variable region which is relatively divergent between the two paralogues, and may have roles in regulating protein localization and DNA binding[172]; at least in the case of DNMT3A, isoforms with a truncated N-terminus region show localization to euchromatin, as opposed to heterochromatin observed in the full length isoform. Isoforms of DNMT3B may also differ in catalytic activity, and in the context of murine HSCs cells, the catalytically inactive splice isoform is most dominant[173].

The MTase domain performs the enzymatic addition of cytosine and contains a number of key residues which facilitate this process. Some of these residues form two loops, known as the target

recognition domain (TRD) and the catalytic loop. Together, the TRD and catalytic loop bind to and enclose the target CpG site[174,175].

DNMT3L

DNMT3L is catalytically inactive due to a lack of several motifs within the MTase domain, including the TRD and catalytic loop[175], and does not directly bind DNA[174–176]. Nonetheless, it plays important structural roles in augmenting the enzymatic activity of DNMT3A and DNMT3B[146] and is involved in DNMT3A/B protein stability and processivity[177,178]. Together with DNMT3A and/or DNMT3B, DNMT3L forms a tetramer in which two DNMT3L-DNMT3A/B heterodimers interact to form a 3L-(3A/B)-(3A/B)-3L complex, which creates a (3A/B)-(3A/B) interface and a (3A/B)-3L interface (Figure 1.5) [150,176,179]. This arrangement forms two sites in which CpGs located on opposite DNA stands may bind to DNMT3A/B[179], and this interaction is stabilized by DNMT3L[174].

Figure 1.5 Model of DNMT3-DNMT3L tetramer formation
Tetrameric complex formed by A) two DNMT3A-DNMT3L heterodimers and B) two DNMT3B-DNMT3L heterodimers. 3L = DNMT3L; 3A = DNMT3A; 3B = DNMT3B.

1.3.1.2 DNA demethylation

The addition of a methyl group to the 5' carbon of cytosine forms a covalent bond that, once established, is chemically stable[180]; under physiological conditions, removal requires enzymatic cleavage, or disruption of methylation maintenance. These two major mechanisms of DNA demethylation are referred to as active or passive DNA demethylation, respectively.

Active removal of 5mC occurs through the action of the ten eleven translocation (TET) family of enzymes, which include TET1, TET2, and TET3. Together, the TET enzymes catalyze oxidation of 5mC to a series of intermediates, which are ultimately converted to 5-carboxylcytosine (5caC)[181]. The oxidized

intermediates are not major substrates for DNMT1[182–184], and are eventually returned to unmodified cytosine through passive dilution (reviewed in[185]). Alternatively, some oxidized intermediates, including 5caC, may be excised by thymine DNA glycosylase through base excision repair mechanisms[186,187].

Passive removal of DNA methylation marks occur during successive rounds of replication in the absence of methylation of newly synthesized DNA strands, eventually resulting in dilution of 5mC by division (reviewed in[188]). Therefore, the relationship between replication and DNA methylation is necessarily linked; to preserve DNA methylation patterns in daughter cells, methylation must be re-established during replication. Altering the equilibrium of cell proliferation and expression/activity levels of maintenance methylation machinery may therefore influence the inheritance of CpG methylation in daughter cells[189] (reviewed in[190]). Since passive DNA demethylation first requires a cell to divide, it is perhaps more suitable for temporally protracted events such as lineage-specification and development, which is typically accompanied by cellular replication, and less appropriate for biological events which require an immediate response[185].

1.3.2 DNA methylation in T-cell differentiation and commitment

Establishment of *de novo* DNA methylation has important consequences during blood cell development. The differentiation of HSCs into mature progeny is accompanied by the acquisition of DNA methylation patterns which are distinct among the hematopoietic lineages[191,192] and which may be used to distinguish different cell types[193]. The importance of DNA methylation patterning may be most evident upon targeted deletion of the *de novo* DNA methyltransferases, which are critical for establishing DNA methylation; in murine hematopoiesis, conditional ablation of *Dnmt3a* and *Dnmt3b* in HSCs using an Mx1-Cre driver results in enhanced self-renewal (measured by increased HSC expansion in up to four serial transplantations) and a block in differentiation of all downstream hematopoietic lineages (measured by a reduction in cell output)[173]. This phenotype is most prominent upon *Dnmt3a* loss; while deletion of *Dnmt3b* alone shows only minor effects on differentiation compared to wildtype HSCs[173,194], *Dnmt3a*-null HSCs demonstrate a measurable differentiation defect accompanied by enhanced self-renewal in comparison to wildtype HSCs[194]. In subsequent studies that focused on the effects of *Dnmt3a* loss on T-cell differentiation, conditional loss of *Dnmt3a* in murine HSCs was associated with an accumulation of immature thymocytes

at the DN2 stage of murine T-cell development. This phenotype, however, was incompletely penetrant; differentiation arrest was only observed after secondary transplantation and in a minority of recipients (9/24 recipients)[195], possibly suggesting that additional events are required for *Dnmt3a* loss to disrupt T-cell development.

Dynamic changes in DNA methylation occur at multiple stages throughout murine and human T-cell development[192,196]. In particular, earlier stages of differentiation are associated with a more pronounced gain of DNA methylation as opposed to loss, implicating *de novo* methylation as a possible important regulator of early T-cell differentiation[192,196]. When focusing on differentially methylated regions during key developmental timepoints throughout T-cell development, including T-cell commitment, β-selection, TCRαβ expression, and positive selection, the greatest increase in DNA methylation can be observed during the process of T-lineage commitment[196]. During the latter development checkpoints, however, DNA demethylation events, as opposed to *de novo* methylation events, are more common[196]. Throughout these later stages of development, progressive demethylation may be important for activation of certain genes specific for T-cell function, such as *Lck*; Lck regulates TCR signaling and is progressively demethylated throughout T-cell differentiation, which is associated with a concomitant increase in gene expression[192]. In contrast, the increase in *de novo* DNA methylation observed during T-cell commitment presents a possible role of DNA methylation in myeloid and T-lineage specification; in murine and human hematopoiesis, committed myeloid lineage cells display a lower level of global DNA methylation compared to cells committed to the lymphoid lineage[191,192], and treatment of murine T-cell progenitors with the DNMT inhibitor, 5-azacytodine, in the presence of both myeloid- and T-lymphoid-promoting cytokines, results in a modest skew towards the myeloid lineage[192]. Taken together, this suggests that DNA methylation may play important roles in establishing or maintaining T-lineage commitment. Yet, the role of *de novo* DNA methylation in T-cell commitment, and the compendia of genes which may be regulated by this process, is not well characterized in human development. When focusing on genes associated with progenitor or alternative lineage programs (i.e., *LYL1, HHEX, SPI1, CEBPA,* and *LMO2*), reduced expression throughout T-cell development does not consistently correlate with an increase in promoter methylation, possibly suggesting that additional mechanisms of regulation may be involved in the repression of alternative lineages[196].

Further insight into how *de novo* DNA methylation may be operative in the context of T-cell development may be gained from diseases in which methylation machinery is mutated. Frequently, *DNTM3A* is altered in T-lymphoid and myeloid leukemias, though to a lesser extent *DNMT1* and *DNMT3B* mutations are also observed[197] (reviewed in[198]). The next section discusses aberrant T-cell development, with a primary focus on T-cell acute lymphoblastic leukemia (T-ALL), and briefly highlights the role of DNA methylation machinery as understood in this context.

1.4 T-cell acute lymphoblastic leukemia

1.4.1 A brief overview of T-ALL

The processes that govern T-cell development must be tightly regulated to orchestrate the production and differentiation of T-cell progenitors into functionally mature subsets. The expansion of immature T-cells is highly controlled and centered around successful rearrangement of T-cell receptor loci. Given the complexity of this highly regulated process, combined with requisite double stranded DNA breakage and rejoining events, it is perhaps inevitable that errors at any step can interrupt developmental progression and ultimately lead to cellular transformation. T-cell acute lymphoblastic leukemia (T-ALL) is the pathologic designation applied to malignant accumulations of immature T-cells, most often involving thymus, bone marrow, and peripheral blood. Current treatment with high intensity multi-agent chemotherapy achieves an event-free survival rate of approximately 85% over 5 years in the pediatric setting (reviewed in[199]). There are, however, long term side effects from cytotoxic chemotherapy and radiation, in addition to dismal outcomes for relapsed disease (reviewed in[199]). In adults, current therapy regimens fare poorly, with disease-free survival rates below 40% (reviewed in[200]), thus advocating for increased understanding of the molecular underpinnings of T-ALL for improved treatment outcomes.

A molecular basis for T-ALL pathogenesis could be first understood from cytogenetic analyses which revealed frequent chromosomal aberrations that altered the expression of a recurrent selection of genes (reviewed in[201,202]). Most commonly, these include deletions affecting chromosome 9 that result in loss of the *CDKN2A* locus[203–205], which encodes the cell cycle genes p16/INK4A and p14/ARF. An outcome of this deletion is resistance to apoptosis and progression through the cell cycle[206,207]. Also common in T-ALL are chromosomal translocations that juxtapose transcription factors (TFs) with active T-cell-specific

promoters or enhancers, such as TCR regulatory elements[208,209], thus resulting in aberrant TF overexpression. Among the TFs frequently translocated and placed under the control of strong T-cell-specific promoters are TAL1, TAL2, LYL1, LMO1/2, TLX1/3, NKX2.1, and NKX2.2, among others (reviewed in[210,211]). Unlike *CDKN2A* loss, which is present in the majority of T-ALLs[203,205], mutations affecting these TFs appear to be typically mutually exclusive.

Also frequent in T-ALL are mutations resulting in constitutive activation of the NOTCH1 receptor[212]. Although NOTCH1 is not commonly overexpressed as a result of gross chromosomal abnormalities, the first evidence that it may have pathogenetic roles in T-ALL was, however, identified by a rare translocation event which placed a truncated form of *NOTCH1* adjacent to the *TCRB* locus[213]. The high frequency of *NOTCH1* mutation in T-ALL pathogenesis was not apparent until more than a decade later; this came with the discovery of mutations that resulted in ligand-independent NOTCH1 activation and/or increased stability of an active form of NOTCH1 among T-ALL cases[212]. Since this discovery, NOTCH1 has emerged as a predominant oncogene in T-ALL biology, with mutations affecting the Notch signaling pathway present in more than 60% of T-ALL cases[214,215] (reviewed in[216]).

1.4.2 Genotypic and phenotypic characterization of T-ALL

Diagnosis of T-ALL is based on a combination of immunophenotypic and morphological analysis, and to a lesser extent cytogenetic assessment (reviewed in[217]). Historically, this has resulted in the classification of T-ALL according to patterns of CD marker expression that are reminiscent of distinct T-cell differentiation stages[218]. The increasing prominence of gene expression profiling and whole-genome sequencing, however, have led to a greater appreciation of the molecular heterogeneity of T-ALL[219–224] (reviewed in[210,225]). Although genome-wide sequencing does not inform T-ALL diagnosis at present, molecular subtyping has emerged as a valuable classifier for greater understanding of disease biology, and has led to the identification of aggressive disease subgroups in T-ALL[226]. Further integration of methods currently employed for T-ALL characterization may improve understanding, and thus treatment, of disease. Two prevailing methods of T-ALL classification, attained through genetic subtyping and immunophenotypic analysis of blasts, are discussed here, although additional methods of disease characterization will also be highlighted.

1.4.2.1 Developmental stage-based classification

Immunophenotypic profiling

The current standard for T-ALL diagnosis and classification is achieved through immunophenotypic analysis of blasts by multi-channel flow cytometry, which allows for the simultaneous detection of several CD markers. Historically, classification of T-ALL was achieved by examination of markers such as CD3, CD1a, CD7, CD2, and CD8. The pattern of expression of these markers was used to orient leukemias along a normal T-cell developmental schema and thus resulted in classification into pro-T, pre-T, cortical, and mature T-ALL subsets[218].

Clinically, immunophenotyping presents a useful tool for leukemia subclassification and minimal residual disease (MRD) detection due to its broad accessibility. Moreover, as immunophenotypic subtypes show variable prognosis, immunophenotypic profiling may be of prognostic relevance. For example, the outcome of cortical T-ALL tend to show superior outcomes[227,228], while other subgroups are associated with worse outcome. Early T-cell precursor (ETP)-ALL is such an example; although first identified from gene expression profiling, ETP-ALL was initially described by CD1a and CD8 negativity, CD5 dim or negative expression, and expression of at least one myeloid or stem marker such as CD117, CD34, HLA-DR, CD13, CD33, CD11b, CD65[226,229], however several alternative proposals for immunophenotypic identification have since emerged[229–232]. ETP-ALL has been associated with unresponsiveness to standard therapies and poor prognosis in initial investigations[226,230], although in recent years improved outcomes have been achieved with more effective therapies[233,234].

Nonetheless, the utility of immunophenotyping as a prognostic indicator is limited by the number of CD markers that may be measured in parallel and thus precludes comprehensive, multi-dimensional phenotypic analysis of individual cells. This is most salient when considering incongruencies in phenotypic definitions of high-risk subgroups such as ETP-ALL, which may benefit from incorporation of additional defining markers for increased precision in disease classification[229–232]. Furthermore, the limited resolution of flow cytometry can result in conflation of genetically distinct T-ALL subgroups into broad developmentally based categories, as discussed in further detail in Chapter 2. Incorporation of sequencing technologies to inform disease subgroups, together with comprehensive immunophenotypic profiling, will improve disease classification and thus may expedite the development of targeted therapies.

T-cell receptor rearrangement status

Treatment efficacy can be monitored by MRD, but this requires highly sensitive and ideally non-invasive methods which can detect low levels of malignant cells. In T-ALL, detection of TCR rearrangements through PCR-based methods may be employed for MRD monitoring, and can achieve high degrees of sensitivity[202,235] (reviewed in[236]). Because TCR rearrangement is thought to occur in a stepwise manner that is intrinsically linked to developmental stage (discussed in detail in Chapter 1.2.4), it may also provide valuable information regarding the stage of arrest in T-ALL cases. Although not currently used in clinical practice to inform stage-based classification of T-ALL, the inclusion of TCR rearrangement status as an additional developmental marker may provide a finer degree of resolution to the current developmental stage-based classification[89].

1.4.2.2 Transcription factor subgroups

From the growing outputs of gene expression profiling and genome sequencing studies, it has become increasingly evident that T-ALL is a heterogenous disease that can be parsed into distinct molecular subgroups[219–224,237]. Each molecular subgroup is typically characterized by unique TF overexpression including, but not limited to, HOXA, LMO2/LYL1, TLX1/NKX2.1 (previously described as the proliferative cluster[219]), TLX3, and TAL/LMO[206,221] (reviewed in[225,238]). ETP-ALL, which lacks a unifying mutation[224,226], is an exception to this general observation, but can be distinguished on a molecular level by high levels of expression of progenitor-associated genes[226]. Overall, however, the gene expression profiles of each T-ALL subgroup are distinct and can be typified by the overexpression of a particular TF, often as a consequence of gross chromosomal aberrations.

Integration of phenotypic and genotypic profiling has contributed to the generally accepted notion that molecular TF subgroups may be associated with specific developmental stages. For example, ETP-ALL is often associated with arrest at the ETP stage of T-cell development owing to the unique compendia of markers expressed (i.e., immature T-cell markers in addition to myeloid/progenitor-associated markers) and a gene expression profile reminiscent of murine ETP cells[226]. The *TAL/LMO* T-ALL subgroup, at the opposite end of the differentiation spectrum, correlates with arrest at a late cortical stage of development due to expression of mature T-cell markers such as CD4, CD8, and CD3 (reviewed in[210]), and accordingly

demonstrates upregulation of genes important for specialized T-cell function such as *RAG* and *PTCRA* (reviewed in[239]).

The rapid rate at which sequencing, and more recently single-cell sequencing, technologies have advanced has led to a greater appreciation of the cellular and molecular contexts in which specific oncogenic TFs may operate, and further studies which integrate high-dimensional analyses of developmental, molecular, and epigenetic profiles will provide valuable mechanistic insight into T-ALL pathogenesis.

1.4.2.3 Mutations in epigenetic modifiers

Although genetic profiling has resulted in the classification of subtypes driven by dysregulation of TFs, mutations within epigenetic modulators also frequently occur in T-ALL[223,224] (reviewed in[240,241]). A landmark study focusing on the epigenetic landscape in over 21 different pediatric cancer contexts revealed that epigenetic regulators are among the most frequently mutated in T-ALL[242], pointing to the relevance of this protein class in T-ALL pathogenesis. Among the most commonly mutated epigenetic modulators in T-ALL are factors important for regulation of DNA methylation, such as *DNMT3A* (up to 20%[243–245]), *TET1* (up to 14%[241]), and *IDH1/2* (up to 9%[241]); polycomb repressive complex 2 (PRC2) members such as *EZH2*, *SUZ12*, and *EED* (up to 25%[241,246]); the histone H3K27 demethylase *UTX* (up to 12%[223,242]); the deubiquitinating enzyme *USP7* (~8%[242]), and *PHF6* (~16-38%[247]), among others (reviewed in[240,241]). The roles of some of these epigenetic modulators in T-ALL pathogenesis have only recently been investigated and are briefly highlighted here.

Mutations in PRC2 components

A high prevalence of loss of function mutations in PRC2 members, including *EZH2*, *SUZ12*, and *EED*, are observed in T-ALL[224,246,248,249]. In murine models, conditional loss of *Ezh2* resulted in spontaneous development of T-ALL[249], and in subsequent studies, deletion of *Ezh2* or *Eed* resulted in accelerated leukemogenesis[250], perhaps supporting a tumor suppressor role of this complex. In normal development, PRC2 establishes transcriptional repression through the deposition of repressive H3K27me3 histone marks (reviewed in[251]), and in the context of T-ALL, PRC2 and NOTCH1 may have antagonistic roles; studies in

T-ALL mouse models and human T-ALL cell lines indicate a reciprocal relationship between NOTCH1 and PRC2 in which the binding of NOTCH1 leads to EZH2 eviction and H3K27me3 loss[246]. This, in addition to enrichment of known PRC2 targets among Notch1 binding sites, perhaps indicate a co-operative role between inactivation of PRC2 components and oncogenic activation of Notch signaling[246].

UTX mutations

The prevalence of gene mutation among PRC2 components might underscore a critical balance of H3K27 methylation in T-ALL pathogenesis. This may be further understood from investigation of the UTX demethylase, which catalyzes demethylation of H3K27me3. The presence of *UTX* mutations in T-ALL, predicted to cause loss-of-function[252,253], are perhaps contradictory to the tumor suppressive role of PRC2 also reported in this disease context[246]. However, although the enzymatic activity of UTX and PRC2 components are diametrically opposed, they may exert their activity towards different loci (reviewed in[241]). The tumor suppressive role of UTX has been demonstrated in mouse models of T-ALL, in which *Utx* knockdown accelerated leukemia progression[252,253].

PHF6 mutations

Loss-of-function mutations in *PHF6* have long been recognized in T-ALL[247], but the functional consequences in this disease context have only recently been described and require further investigation. Initial studies in murine models of T-ALL suggest that *Phf6* loss may enhance hematopoietic stem cell renewal and may enrich for a leukemia-initiating cell population, but the molecular mechanisms driving these phenotypes have yet to be elucidated[254].

USP7 mutations

USP7 is frequently upregulated in T-ALL[255], and its suppression inhibits T-ALL proliferation in *in vivo* mouse models and human T-ALL cell lines[255,256]. This may occur by removal of ubiquitination marks and subsequent stabilization of NOTCH1 through the activity of USP7, which functions as a ubiquitin-specific protease[255,256]. These early studies might advocate for USP7 inhibition in T-ALL treatment. Nonetheless, reports of *USP7* mutations in T-ALL[223,242], purported to cause protein loss-of-function[223],

indicate otherwise, and altogether highlight that further investigation of USP7 in T-ALL pathogenesis is needed.

1.4.2.4 *DNMT3A* mutation and consequences of *DNMT3A* loss in T-ALL

DNMT3A mutations in T-ALL

Frequently, mutations affecting DNA methylation machinery are observed in T-ALL (reviewed in[240]); this most commonly involves the *de novo* DNA methyltransferase DNMT3A[244,257,258], but to a lesser extent mutations are also observed in DNA demethylases such as *TET1*[259] and *TET2*[260]. Rarely, mutations in *IDH1/2*, which are thought to result in inhibition of TET enzyme activity[261,262], are reported. In the T-cell lineage, mutations in *DNMT3A* are present in up to 20% of T-ALL cases[244,257,258] and are often enriched in ETP-ALL and immature T-ALL[243–245]. Mutations in the other *de novo* methyltransferase, *DNMT3B*, are rarely reported.

The majority of *DNMT3A* mutations reported in T-ALL are biallelic[244,263] and generally occur throughout the gene body[244,245,257]; this is in contrast to mutations observed in other hematological malignancies, such as acute myeloid leukemias, in which *DNMT3A* mutations tend to localize to R882[264–266]. In both disease contexts, however, *DNMT3A* mutation is thought to result in a loss of protein function[198,244,263] (reviewed in[198]). In a screen of more than 200 disease-associated *DNMT3A* mutations, 74% were implicated in a loss of protein function, with many variants resulting in reduced protein stability[267]. Among those mutations reported in T-ALL, approximately 88% decreased methyltransferase activity relative to wildtype *DNMT3A*[267]. These results, together with the mutational spectra observed in sequencing studies, implicate *DNMT3A* as a tumor suppressor in the context of T-ALL.

DNTM3A loss in T-ALL

Functional evidence that *DNMT3A* loss can predispose to T-ALL was demonstrated by studies in which deletion of *Dnmt3a* in the murine HSC compartment resulted in hematological malignancies that affected both the myeloid and T-cell lineages[268]. Although the majority of mice developed myeloid malignancies, development of T-ALL was also observed and moribund mice had enlarged thymuses, splenomegaly, and an accumulation of DP blasts in the bone marrow[268]. Progression to T-ALL was

associated with acquisition of additional mutations; frequently, activating mutations in *Notch1* were observed. This suggests that co-operation between *Dnmt3a* loss and *Notch1* activation may result in progression to T-ALL[268].

Further understanding of the interplay between DNMT3A and NOTCH1 in T-ALL could be understood from mouse models in which concurrent activation of *Notch1* and *Dnmt3a* deletion in HSCs resulted in rapid T-ALL progression[195]. The leukemias generated showed upregulation of several genes characteristic of myeloid lineage cells (e.g., *Elane, Mpo, Csf2rb and Gfi1b*), thus bearing resemblance to ETP-ALL. *Dnmt3a* knockout was also associated with global hypomethylation in comparison to *Dnmt3a* wildtype T-ALL and to normal DP T-cell populations, with the most prominent methylation loss observed in exons and enhancers[195]. These leukemias also showed resistance to apoptosis as a result of downregulation of the orphan nuclear receptor *Nr4a1*, which in a T-cell context can promote apoptosis through interaction with Bcl-2[195]. When expression of *Nr4a1* was induced using a chemical agonist, apoptosis was restored both in *Dnmt3a*-knockout murine thymocytes and in a human T-ALL cell line treated with shRNA targeting *DNMT3A*. The mechanism of *Nr4a1* downregulation upon *Dnmt3a* loss is yet to be elucidated[195], but further investigation may lend insight to the role of DNMT3A in normal T-cell development and leukemia progression.

Beyond the context of Notch, co-operation between DNMT3A and other oncogenes has been observed in murine leukemias generated by internal tandem duplication of FLT3 (FLT3-ITD)[269]. FLT3-ITD results in constitutive activation of the FLT3 cytokine receptor[270,271], and is commonly reported in hematological malignancies such as AML[272]. Mutations affecting FLT3 have also been observed in T-ALL[273]. In the context of *Dnmt3a*-null BM progenitors, expression of FLT3-ITD resulted in progression to T-ALL. The resulting leukemias resembled ETP-ALL based on CD marker expression and gene expression profile[269], and genome-wide DNA methylation analysis revealed global hypomethylation in comparison to DP T-cells or T-ALLs generated by FLT3-ITD alone. Notably, hypomethylation was pronounced in enhancer regions[269], consistent with previous reports in the context of murine T-ALL[195]. Many of the hypomethylated enhancer regions were identified to be putative binding sites for TFs including FLI1, GFI1B, and PU.1, possibly suggesting that DNMT3A acts as a tumor suppressor by regulating enhancer methylation of progenitor-associated genes[269].

Together, these functional studies emphasize the importance of DNMT3A in T-ALL pathogenesis. However, further studies are required to understand the complex role of DNMT3A during normal T-cell development and in disease progression. For example, upregulation of genes associated with progenitor/myeloid-lineage programs is consistently reported upon *Dnmt3a* loss[195,269], but the mechanism is unclear; this observation might suggest that *Dnmt3a* knockout induces a differentiation arrest at an ETP stage, or that *Dnmt3a* knockout results in derepression of genes associated with progenitor/myeloid-lineage programs. Yet, differentiation arrest was inconsistently induced[195], and DNA methylation changes do not consistently correlate with gene expression changes[173,194,195]. This paradox might suggest that there are additional roles of DNMT3A that are yet to be defined in T-cell development and in progression to T-ALL. A better understanding of how mutations may be operative in distinct developmental contexts will certainly lend insight into the molecular underpinnings of T-ALL pathogenesis and may yield better treatment of this disease for improved patient outcomes.

1.4.3 Current treatment and challenges

The standard of care for the majority of individuals with T-ALL is multi-agent chemotherapy, in which different treatment regimens are administered in three phases (i.e., induction, consolidation, and maintenance) aimed to prevent development of therapy resistance (reviewed in[274]). Treatment is often risk-intensified according to treatment response, which may be monitored by MRD assessment. MRD monitoring involves screening for the presence of T-ALL blasts remaining after the induction and consolidation phases. MRD response at the end of consolidation is a major determinant of prognosis in T-ALL, with MRD negativity indicating a generally favorable outcome (reviewed in[199]). Although strategies which implement risk-based treatment intensification have been met with success, particularly in subtypes associated with poor prognosis such as ETP-ALL (reviewed in[199]), long-term side effects from cytotoxic chemotherapy, in addition to dismal outcomes for relapsed disease (reviewed in[199]), advocate for the use of more targeted therapies.

Targeted therapy for use in T-ALL treatment is impeded by a lack of druggable targets. Inhibition of Notch signaling is an attractive candidate due to the oncogenic role of NOTCH1 in T-ALL, and a number of clinical trials have investigated NOTCH1 inhibition, primarily with γ-secretase inhibitors (GSIs). GSIs

prevent NOTCH1 activation through inhibition of the γ-secretase complex, which proteolytically cleaves and activates the NOTCH1 receptor (reviewed in[275]); in a subset of T-ALL cell lines, treatment with GSI induced G_0/G_1 cell cycle arrest[212]. In clinical trials, however, success has been limited due to significant gastrointestinal toxicity[276] (reviewed in[277]). Recent clinical trials aim to combine GSIs with additional agents, such as dexamethasone[278], to relieve off-target tissue effects. Alternative approaches to targeting the Notch signaling pathway include monoclonal antibodies directed against Notch[279], or by targeting factors downstream of Notch signaling, such as CXCR4 (reviewed in[280]).

Also frequently deregulated in T-ALL are cell cycle components; the majority of T-ALLs exhibit *CDKN2A* loss and/or *CCND3* overexpression, resulting in upregulation of *CDK4/6*[281] (reviewed in[282]). Thus, inhibition of the cell cycle regulators, CDK4 and CDK6, presents a possible therapeutic opportunity. In T-ALL mouse models and human cell lines, CDK4/6 inhibition induces cell-cycle arrest[282,283], and may act synergistically with glucocorticoids and other molecular inhibitors[283]. The use of CDK4/6 inhibitors is currently under investigation for the treatment of T-ALL, as well as other disease contexts in which CDK4/6 are upregulated (reviewed in[284]). There is, however, significant myelotoxicity and neutropenia associated with some CDK4/6 inhibitors, such as palbociclib, thus requiring stringent dose administration (reviewed in[285]).

In the absence of a pan-T-ALL target, precise classification of T-ALL subtypes may guide novel, subtype-specific therapeutic strategies. As a goal of targeted treatment is elimination of malignant cells with few adverse effects on normal tissues, it will be important to understand the specific developmental contexts in which various mutations operate. In T-ALL, mutation of certain factors correlates with particular stages of T-cell differentiation, which might suggest co-operation between stage of development and factor mutation, and furthermore may indicate a unique dependency of certain factors on the transcriptional programs that are rooted in development. This perspective would then imply that while mutations may drive disease, the mutations are intrinsically dependent on the developmental contexts in which they operate. Thus, a greater understanding of the complex interplay between various mutations and developmental programs may lend insight into targeting disease in a subtype-specific fashion.

1.5 Thesis objective and chapter summaries

T-ALL has long been characterized as a malignancy of developmentally arrested T-cells, but this central tenet has yet to be leveraged into improved therapies. The overarching objective of this thesis is to establish a greater understanding of how genetic and molecular events might operate in specific developmental contexts in T-ALL. The work presented in this thesis was motivated by the observation that there is synchrony between specific mutations and the developmental framework within which they are embedded, and that further understanding of disease pathogenesis could be understood by first probing the developmental contexts in which these factors operate. In the studies that will be discussed here, an emphasis was placed on understanding the role of *DNMT3A* in the context of early T-cell development. This was approached (1) by establishing a comprehensive developmental landscape of T-ALL, using normal intrathymic populations as a baseline for differentiation; (2) through the multi-omic and functional characterization of a chemically-defined *in vitro* T-cell model to then interrogate the consequence of *DNMT3A* loss on T-cell development, and (3) by the application of findings derived from *DNMT3A* loss in normal T-cell development to T-ALL.

In Chapter 2, I present analyses that characterize the developmental states of T-ALL through incorporation of RNA-seq data from normal human intrathymic subsets. To this end, I first generated a gene signature of normal T-cell development using publicly available expression data from sorted intrathymic subsets[286], and then applied this signature to two publicly available RNA-seq datasets of diagnostic T-ALL samples[222,223]. The goal of this chapter was to view T-ALL on a more granular level through the framework of normal T-cell development. The results of these analyses reveal that the use of high-dimensional data to classify developmental stage can improve resolution achieved from phenotypic analysis alone, and that this increased dimensionality can enhance sensitivity in detecting a subgroup of T-ALL that has previously been associated with poor prognosis. I also present observations on the coordination between specific transcription factor subgroups and the limited developmental contexts in which they appear. Overall, the results presented in this chapter underscores the value in approaching T-ALL classification with higher-dimensional methodology for improved understanding of the developmental contexts in which mutations may operate.

In Chapters 3 and 4, I characterize a chemically-defined model for *in vitro* T-cell development to then probe the consequence of *DNMT3A* loss in specific developmental contexts. This was achieved through multi-omic profiling (i.e., RNA-seq and whole-genome bisulfite sequencing) of T-cell subsets collected at various time points throughout differentiation, and through functional assessment of the developmental checkpoints that are characteristic of T-cell differentiation, such as T-cell commitment and T-cell receptor rearrangement. After a baseline of normal *in vitro* development was established, I then introduced genetic perturbation through CRISPR/Cas9-mediated knockout and shRNA-mediated knockdown experiments. Although *DNMT3A* loss did not demonstrably affect T-cell differentiation, I observed that knockdown and knockout of *DNMT3A* in committed T-cell populations was associated with an altered response to cytokines which typically contribute to the proliferation and differentiation of myeloid populations (‘myeloid’ cytokines). This was observed by an outgrowth of cells which appeared myeloid-like by immunophenotypic profiling. When this phenotype was examined in the context of human T-ALL cells, I observed that the addition of these ‘myeloid’ cytokines to culture media resulted in enhanced proliferation upon *DNMT3A* knockdown. The findings presented in this chapter shed novel insights into a possible role of DNMT3A in constraining lineage-specific signal responses in T-cell development and in T-ALL.

Chapter 2: Characterization of T-cell acute lymphoblastic leukemia in the context of normal T-cell development

2.1 Introduction

The processes that govern T-cell development must be tightly regulated to orchestrate the production and differentiation of T-cell progenitors into functionally mature subsets. In early T-cell ontogeny, a small number of hematopoietic progenitors emigrate from the bone marrow to the thymus. Exposure to the thymic environment restricts the alternative lineage potential of early T-lineage precursors and eventuates commitment to the T-cell fate[43,97]. The majority of cells adopt an $\alpha\beta$ T-cell fate, where commitment is punctuated by proliferative expansion of select cells in a process termed β-selection, and results in progression to $CD4^+CD8^+$ DP T-cells. The expansion of immature T-cells is highly controlled, and the survival of cells is dependent upon successful rearrangement of T-cell receptor loci. Errors throughout this process, which involves double stranded DNA breakage and re-ligation, can trigger gross chromosomal rearrangements that result in aberrant factor expression and ultimately lead to aberrant development and cellular transformation.

T-cell acute lymphoblastic leukemia (T-ALL) is the pathologic designation applied to malignant accumulations of immature T-cells. Although current chemotherapy achieves an event-free survival rate of approximately 85% over 5 years[199], long term side effects from cytotoxic chemotherapy and radiation, in addition to dismal outcomes for relapsed disease, advocate for the use of more targeted therapies and thus improved disease classification. Previous efforts to characterize T-ALL by immunophenotyping using a few select CD markers and TCR rearrangement status previously led to classification according to developmental stage. However, with the advent of gene expression/RNA profiling, classification by TF signature has become increasingly common. While TF clusters correspond roughly to normal T-cell developmental stages, there are frequent discordant samples that suggest additional genetic features influence the stage of developmental arrest.

The relevance of characterization by developmental stage has been highlighted in studies describing an ETP-like subgroup of T-ALL with a distinct immunophenotype ($CD1a^-$ $CD8^-$ and weak CD5,

in addition to expression of myeloid or stem markers including CD117, CD34, HLA-DR, CD13, CD33, CD11b, CD65). ETP-ALL was initially identified based on similarity to a murine ETP genetic signature[226], though additional studies have similarly pointed to T-ALLs marked by an early developmental arrest[89,206,221,248]. In some studies, ETP-ALL has been associated with unresponsiveness to standard induction therapy based on MRD, as well as poor prognosis[226,287]. These observations suggest a detailed, more highly resolved classification of T-ALL according to stage of differentiation arrest could provide added value in refining diagnostic subclassification and risk stratification.

Methods currently employed to infer T-cell developmental stage, such as clinical immunophenotyping or TCR rearrangement status, are relatively limited in dimensionality and thus are particularly sensitive to errors introduced by aberrant expression of individual component markers. Indeed, previous studies have highlighted T-ALL cases with a high degree of transcriptional similarity to the mouse ETP gene signature originally used to describe ETP-ALL, with only a subset of these cases fulfilling the immunophenotypic criteria defining ETP-ALL[231,288]. For example, Coustan-Smith *et al*. (2009)[226] reported a subgroup of T-ALL cases which were transcriptionally similar to the mouse ETP gene signature, but with higher expression of CD5, and this group has been subsequently classified as near-ETP-ALL[223]. The outcomes of near-ETP-ALL, however, are not well characterized. Taken together, this points to the need for a higher dimensional, more robust approach to developmental classification of T-ALL. To address this, we sought to derive a comprehensive developmental classification of T-ALL based on NGS-derived gene expression profiling data (i.e., RNA-seq) from purified subsets of normal human thymic T-cell progenitors.

By integrating publicly available RNA-seq datasets from normal intrathymic T-cell subsets and diagnostic T-ALL samples, we generated a developmental based classification of human T-ALL samples that closely recapitulates stages of progressive T-cell differentiation. We found that developmentally immature T-ALLs (corresponding to normal $CD34^{+}CD7^{+/-}CD1a^{-}CD4^{-}CD8^{-}$ T-cells) show inferior response to induction therapy and may thus help to identify high-risk patients. These immature T-ALLs share a common transcriptional signature with ETP-ALLs, but many do not fit immunophenotypic criteria for ETP-ALL, suggesting that a molecular signature of immature T-ALLs is more inclusive and captures a greater proportion of high-risk cases than the more restrictive immunophenotypic definition of ETP-ALL. In sum, we present a developmentally based classification of human T-ALL derived from high dimensional RNA-

seq data that defines clinically relevant disease subsets and exceeds upon existing lower dimensional strategies in clinical use such as flow cytometric immunophenotyping.

2.2 Materials and methods

2.2.1 Public data resources

The TARGET RNA-seq dataset as originally published by Liu et al. (2017)[223] was obtained from dbGaP under the accession numbers phs000218 and phs000464. The CGAD RNA-seq data as originally published by Chen et al. (2018)[222] was obtained from the Chinese Leukemia Genotype-Phenotype Archive under accession number CGAS00000000002. The RNA-seq dataset of human thymic populations as originally published by Casero *et al.* (2015) was obtained from the National Centre for Biotechnology Information Gene Expression Omnibus (GEO) database under the accession code GSE69239. The microarray dataset of human thymic populations as originally published by Lee *et al.* (2004)[289] was obtained from the GEO database under the accession code GSE1460.

2.2.2 RNA-seq alignment

RNA-seq data was quality assessed using FastQC (version 0.11.9). Quality read trimming was performed using Trimmomatic (version 0.40)[290]. Trimmed reads were aligned to the human reference genome (GRCh38/hg38) using STAR (version 2.7.8a) using default parameters[291]. To obtain read counts, the featureCounts function in Rsubread (version 2.8.2)[292] available in R (version 4.1.1) was used. Genes with $\leq$1 read count across all samples were removed for all downstream analysis. To identify differentially expressed genes, the DESeq2 package (version 1.34.0) in R (version 4.1.1) was used[293]. Differentially expressed genes were identified based on a BH-adjusted p-value < 0.05 and a $\log_2$ fold change $\geq$1 unless otherwise indicated.

2.2.3 Principal component analysis

We performed variance stabilizing transformation using the DESeq2 R Package (version 1.26.0) on protein-coding genes which were filtered to remove genes with low expression. We then performed principal component analysis using the top 500 variable genes. In cases where analysis was limited to

specific gene sets, we used the top 500 variable genes pertaining to that gene set. To perform data projection, we performed matrix multiplication of the samples to be projected (TARGET or CGAD cohorts) and the eigenvector matrix determined from the reference dataset (sorted thymic populations). To obtain a list of genes that contribute to variance, we extracted the gene loadings associated with the first two principal components.

2.2.4 Unsupervised clustering

We performed k-medoid and k-means clustering using the factoextra R Package (version 1.0.7). The number of optimal clusters (between 2-10 clusters) was determined using the silhouette and gap statistic. Clusterwise stability was determined by bootstrap resampling using the fpc package (version 2.2.7) and the cluster stability was determined by the mean Jaccard coefficient over 100 bootstrap iterations. Hierarchical clustering was performed using the hclust R package (R version 3.6.0) using Ward's linkage and Euclidean distance.

2.2.5 Differential gene expression

Differential expression analysis was performed using the DESeq2 R package (version 1.26.0) on read count data outputted from featureCounts as described above. To determine significantly differentially expressed genes, we used a threshold of a log2 fold change > 1 with FDR < 0.05, unless otherwise specified throughout the text.

2.2.6 T-cell receptor analysis

TCR rearrangements were determined from the RNA-Seq data using MiXCR as previously described (Kusakabe et al., 2019)[294]. We then filtered for unique sequences supported by at least 10 reads and comprising at least 2% of the total abundance per sample. T-ALLs were classified based on the extent of rearrangement observed; that is, a sample showing evidence of TRD, TRG and TRB rearrangement would be classified as TCRβ rearranged.

2.2.7 Transcription factor subgroup classification

We used a combination of chromosomal translocation and gene expression to classify patients from the CGAD cohort by transcription factor subgroup. Individuals were assigned to subgroups based on the presence of translocations affecting transcription factors frequently observed in T-ALL, as reported by the authors. We additionally classified individuals using an expression threshold corresponding to the top 10th percentile for each transcription factor.

2.2.8 Gene set variation analysis

Gene set variation analysis was performed using the GSVA package (version 1.34.0). Gene sets for each intrathymic subset (referred to as “Thy” throughout) were created by performing differential gene expression between the subset of interest and the remaining subsets (see Figure 2.2A). Enrichment of gene sets was determined using the default settings for GSVA. The enrichment scores for each gene set were visualized using heatmaps.

2.2.9 Diffusion map

To plot RNA-seq data from different sources onto the same diffusion map, we first performed rank normalization of the 2,824 genes implicated in T-cell differentiation using the CGAD, TARGET or Crooks datasets. The results were plotted using the destiny R package (version 3.0.1).

2.2.10 Statistical analysis

To determine the specificity and sensitivity of T-ALL classification methods (immunophenotypic ETP-ALL classification vs. differentiation signature classification), we performed paired McNemar’s tests. Patients were dichotomized into MRD positive or MRD negative categories using a threshold of $\geq$1% to indicate MRD positivity, or <1% to indicate MRD negativity. All statistical analyses were performed using R (version 3.6.0).

2.3 Results

2.3.1 Creation of a gene signature associated with human intrathymic T-cell development

We first focused our analysis on thymic populations which reflect the developmental window in which T-ALLs are arrested. We accessed publicly available RNA-seq data from human intrathymic subpopulations (Thy1-Thy6), which captures the progressive differentiation of normal T-cells[286]. Principal component analysis (PCA) of these subsets using the top 500 variable genes placed each subpopulation along an ordered differentiation trajectory, with the highest variance associated with differentiation along the first principal component (PC1) (Figure 2.1A). To pare down our analysis to the T-cell subsets which are most relevant to T-ALL, we projected patient T-ALL samples from two independent RNA-seq datasets accessed online (TARGET and CGAD studies) into PCA space defined by normal thymic subpopulations. We observed that T-ALL samples tend to fall among the immature T-cell subsets (Thy1-4), but not the terminally differentiated $SP4^+$ (Thy5) or $SP8^+$ (Thy6) populations, and thus determined Thy1-4 to be most suitable for our analyses (Figure 2.1B)[222,223].

We then defined gene expression patterns associated with T-cell differentiation by performing pairwise differential expression analysis among the various thymic subsets as shown in Figure 2.2A. We chose this approach, as opposed to a sequential pairwise DE analysis (ie: Thy1 vs Thy2, Thy2 vs Thy3, Thy3 vs Thy4), in order to capture the major gene expression changes that occur as cells establish and elaborate upon their developmental programs, since a sequential pairwise analysis might exclude relevant gene expression changes. This analysis revealed a set of gene expression signatures altogether comprising 2,824 unique genes. Among the 2,824 T-cell differentiation gene set, we grouped genes which co-varied across the Thy1-4 subsets using *k*-means clustering, yielding 5 distinct gene expression patterns (absolute $\log_2$ fold-change ≥ 1, BH-adjusted p-value < 0.05; Figure 2.2B); gene cluster 1-5 (GC1-5). Genes within each cluster showed high correlation (median Pearson correlation = 0.93), even when examined within an independent microarray dataset of sorted human thymic populations (McCune dataset; median Pearson correlation = 0.78), thus validating the composition of the gene expression clusters (Figure 2.2C)[289]. Among the five gene expression clusters, we noted downregulation of factors important for alternative lineage programs, such as *SPI1* (GC2), and upregulation of key regulators associated with T-

cell commitment and/or development including *PTCRA*, *RAG1/2*, and *GATA3* (GC4 and GC5) (Figure 2.2C).

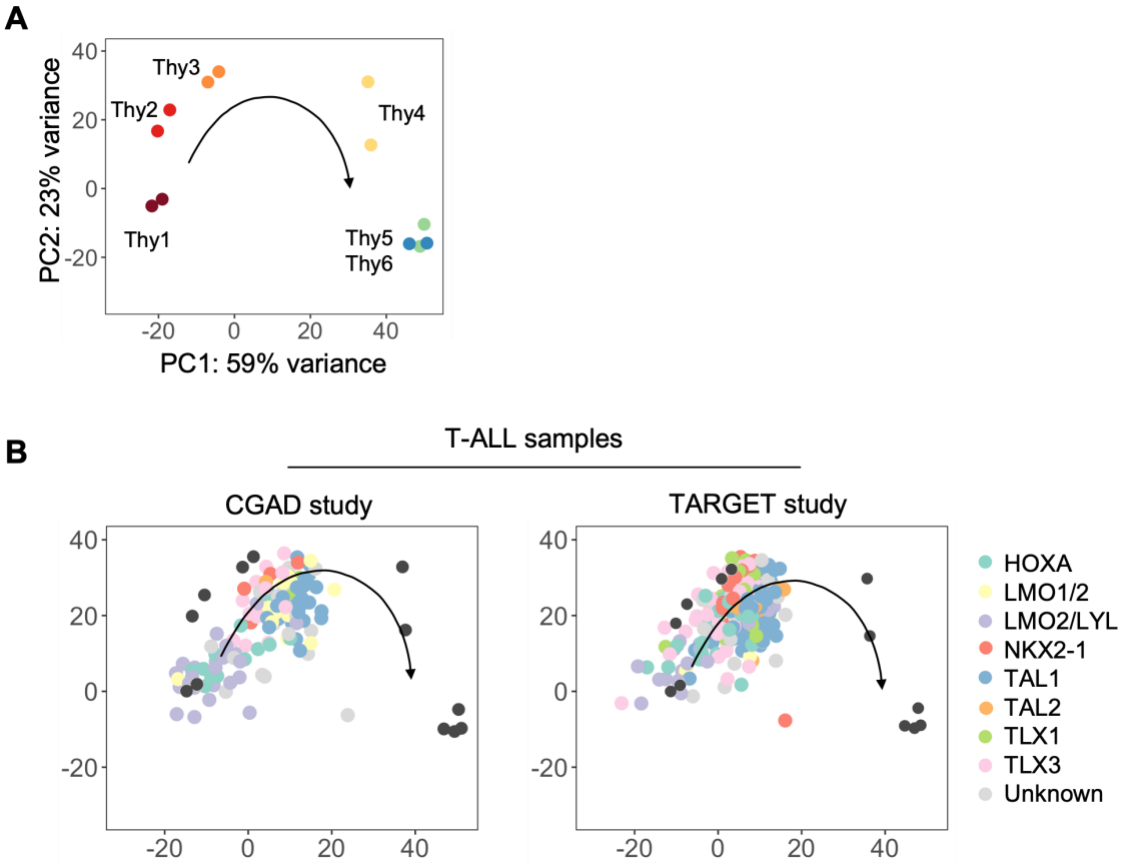


Figure 2.1 PCA projection of intrathymic T-cell subsets and T-ALL.

A PCA of sorted human thymus subsets as published by Casero *et al.*, 2015[286].
B Patient T-ALL samples from CGAD and TARGET studies (n = 129 and 264, respectively) projected into PCA space defined by normal T-cell subsets (Thy1-6). Black dots indicate the PCA projections of normal sorted thymic populations. T-ALL samples are coloured based on transcription factor (TF) subgroup.

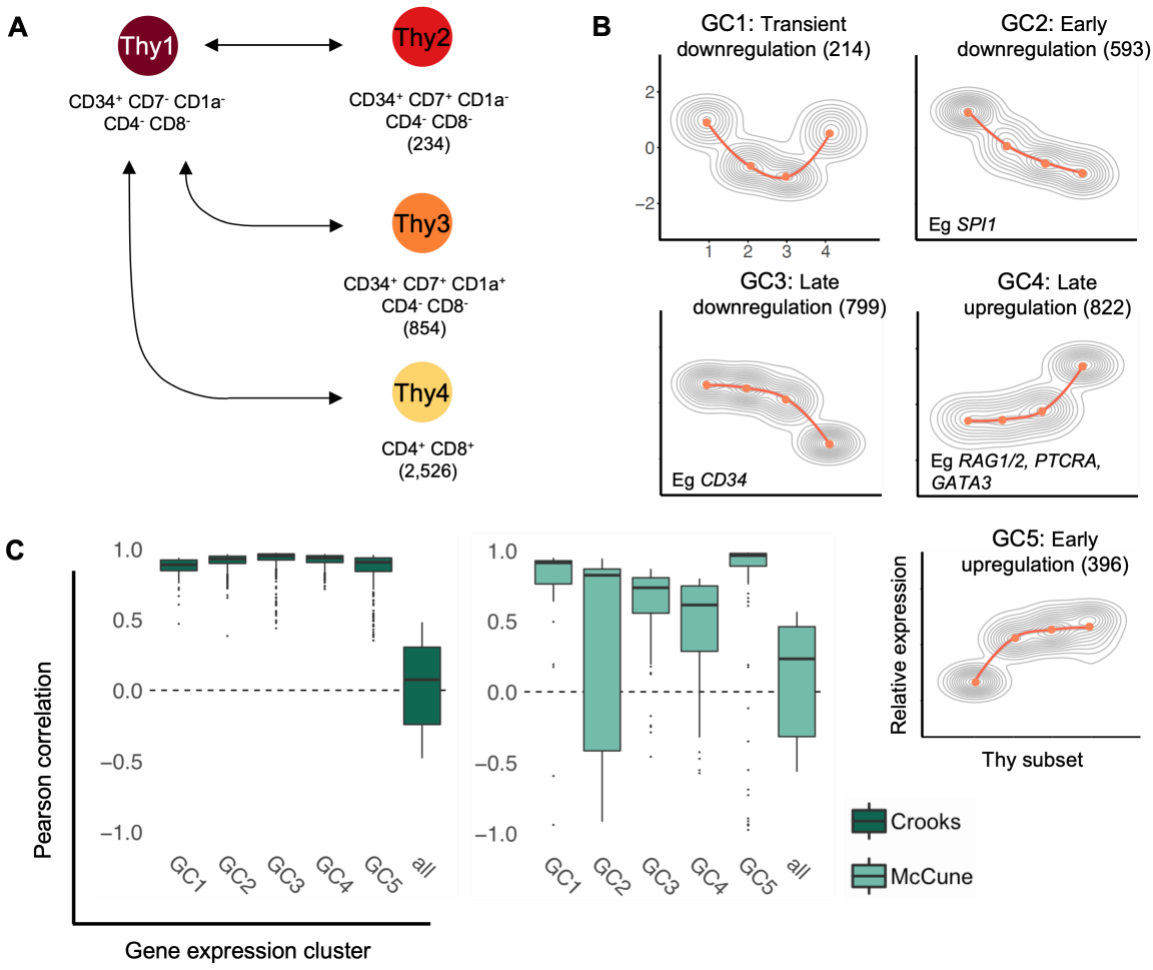


Figure 2.2 Cluster analysis of normal thymic T-cells.

A Differential gene expression comparisons employed to generate the 2,824 gene signature. The number of unique differentially expressed genes for each comparison is shown in brackets.
B k-means gene clustering of 2,824 genes implicated in T-cell development based on differential gene expression analysis. The relative gene expression is shown for each sorted human thymus subset (Thy1-4), with the mean expression for each gene cluster shown as a red line. The number of genes within each gene cluster is shown in brackets. GC = gene cluster.
C Intra-cluster gene correlation of the five k-means gene clusters of genes implicated in T-cell development. Intra-cluster correlations were also tested on an independent microarray of sorted thymic populations (McCune dataset) using the *k*-means gene clusters identified by Casero *et al*., 2015 (Crooks dataset)[286]. A high correlation is observed within each of the five k-means gene clusters (GC1-GC6) when compared to the correlation across all genes ("all").

2.3.2 Classification of T-ALL according to gene signatures that correlate with normal T-cell development

Having pared down to the subset of genes most informative with respect to thymic T-cell differentiation, we sought to parse the spectra of patient T-ALL samples against this background and, in essence, to perceive T-ALL though the lens of normal differentiation. We thus performed unsupervised *k*-medoid clustering of the TARGET and CGAD T-ALL datasets with analysis limited to the 2,824 gene set. Based on average silhouette width and gap statistic, we determined the datasets were optimally divided into 6 clusters (Figure 2.3A, Appendix G.1). The transcriptomic profiles of the sample clusters were highly correlated across both the TARGET and CGAD T-ALL datasets, supporting consistency of the cluster assignments (Figure 2.3B,C).

In an attempt to order the 6 clusters along a linear progression of intrathymic T-cell development, we looked for correlation with TCR rearrangement status as determined by MiXCR analysis of the RNA-seq reads[295]. Cluster Teal was notably enriched for samples showing no TCR rearrangements, while Cluster Yellow was most enriched for *TRA* rearrangements, suggesting that Teal and Yellow represent the extremes of most immature and mature stages, respectively (Figure 2.4, Appendix G.2)[87]. Further, ETP cases were strongly enriched in Cluster Teal, supporting its assignment as the most immature stage. Based on the next highest abundance of *TRA* rearranged samples, Clusters Green and Pink would presumably represent the 2nd/3rd most mature clusters, thus leaving Blue and Orange as 4th/5th most mature. Inspection of TF subgroup assignments suggested *LMO2/LYL1* cases were least mature, whereas *TAL1* cases were most mature.

We next attempted to define developmental order from RNA-seq data first by projecting normal T-cells onto the same PCA space as T-ALL cases. This revealed an arc-like trajectory of progressive differentiation states (Figure 2.5A, Appendix G.3). We also applied non-linear dimensionality reduction (Appendix G.4), which revealed a similar trajectory, thus supporting that a developmental-based ordering was possible. We thus proceeded to impute a differentiation-based order based on transcriptional signatures of the normal thymic subsets. We first defined gene sets corresponding to each of the Thy1-4 subsets (Appendix G.5), then used Gene Set Variation Analysis (GSVA)[296] to generate Thy1-4 signature scores for each of the samples in TARGET and CGAD datasets. We found that the 6 clusters could be

ordered along a developmental continuum based on these GSVA signature scores with Cluster Teal most aligned with Thy1/2, followed by Orange aligned with Thy2, then Blue with Thy3, Green with Thy3/4, and finally Yellow with Thy4 (Figure 2.5B,C, Appendix G.5). Cluster Pink was more difficult to place by this approach as its scoring pattern differed between the two datasets, aligning most with Thy1 in TARGET vs. somewhat ambiguously near Thy4 in CGAD. Incorporation of TCR rearrangement status information from both datasets would however support assignment of Cluster Pink to a more mature stage. This ambiguity in placement of Cluster Pink would suggest that it describes a subset of T-ALLs with more pronounced dysregulation of developmental gene expression programs.

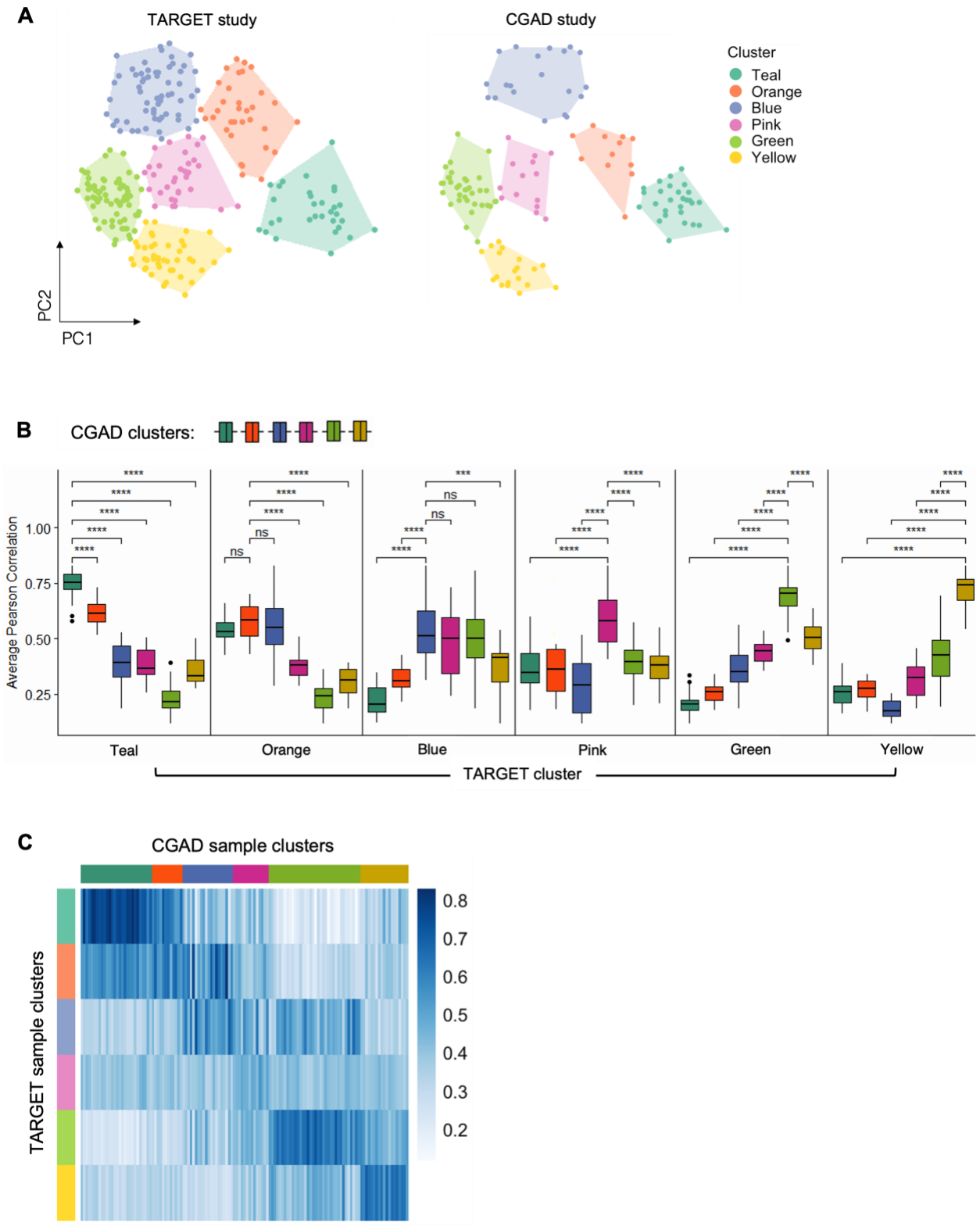


Figure 2.3 *k*-medoid clustering of T-ALL samples.

A PCA plot with optimal k=6 clustering for TARGET (left) and CGAD (right) T-ALL datasets. Cluster stability was calculated for each cluster based on 100 bootstrap iterations.
B Cluster-wise sample correlations across the TARGET and CGAD RNA-seq datasets. The Pearson correlation between each CGAD and TARGET T-ALL dataset is shown as a boxplot or heatmap and grouped by *k*-medoid cluster membership (t-test with Holm-Sidak correction for multiple comparisons). ns: not significant; *** p<0.001; **** p<0.0001.

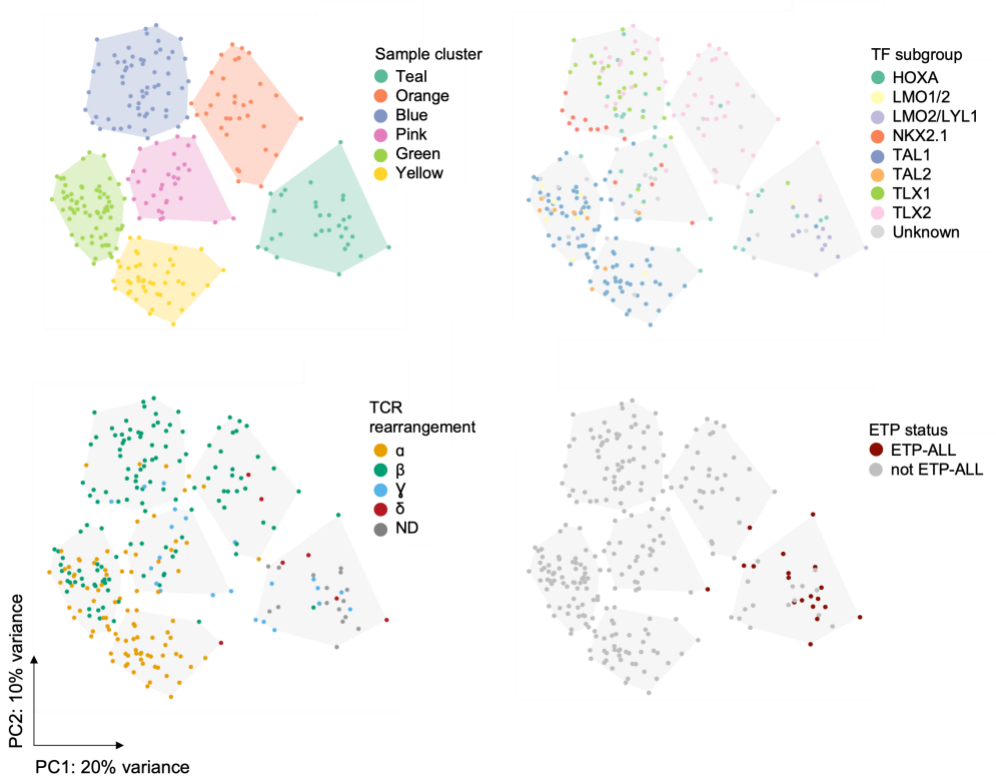


Figure 2.4 Developmental-based clustering of patient T-ALLs.

PCA plots of the TARGET T-ALL cohort (n=264) showing six *k*-medoid clusters in light grey with annotations according to transcription factor (TF) subgroup, T-cell receptor rearrangement (TCR) status, and ETP-ALL immunophenotype. ND = not detected.

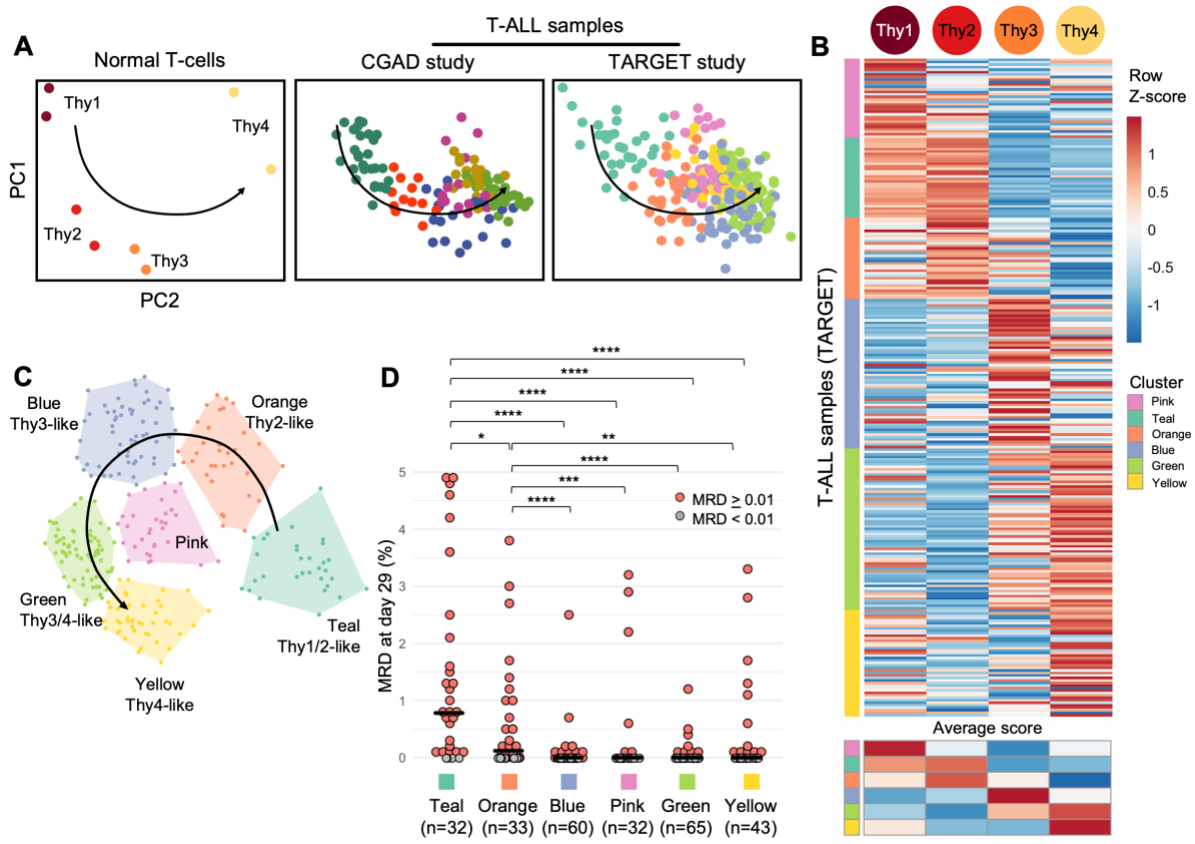


Figure 2.5 Developmental ordering and clinical significance of *k*-medoid clusters.

A Projection of CGAD and TARGET T-ALL cohorts into PCA space defined by normal T-cell subsets (Thy1-4). Samples are colored based on *k*-medoid cluster membership. The black arrow depicts an inferred developmental trajectory.
B Gene set variation analysis (GSVA) was performed on TARGET cohort to determine relative enrichment for each gene set (Thy1-4).
C Consensus developmental ordering of the six *k*-medoid sample clusters based on GSVA, TCR rearrangement and immunophenotype.
D Minimal residual disease (MRD) at 29 days post-induction therapy for the TARGET cohort. Horizontal bars indicate median MRD values for each subgroup. *, p<.05; **, p<.01; ***, p<.001; ****, p<.0001 (Dunn's test with Benjamini-Hochberg multiple test correction).

2.3.3 Developmental classification reveals patient subgroups with high minimal residual disease

Finally, we considered whether the 6 differentiation-based clusters held any clinical significance. Based on clinical outcome information available for the TARGET T-ALL dataset (n = 264), we found that the 6 clusters showed no significant association with event-free survival (Appendix G.6); however, we did find that Cluster Teal was associated with significantly higher MRD level at day 29 post-induction therapy

(Figure 2.5D). This is perhaps not surprising given the overlap with ETP-ALL which is already known to be associated with increased MRD (Appendix G.6); however, only 53% (17/32) of samples within Cluster Teal were immunophenotypically classified as ETP-ALL (Figure 2.4, Appendix G.6). Interestingly, segregation of Cluster Teal into ETP/near-ETP/non-ETP types revealed non-significant differences in MRD level (Appendix G.6), supporting that Cluster Teal identifies a greater proportion of high-risk patients than does ETP classification. To assess the relative utility of ETP-ALL vs. Cluster Teal designations, we performed a pairwise McNemar's test which revealed the Cluster Teal designation was significantly more sensitive than the ETP-ALL designation (p = 0.00079), while maintaining an equivalent degree of specificity (Table 2.1). These findings support that the RNA-seq-based differentiation classifier is capable of segregating patient T-ALL samples into clinically relevant subgroups.

Table 2.1 Sensitivity/specificity of Cluster Teal vs ETP-ALL classifier in identifying day 29 post-induction MRD positive cases.

The sensitivity (top) and specificity (bottom) of identifying day 29 post-induction minimal residual disease positive cases are shown. Statistical p-values are calculated by McNemar's test.

MRD positive on day 29 post-induction therapy (sensitivity):

	Cluster Teal group		
ETP-ALL group	Yes	No	
Yes	14	1	15 (16.13%)
No	14	64	78 (83.90%)
	28 (30.10%)	65 (69.90%)	93

Sensitivity difference: 14%
p = 0.00079

MRD negative on day 29 post-induction therapy (specificity):

	Cluster Teal group		
ETP-ALL group	Yes	No	
Yes	3	1	4 (2.33%)
No	1	167	168 (97.70%)
	4 (2.33%)	168 (97.70%)	172

Specificity difference: 0%
p = 1.0

2.4 Discussion

Our approach to characterize T-ALL according to developmental stage is ostensibly orthogonal to other studies which have emphasized classification based on distinct genetic transcription factor subgroups[219,222,223]. Indeed, some studies have specifically identified TF-mediated mechanisms of developmental arrest[297]; however, the analysis presented here emphasizes the clinical relevance of a dedicated approach to developmental-based classification of disease that is not fully captured by TF or CD marker-based clustering.

Although genetic/TF subgroup-based classification provides unique insight into transcriptional signatures that likely drive leukemogenesis, it remains difficult to segregate out the operative oncogenic effectors from those genes more generally associated with the prevailing stage of differentiation arrest. For example, among 18 *LMO2/LYL1* cases in the TARGET cohort, 17 segregated into either Cluster Teal (13) or Green (4) which reside at opposite ends of PC1 (Figure 2.5C), thus prompting investigation as to whether these TFs activate the same or different sets of genes in these two rather dissimilar developmental contexts. Further, since 3 of 4 cases in Cluster Green were associated with *LMO2* translocation (as compared to 0 of 13 in Cluster Teal) another issue may be whether the *LMO2/LYL1* signature in Cluster Teal is due to

oncogenic dysregulation or alternatively a secondary consequence of developmental arrest at a stage where *LMO2/LYL1* are normally expressed at high levels. The observation that certain genetic/TF subgroups tend to cluster within specific developmental windows may indicate a unique requirement of specific TFs for the developmental programs within which they are embedded. Distinguishing between those factors which necessarily depend on specific developmental contexts for survival, and those which are simply a consequence of developmental context, may have implications for differentiation stage-specific therapies, which have shown benefit in AML[298].

We note that the analyses presented here are based on bulk RNA-seq data and do not fully capture developmental or intra-tumor heterogeneity, which may be better illustrated by methods such as single-cell RNA-seq. We speculate that normal T-cell development represents a continuum of transient differentiation states, and that comparison of T-ALL cases to bulk thymus may result in artificial 'binning' of T-ALLs to the nearest normal counterparts as a result of capturing only a limited number of normal T-cell developmental states. Similarly, the existence of sub-clonal genetic events in T-ALL may influence, or be influenced by, the developmental context in which leukemia cells are arrested, and these dynamics are not fully reflected by bulk RNA sequencing of T-ALL. Our analyses suggest that T-ALL may be placed along a trajectory of aberrant T-cell development, which is in contrast to current classification methods which place T-ALL into discrete bins. Accordingly, it may be that even greater resolution afforded by single cell analyses will yield further refinement in the classification of tumors by revealing both inter- and intra-tumor heterogeneity.

In sum, patient T-ALLs can be clustered and ordered along a developmental trajectory with improved resolution using transcriptomic data and which outperforms both TF and CD marker-based clustering in prediction of MRD level, a highly relevant clinical outcome measure. We would postulate that the functional outcomes and/or sets of oncogenic effectors induced by a given TF potentially differ based on the developmental context of the cells in which they operate, and thus establishing a robust developmental schema may provide a useful backdrop in interpreting gene expression data and may facilitate design of differentiation state-specific therapies. Leveraging this information will, however, require the use of developmental models to interrogate the effects of genetic perturbations in specific developmental contexts.

Chapter 3: Normal human T-cell development using a feeder-free *in vitro* culture system

3.1 Introduction

To study the development of T-cells and the effect of genetic perturbations on normal T-lymphopoiesis, a robust system that effectively models the complex thymic environment and allows for genetic manipulation of developing T-cells is required. Although experimental mouse models provide a feasible means to explore and manipulate mechanisms underlying T-cell differentiation *in vivo*, several recent studies have highlighted differences between murine and human T-cell development (reviewed in[299]). These species-specific differences advocate for the use of human T-cell culture systems to study processes related to human T-cell development. While the most accurate models might be those that best approximate primary tissues, such as thymocytes isolated from human thymus, the use of primary tissue is restricted by limited availability and by challenges in manipulation of human thymocytes *in vitro*. Accordingly, many culture systems have instead focused on the generation of T-cells from human hematopoietic stem/progenitor cells *in vitro.*

Early iterations of *in vitro* human T-cell culture were limited by the use of cumbersome fetal thymic organ cultures (FTOCs), which required careful extraction of thymic lobes from a fetal mouse and reconstitution with human progenitors[300]. Ground-breaking efforts have since shown that T-cells can be derived *in vitro* by culturing human progenitor cells from various sources (e.g., umbilical cord blood, bone marrow, and mobilized peripheral blood) on a monolayer of murine OP9 stromal cells that overexpress Notch ligands (e.g., Delta-like ligand 1; DLL1) and are thus termed OP9-DL1 cells[29,301,302]. While these cultures may be supplemented with exogenous cytokines to promote T-cell differentiation, such as IL-7, SCF, and Flt3L, parental OP9 cells themselves also secrete factors which promote T-lymphocyte development, including IL-7, SCF, and CXCR12[303,304]. The OP9-DL1 system efficiently supports the production of $CD4^+CD8^+$ double positive (DP) T-cells, and to a lesser extent, mature single positive (SP) $CD8^+$ T-cells with rearranged T-cell receptors (TCRs)[305]. Importantly, these *in vitro*-generated T-cells may be readily altered and thus provide a foundation for T-cell production and for exploring processes related to T-cell development. The ease of manipulation afforded by *in vitro* models, either through genetic

perturbation or by addition of exogenous reagents, thus makes them an attractive choice for modeling T-cell development to study the contexts in which normal development may go astray.

Although the OP9-DL1 system represents a major breakthrough in the *in vitro* generation of T-cells and has revolutionized *in vitro* T-cell culture, it does not come without limitations. The simplicity of the OP9-DL1 monolayer system translates to convenience in cell culture and manipulation, but does not fully appreciate the complex three-dimensional architecture in which normal T-cells develop, nor the distinct cellular zones that are characteristic of the thymus (reviewed in[24]), and thus may not be suitable for modeling processes such as intrathymic T-cell migration. Furthermore, deviations between variant OP9-DL1 cell lines as a result of extensive *in vitro* culture may introduce additional uncontrollable heterogeneity; variations in culture conditions, such as the levels of expression of Notch ligand by feeder lines, can elicit demonstrable differences in T-cell outputs[306]. This may be further compounded by the use of fetal bovine serum, which varies in composition across different batches and sources. Taken together, these limitations may culminate in defects in faithful T-cell differentiation; for example, while the OP9-DL1 system can efficiently generate $CD8^+$ SP cells, the $CD4^+$ SP output is hindered[307] (reviewed in[308,309]).

The development of alternative *in vitro* T-cell differentiation models in more recent years, such as 3D organoid systems, may mitigate these effects by more accurately mimicking the complex thymic architecture and hold promise in generating T-cell populations which faithfully resemble those observed in human thymus[300,310–314]. The use of these systems have also achieved noticeable improvement over the OP9-DL1 system in the generation of DP and SP cells[312]. Nonetheless, these models remain limited by the use of stromal cells (e.g., artificial thymic organoids) or primary thymic tissue which may deliver exogenous factors of variable composition and concentration. Although 3D organoid systems are an invaluable tool for modeling T-cell development within thymic tissue, the additional uncontrolled variability introduced by variant feeder lines or primary thymic tissue may confound interpretations when attempting to assess the effect of genetic or molecular manipulation on T-cell development. For this purpose, culture systems with defined components may be more suitable for certain applications.

Although primarily developed for clinical use, several groups have developed chemically-defined, feeder-free conditions that support the generation and expansion of T-cells *in vitro* from umbilical cord blood (CB)-derived progenitors[315–317]. These systems typically include recombinant human Notch ligand bound to

a tissue culture plate and cytokines that encourage T-lineage growth, such as SCF, IL-7, and Flt3L, and provide a scalable means for efficient and reproducible T-cell maturation. Similar to the OP9-DL1 system, these are monolayer culture systems and thus are not suitable for modeling processes related to T-cell migration, but given the tractability and use of synthetic media, may be valuable for modeling the consequences of specific genetic perturbations. Another important limitation of these cultures systems is the relative paucity of SP8 and SP4 T-cells, which is likely due to the absence of thymic epithelial cells bearing MHC molecules which are required for positive and negative selection.

In the studies described here, T-cells were generated using an off-the-shelf feeder-free culture system. This culture system yields large numbers of T-cells from $CD34^+$ progenitors from human CB which differentiate in a reproducible manner and are amenable to genetic manipulation. T-cells generated from this system may be captured in large quantities at any stage of T-cell development, therefore allowing precise dissection of events in early T-cell ontogeny, which can be difficult to capture due to limited cell numbers in primary tissues. Importantly, the T-cells derived from this system resemble human thymic populations both phenotypically (based on transcriptomic and immunophenotypic analyses), and functionally (based on the presence of TCR rearrangements and T-cell lineage commitment), and thus may be suitable for use to further understand factors important for early events in T-cell differentiation. As these cells are easily manipulated (i.e., by shRNA-mediated knockdown or CRISPR-mediated modulation), this system may be used to explore the effect of genetic perturbation on T-cell development, as discussed in subsequent text. This chapter describes the molecular and phenotypic characterization of T-cells generated from this system in comparison to primary intrathymic T-cell populations through the use of RNA-seq, whole genome bisulfite sequencing, and immunophenotypic profiling.

3.2 Materials and methods

3.2.1 Cells and cell culture

3.2.1.1 Primary cells

Human umbilical cord blood samples were obtained from Stem Cell Assay (Terry Fox Laboratory, BC Cancer Research Centre, Vancouver, British Columbia, Canada) with informed consent from donors. These samples contained pools of cord blood samples from multiple donors (2-10 individual donors). A list

of cord blood pools accessed, and the experiments performed for each cord blood pool, can be found in Appendix A.

3.2.1.2 Primary cell culture

Cells were cultured *in vitro* using the StemSpan T cell Generation Kit (STEMCELL Technologies) according to manufacturer's instructions and maintained in a 37°C incubator with 5% CO_2. Briefly, non-treated culture dishes were coated with StemSpan Lymphoid Differentiation Coating Material (100X), diluted 1:100 in PBS and left overnight at 4°C, or at room temperature for 2 hours, before the coating material was removed and plates were rinsed with PBS. Cord blood samples were resuspended in StemSpan Lymphoid Progenitor Expansion Supplement (10X), diluted 1:10 in StemSpan Serum-Free Expansion Medium II (SFEM II, STEMCELL Technologies), or in StemSpan T cell Progenitor Maturation Supplement (10X), diluted 1:10 in SFEM II (STEMCELL Technologies), according to manufacturer's protocol. For the myeloid media transfer experiments, T-cells were sorted into non-treated culture dishes containing StemSpan Myeloid Expansion Supplement (100X) or StemSpan Myeloid Expansion Supplement II (100X), diluted 1:100 in StemSpan SFEM II (STEMCELL Technologies).

3.2.2 Limited dilution analysis

Cells were sorted at various cell concentrations into non-treated 96-well round bottom plates (Falcon). For all experiments, 3-8 replicates were performed for each media condition and cell concentration. For populations cultured in conditions supportive of granulocytic cell expansion, cells were sorted into wells containing StemSpan Myeloid Expansion Supplement (100X), diluted 1:100 in StemSpan SFEM II (STEMCELL Technologies). For populations cultured in conditions promoting monocyte cell expansion, cells were sorted into wells containing StemSpan Myeloid Expansion Supplement II (100X), diluted 1:100 in StemSpan SFEM II (STEMCELL Technologies). For populations cultured in conditions supportive of T-cell differentiation, cells were sorted into wells coated with StemSpan Lymphoid Differentiation Coating Material (100X), diluted 1:100 in PBS, containing StemSpan Lymphoid Progenitor Expansion Supplement (10X), diluted 1:10 in StemSpan SFEM II (STEMCELL Technologies). Data analysis

was performed using R (version 4.1.1) using the elda package (version 1.4.36) using default parameters except for “observed”, which was set to “TRUE”.

3.2.3 Flow cytometry and cell sorting

3.2.3.1 Flow cytometry

Samples were stained with fluorochrome-conjugated antibodies as listed in Appendix B.1 and B.6. Data was acquired on the BD LSRFortessa Cell Analyzer (BD Biosciences) and the FACSymphony (BD Biosciences). To obtain cell numbers for each sample, AccuCheck counting beads (Invitrogen/ThermoFisher Scientific) were added prior to sample acquisition. In experiments where samples were acquired over several weeks, per-channel normalization was performed by staining a bead standard with the antibody mastermix at each timepoint.

3.2.3.2 Cell sorting

Samples were stained with fluorochrome-conjugated antibodies as listed in Appendix B.4 and B.5. Cell sorting was performed on the BD FACSAria II (BD Biosciences) and the BD FACSAria Fusion (BD Biosciences). For samples collected for RNA-seq and whole-genome bisulfite sequencing, cell populations were cultured *in vitro* as per the methods outlined in Section 3.2.1.2. At different indicated time points, cells were harvested and sorted according to the markers as listed in Appendix B.4, Appendix C, and Figure 3.3. Sorted populations were pelleted and snap frozen pellets were processed by the Genome Sciences Centre (Vancouver, British Columbia, Canada) for RNA/DNA extraction as outlined in Section 3.2.4.1 and Section 3.2.6.1. For the isolation of DN and DP thymus subsets, CD4 and CD8 selection was performed using the EasySep Human CD8+ T Cell Positive Selection Kit II and the EasySep Human CD4+ T Cell Positive Selection Kit II (STEMCELL Technologies). The eluent (DP-debulked) and magnetic-bound (DP-enriched) fractions were subsequently sorted by FACs.

3.2.3.3 Flow cytometry data analysis

Data was analyzed using FlowJo v10 software (Tree Star) and in R (version 4.1.1) using custom scripts. Single-cell data collected from 16 dimensions (FSC-A, SSC-A, CD4, CD8, CD7, CD5, CD1a, CD3,

CD13, CD33, HLADR, CD15, CD11b, CD44, CD34, and CD38) were reduced to 2 dimensions using the R package 'Monocle3' (version 1.0.0)[318] or 'umap' (version 0.2.28). Scale data was exported from FlowJo and was subsequently transformed using the arcsinh function from the 'flowCore' package (version 2.6.0)[319] in R (version 4.1.1) using the cofactor set to 150. For experiments where data was acquired on separate days, per-channel normalization was performed prior to archsinh transformation. Sample normalization was performed by collecting a bead standard stained with the antibody master mix at each timepoint. For generation of UMAP projections, data were generated using the predict function in 'umap' (version 0.2.28). All data was visualized using ggplot2 (version 3.3.6).

3.2.4 RNA-seq

3.2.4.1 RNA-seq sample processing

RNA-seq samples were processed as part of a collaboration with the Canadian Epigenetics, Environment and Health Research Consortium (CEEHRC) by the Genome Sciences Centre (GSC) in Vancouver, British Columbia, Canada. Briefly, polyadenylated (polyA+) RNA was purified from total RNA obtained from sorted flash frozen pellets using the NEBNext Poly(A) mRNA Magnetic Isolation Module from New England Biolabs. Complementary DNA (cDNA) was then generated from the purified polyA+ RNA for strand-specific sequencing. First-strand cDNA synthesis was performed using Maxima H Minus Reverse Transcriptase (Thermo Fisher) according to manufacturer's protocol but with the addition of 0.04μg/μL Actinomycin D and random primers. Second-strand cDNA synthesis was performed using NEBNext Second Strand Synthesis Module (New England Biolabs). Bead clean-up was performed following first-strand and second-strand cDNA synthesis using Ampure XP Beads (Beckman Coulter). Libraries were sequenced on the Illumina HiSeq 2500 instrument (150bp paired-end reads) according to manufacturer's protocols (Illumina, Haward, CA). On average, approximately 100-200 million reads per sample were achieved.

3.2.4.2 RNA-seq analysis

RNA-seq reads which passed the Illumina Chastity filtering were quality assessed using FastQC (version 0.11.9). Quality read trimming was performed using 'Trimmomatic' (version 0.40)[290]. Trimmed reads were aligned to the human reference genome (GRCh38/hg38) using 'STAR' (version 2.7.8a) using

default parameters[291]. To obtain read counts, the 'featureCounts' function in 'Rsubread' (version 2.8.2)[292] available in R (version 4.1.1) was used. Genes with $\leq$1 read count across all samples were removed for all downstream analysis. To identify differentially expressed genes, the 'DESeq2' package (version 1.34.0) in R (version 4.1.1) was used[293]. Differentially expressed genes were identified based on a BH-adjusted p-value < 0.05 and a $\log_2$ fold change $\geq$1 unless otherwise indicated. Clustering of gene expression patterns was performed using the R package 'MBCluster.Seq' (version 1.0)[320]. All data was visualized using ggplot2 (version 3.3.6).

3.2.4.3 Public data resources

The RNA-seq dataset of human thymic populations as originally published by Casero *et al.* (2015)[286] can be obtained from the National Centre for Biotechnology Information Gene Expression Omnibus (GEO) database under the accession code GSE69239.

3.2.5 TCR analysis

TCR rearrangements were determined from the RNA-seq data using MiXCR[295], as previously described[294,321], and visualized using ggplot2 (version 3.3.6).

3.2.6 Whole-genome bisulfite sequencing

3.2.6.1 WGBS sample processing

Samples for WGBS were processed as part of a collaboration with CEEHRC by the GSC in Vancouver, British Columbia, Canada. Briefly, 1μg of DNA was spiked using 10ng of unmethylated lambda DNA (Promega) to measure the efficiency of bisulfite conversion. Samples were sheared to an average fragment size of ~300bp by sonication using a Covaris LE220. End-repair was performed using the End-Repair Premix (New England Biolabs) at 20°C for 30 minutes. Following end repair, A-tailing was then performed using the dA Tailing Reaction Mix (New England Biolabs) at 37°C for 30 minutes, 70°C for 5 minutes, and 4°C for 5 minutes. Methylated PE adapters were then ligated to the end-repaired DNA using the 2X NEB Ligation Premix (New England Biolabs). The methylated adapters were modified to contain 5-methylcytosine bases in place of cytosine bases. Bisulfite conversion was then performed using the EZ

DNA Methylation-Gold Kit (Zymo Research). Enrichment for adapter-ligated DNA was performed with 5 cycles of PCR using KAPA HiFi Uracil+ ReadyMix (Roche) using the following protocol: 98°C 1min; 5 cycles of 98°C 15sec, 65°C 30sec, 72°C 30sec; 72°C 5min. The final library was analyzed on the Agilent Bioanalyzer and quantitated by Quant-iT. Bead clean-up was performed between each sample preparation step using PCRClean DX (Aline Biosciences). Paired-end DNA sequencing (2x125nt) was then performed using Illumina sequencing technology.

3.2.6.2 WGBS sample analysis

Reads were mapped to the human reference genome (version GRCh38/hg38) and estimation of CpG methylation was performed using 'gemBS' (version 3.5.0)[322] using default parameters. Analysis was restricted to uniquely mapped reads with a minimum MAPQ score > 20. CpGs overlapping with known SNPs, present on XY chromosomes, or supported by fewer than 10 reads were filtered prior out to downstream analysis. Only CpG sites with methylation estimates present in all samples were included in downstream analysis. Approximately 24 million CpGs passed filtering for all samples, with an average mean depth per between 22-28X. Functional annotation of gene regulatory regions was performed using custom scripts. CpG methylation of promoters was determined by calculating the average CpG methylation of all CpG sites located 1500bp upstream and 500bp downstream of the transcriptional start site of protein-coding genes. The location of CpG islands was determined from the UCSC genome browser. Genome-wide CpG methylation was determined by taking the average CpG methylation of all CpG sites within 5kb bins. Plots indicating the level of methylation of individual CpGs in genome browser view were generated using the Gviz package (version 1.38.4) in R(version 4.1.1).

3.3 Results

3.3.1 Phenotypes of *in vitro* CB-derived T-cell populations

To characterize the phenotype of T-cells generated *in vitro* from $CD34^+$ CB, I collected cells cultured in feeder-free T-cell expansion media for analysis by flow cytometry using a panel of markers chosen to resolve various early T-cell populations (Appendix B.6). Included among the markers analyzed were CD7, CD5, CD1a, CD4, and CD8, which are sequentially expressed in developing thymocytes (Figure 1.1). Cells

were collected at several timepoints (up to day 42 of *in vitro* culture) to assess the kinetics of marker acquisition over the course of *in vitro* differentiation. From these experiments, I observed expression of CD7 and CD5 markers by day 14, followed by CD1a expression by day 28, and finally surface expression of TCR$\alpha\beta$ by the end of the culture period (Figure 3.1). Additional repetitions with more timepoints on a subset of surface markers indicated that the kinetics of CD7/CD5 gain and CD44 loss is reproducible, even when using a different cord blood pool, and furthermore demonstrate the acquisition of CD7 prior to CD5 at day 7 (Appendix F). These results are in accordance with previously published reports using the same culture system[323].

Importantly, the feeder-free culture conditions supported the growth of double positive (DP) $CD4^+CD8^+$ T-cell populations, with the majority of cells displaying CD4 and CD8 co-expression by day 42. This contrasts with T-cells generated using the OP9-DL1 system, where very few DP cells were observed after 42 days in culture (Ann Sun, personal communication; Figure 3.2). While it is possible this discrepancy could be due to a delay in the kinetics of marker acquisition in the OP9-DL1 culture system, no DP cells were present even when cultured for an additional 9 days (Ann Sun, personal communication; Figure 3.2), suggesting that, at least in our hands, the OP9-DL1 culture system does not support the full range of T-cell differentiation. This is in contrast to previous reports, where DP generation from $CD34^+$ CB can be first detected after ~24 days of culture on OP9-DL1 feeders[29]. This discrepancy may point to variations in OP9-DL1 feeder lines or culture conditions between different research settings.

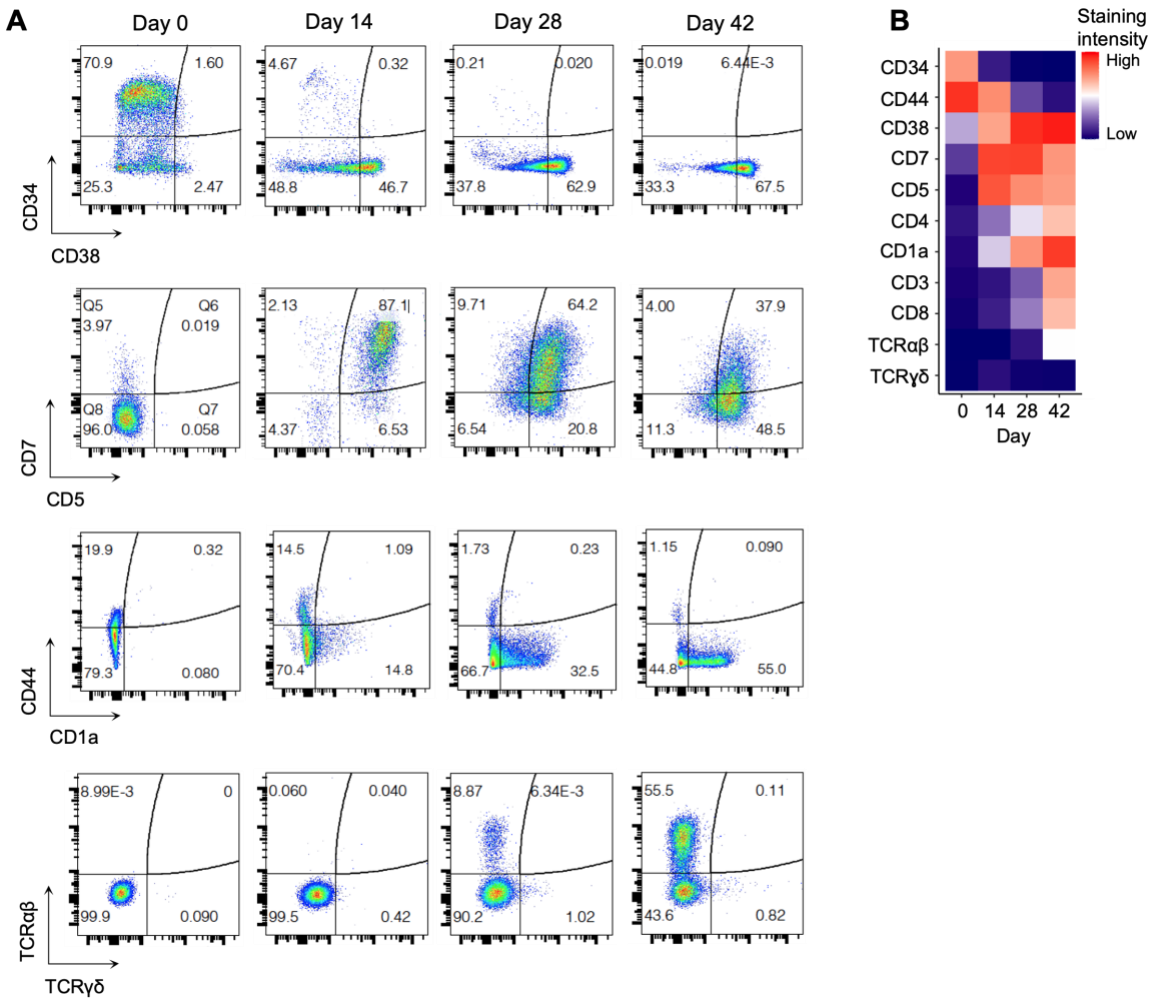


Figure 3.1 Flow cytometric analysis of T-cell subsets derived *in vitro* from CD34+ cord blood.

A Pseudocoloured flow cytometry plots representative of 2 different cord blood donors. Cells were harvested for flow cytometry at the indicated time points throughout *in vitro* T-cell differentiation (day 0, day 14, day 28 and day 42).
B Heatmap representation of CD marker expression at time point.

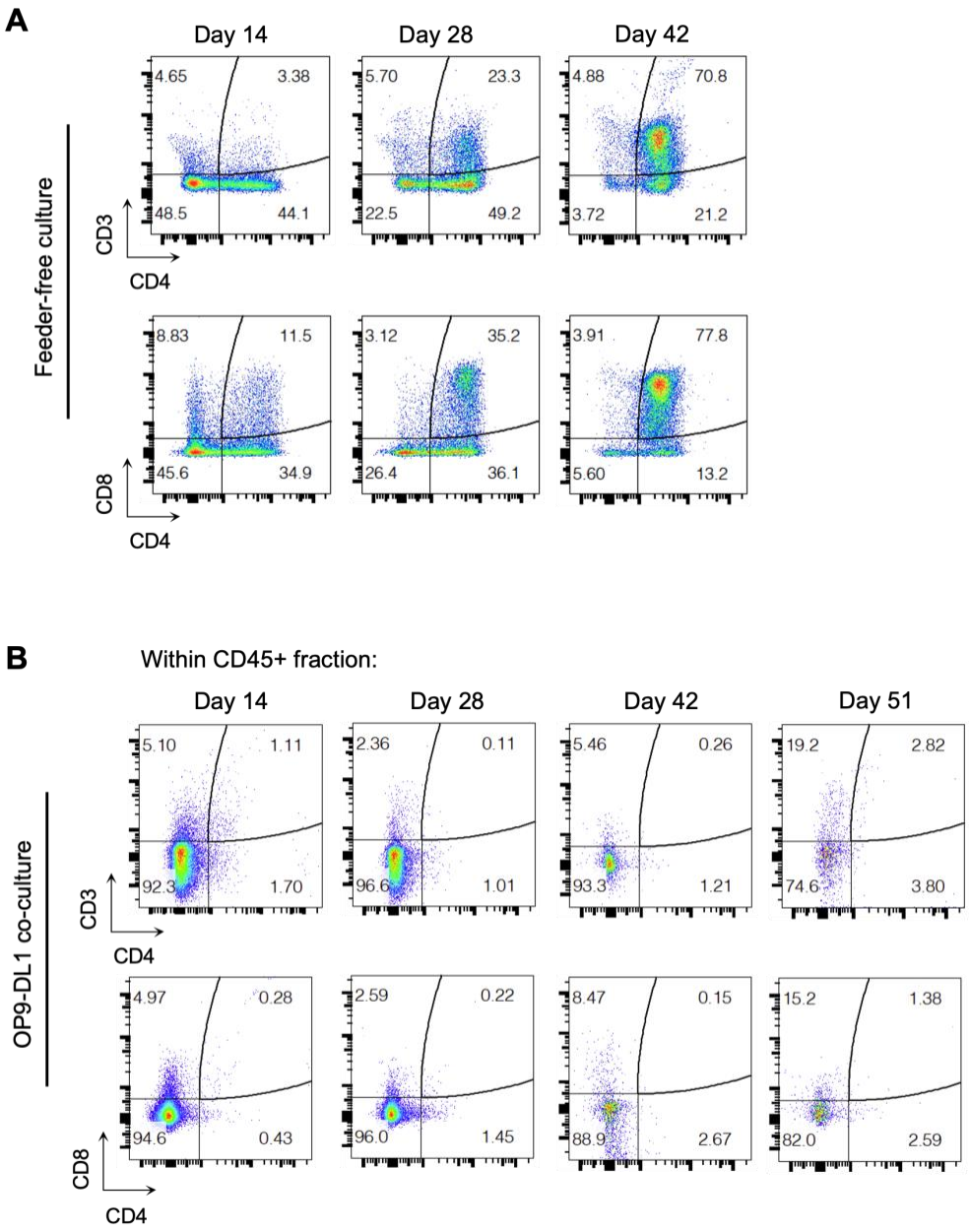


Figure 3.2 Comparison of surface marker expression of cord blood-derived T-cell subsets generated in feeder-free or OP9-DL1 culture systems.

A *In vitro*-derived T-cells generated in a feeder-free culture system.
B *In vitro*-derived T-cells generated from the OP9-DL1 co-culture system.

3.3.2 *In vitro* generated T-cell subsets are transcriptionally similar to human thymic populations

To gain an understanding of the transcriptional changes that occur throughout differentiation, T-cell subsets were isolated based on several surface markers at various timepoints throughout *in vitro* culture for RNA-seq and whole genome bisulfite sequencing (WGBS) analysis as shown in Figure 3.3A (top panel) and Appendix C. Principal component analysis of the top variably expressed genes among these subsets indicated that samples arrange according to the number of days cultured *in vitro* (Figure 3.3B). To determine if our T-cell subsets recapitulate aspects of T-cell development as observed in human thymocyte populations, I next sought to compare RNA-seq expression profiles of our *in vitro* generated T-cells to a publicly available RNA-seq dataset of sorted T-cell subsets isolated from human thymus generated by Casero *et al*. (2015)[286], as previously described in Chapter 2 (Figure 3.3A, bottom panel). To first determine which intrathymic populations corresponded to each of our *in vitro* generated T-cell populations, subsets across both datasets were roughly matched based on immunophenotype as shown in Figure 3.3A. Projection of both datasets into the same principal component space showed similarity along PC1 among the matched samples, suggesting that the sample comparisons were appropriate (Figure 3.3C).

Having determined the closest developmental equivalents between our *in vitro*-derived T-cell subsets and human thymocytes populations, we next examined the gene expression changes associated with T-cell development. Focusing on a set of 2,824 genes related to T-cell development as determined from intrathymic populations (discussed in Chapter 2), I first clustered the 2,824 gene set into a series of 18 expression patterns. These expression pattern clusters were then compared to the *in vitro* generated T-cell populations (Figure 3.4A). Overall, there were similarities in the gene expression patterns of the *in vitro* generated T-cells and the intrathymic subsets throughout development. When this analysis was extended to a wider selection of genes (i.e., not limited to the 2,824 gene set), a high correlation among the matched populations was observed (Figure 3.4B). These results suggest that the *in vitro*-derived T-cell subsets recapitulate certain aspects of human thymic populations.

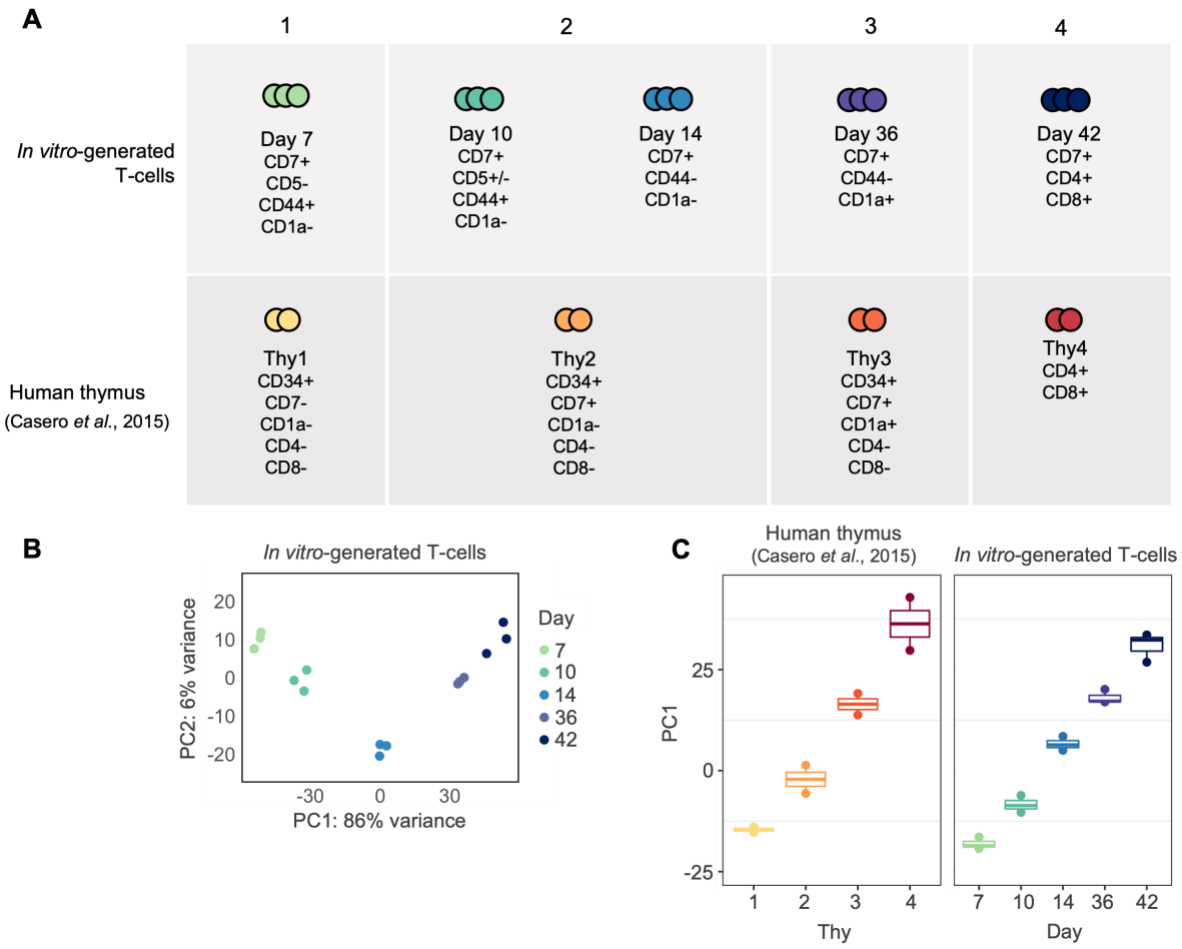


Figure 3.3 Overview of datasets used for the analysis of human T-cell differentiation.

A Sample comparison between *in vitro*-derived T-cell subsets (top, n = 3 replicates for each subset) and human intrathymic subsets (bottom, n = 2 replicates for each subset; data from Casero *et al*., 2015)[286]. The defining surface markers for each population are indicated.
B Principal component analysis of *in vitro*-derived T-cell subsets.
C Comparison of the *in vitro*-derived T-cell subsets and human intrathymic subsets when projected onto PC1.

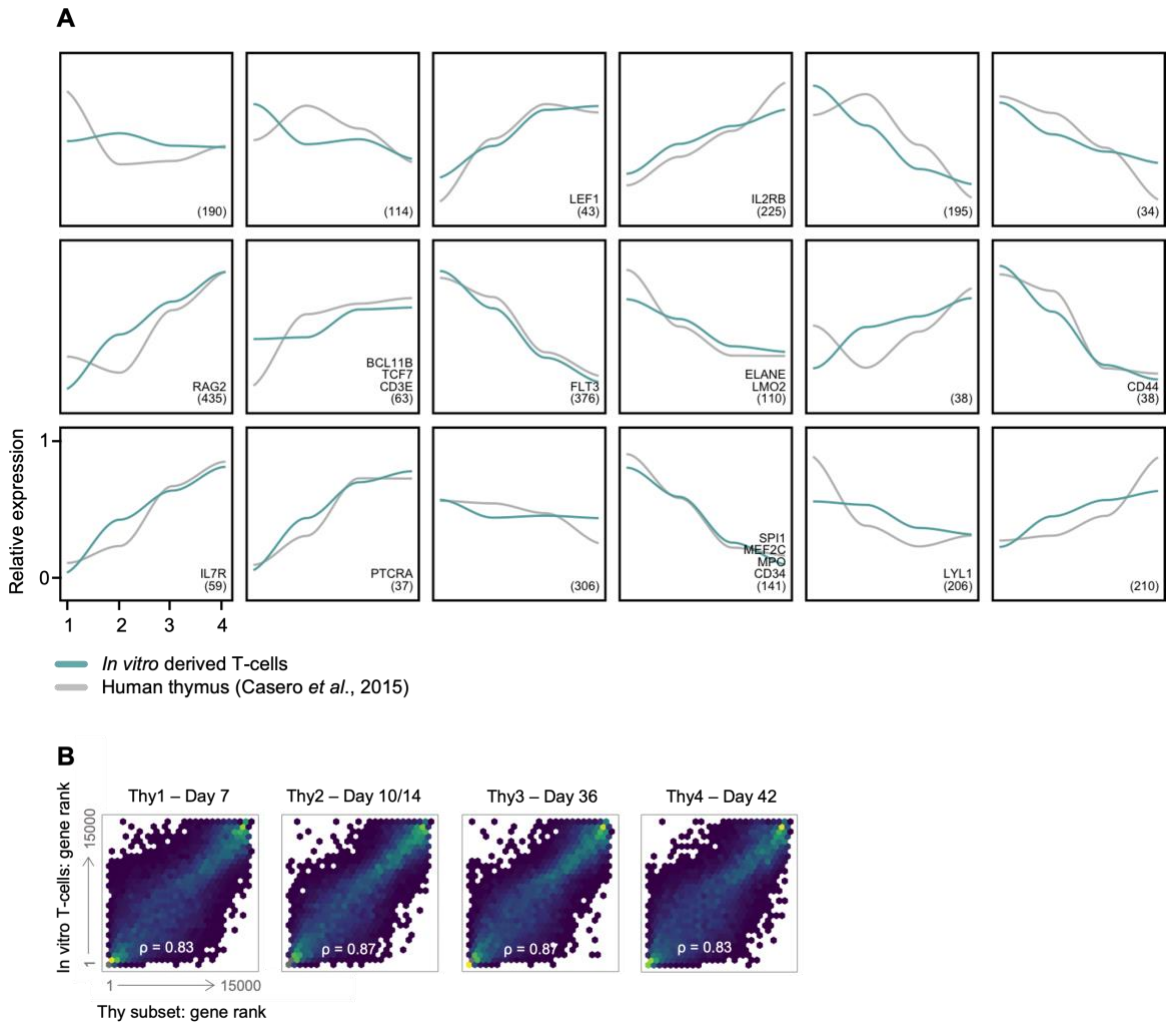


Figure 3.4 Transcriptomic comparison of *in vitro* generated T-cells and human intrathymic populations.

A Mean expression profiles of *in vitro*-derived T-cell subsets (teal) compared to human intrathymic subsets (grey). The number of genes in each gene expression cluster are indicated in brackets. Genes known to be involved in processes related to T-cell differentiation are highlighted. RNA-seq of human intrathymic populations was obtained from GEO accession GSE69239 (published by Casero *et al.*, 2015)[286].
B Spearman rank correlation of gene expression between matched *in vitro*-derived T-cell subsets and human intrathymic subsets.

3.3.3 *In vitro* generated T-cell subsets exhibit T-cell receptor rearrangement and myeloid-lineage restriction

Although T-cells generated *in vitro* appeared similar to intrathymic T-cell populations based on expression profile and surface marker expression, we considered whether these cells exhibit developmental checkpoints that are characteristic of T-cell development. Two processes during T-cell ontogeny which can be readily interrogated *in vitro* include T-cell commitment and TCR rearrangement.

T-cell commitment is a process whereby once-multipotent cells selectively restrict access to alternative fates until the T-lineage remains the only available option. To examine if CB-derived T-cells undergo aspects of T-cell commitment, we assessed the growth of various T-cell subpopulations by transferring them to media supportive of the growth of granulocytic and monocytic ("myeloid") lineages. We specifically chose to test myeloid potential at least in part due to the relatively late divergence of myeloid and T-lymphoid lineages during human T-cell development[107]. In contrast, the potential of B-cell or erythroid fates are thought to be lost quite early in human and murine T-cell development based on lineage potential assays in which T-cells are plated into conditions permissive of alternative lineages (reviewed in[104,324]).

To this end, cells were sorted into myeloid conditions on the basis of CD7, CD44 and CD1a expression (Figure 3.5A,B). These markers were chosen as previous studies have correlated T-cell commitment in human thymocytes with expression of CD7 and CD1a[43,101], and more recently, the loss of $CD44^{dim}$ expression, which precedes gain of CD1a[95,107]. Consistent with these prior studies, the *in vitro* generated $CD7^{+}$ $CD44^{+/-}CD1a^{+/-}$ T-cell subpopulations had nonequivalent myeloid potential as measured by growth and phenotypic changes upon culture in myeloid conditions (Figure 3.5C). Whereas $CD7^{+}CD44^{+}CD1a^{-}$ cells had virtually indistinguishable myeloid and T-lymphoid potential based on limiting dilution analysis, myeloid developmental potential was reduced upon loss of CD44 expression, and even further so upon gain of CD1a. When examining the immunophenotype of myeloid outputs generated from $CD7^{+}CD44^{-}$ cells, the surface marker expression was consistent with granulocytic or monocytic lineages based on the expression of markers including HLA-DR, CD15, CD13, CD11b, CD7 and CD5, and we observed a clear separation of T-cell and myeloid populations by Uniform Manifold Approximation and Projection (UMAP; Figure 3.6).

We next wondered if there was a temporal difference in myeloid potential within the $CD7^+CD44^-$ subset. Our initial experiments were performed when cells were cultured for a total of 14 days *in vitro*, as this timepoint includes the earliest emergence of $CD7^+CD44^-$ cells. As cells are cultured for longer, however, the $CD7^+CD44^-CD1a^+$ population becomes more appreciable (Figure 3.1A). In our next set of experiments, cells were instead cultured for a period of 28 days in order to assess myeloid lineage output among the CD44 and CD1a subpopulations (Figure 3.7A,B). Although the $CD44^+$ fraction was too rare to feasibly examine at this timepoint, the myeloid potential of $CD7^+CD44^-CD1a^-$ cells was greatly reduced when compared to the same phenotypic population cultured for only 14 days (20 – 30-fold decrease; Figure 3.7C). These results suggest that although CD44 loss may correlate with loss of myeloid potential, there is also a temporal factor associated with this lineage exclusion process. Importantly, the T-cell lineage output did not appear to be affected (1 in 4 – 8 cells compared to 1 in 7 cells after culture for 14 days and 28 days, respectively), which might suggest that the differences observed in myeloid output are not related to differences in clonogenic potential across the two timepoints tested. Overall, these findings suggest that myeloid lineage exclusion in *in vitro* derived T-cells occurs in a $CD7^+CD44^-$ population, which is similar to what has been described in primary human thymocytes[95,107], and this population is thus referred to as "T-lineage restricted" throughout.

We then performed a differential gene expression analysis of subsets before and after the T-lineage restriction timepoint (i.e., day 7 and 10 vs day 14, 36 and 42), yielding a set of ~4000 significantly differentially expressed genes (absolute $\log_2$ fold change $\geq$ 1, BH-adjusted p-value < 0.05). Pathway analysis of these differentially expressed genes revealed that programs associated with myeloid differentiation were among the top pathways upregulated in the non-restricted T-cell subsets (i.e., day 7 and day 10), whereas programs associated with T-cell differentiation were upregulated in the restricted T-cell subsets (i.e., day 14, 36 and 42) (Figure 3.8). These results support our functional experiments and further suggest that myeloid lineage exclusion may be important for T-cell differentiation.

We next sought to determine if T-cells generated *in vitro* also undergo TCR rearrangement. To this end, MiXCR analysis was performed to determine the relative proportion of rearrangements within the *TRG*, *TRD*, *TRB* and *TRA* loci in T-cell subsets collected throughout *in vitro* culture. MiXCR can be used to infer TCR rearrangement status from RNA-seq data by extracting reads that align to the CDR3 region[295]. I

compared the extent and proportion of rearrangement observed in *in vitro*-derived T-cells to human intrathymic T-cell subsets (Figure 3.9). Our data suggest that the first instance of reproducible detection of TCR rearrangement occurs by day 10 of *in vitro* culture, with most rearrangements occurring within the gamma chain. Although rearrangements were present even by day 7, these were supported by very few unique rearrangements, and high variability was observed across sample replicates, which may be suggestive of contaminating mature T-cell populations. Comparison of the percent of total rearrangements between our data and publicly available RNA-seq data of intrathymic human T-cell subsets similarly showed rearrangement of *TRG* and *TRD* loci in Thy3, followed by *TRB* and *TRA* rearrangement emerging in Thy4. This is further supported by additional studies which report that TCR rearrangement occurs in a sequential order from *TRG* > *TRD* > *TRB* > *TRA*[87–89]. Overall, this data suggests that TCR rearrangement occurs in the expected fashion in T-cell subsets derived from *in vitro* cultures.

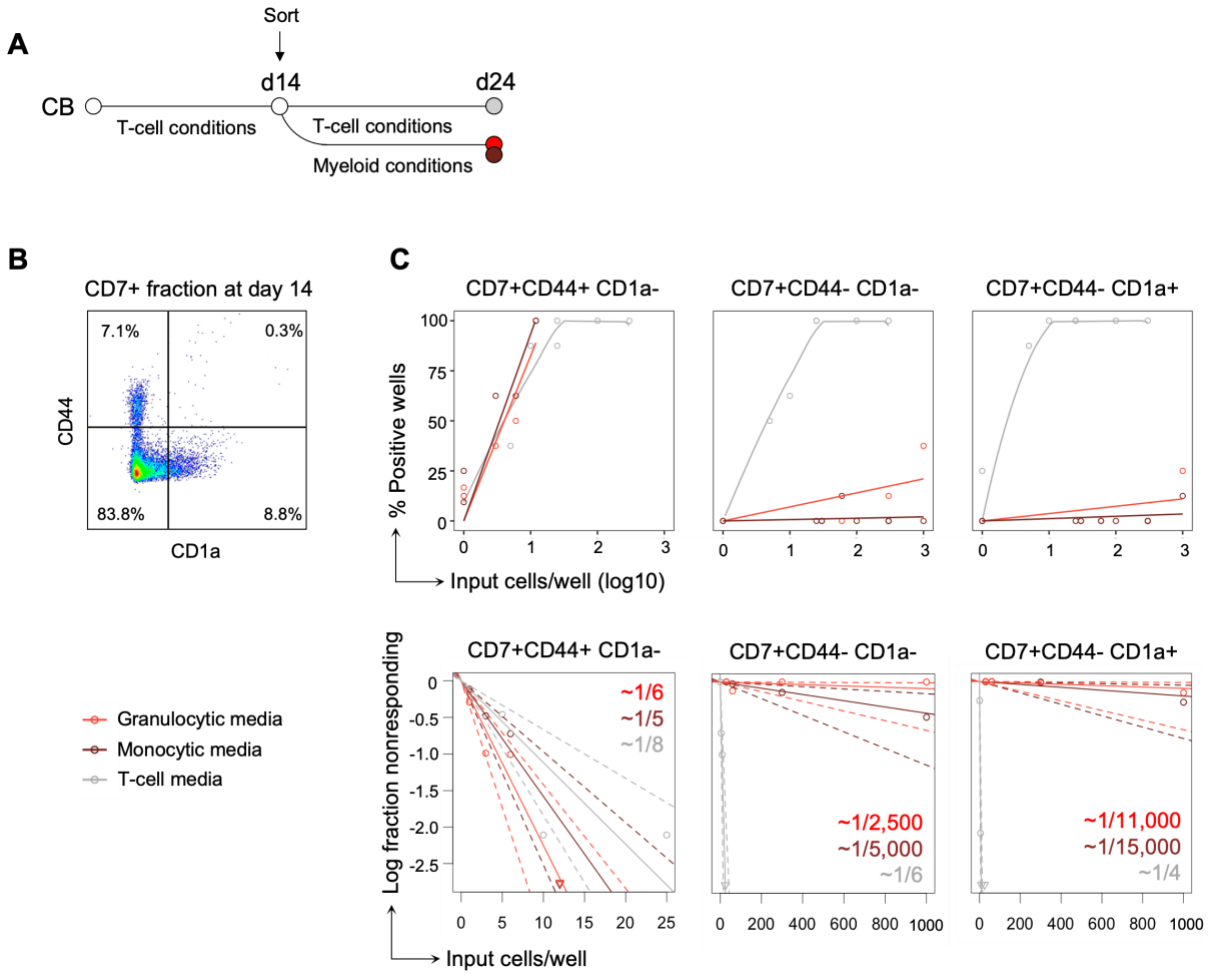


Figure 3.5 CD44 loss in T-cells correlates with myeloid-lineage exclusion.

A General experimental design of media transfer experiments. T-cells were derived *in vitro* from CD34+ cord blood (CB) and cultured for 14 days in T-cell growth conditions. CD7+ cells were sorted upon the basis of CD44 and CD1a expression into conditions supportive of T-cell (grey) or granulocytic/monocytic lineages (red and burgundy). Cells were scored for growth and harvested for flow cytometry at day 24.
B Representative flow cytometry plot indicating the levels of CD44 and CD1a expression within the CD7+ subset after 14 days of culture.
C The percent of positive wells for each phenotypic population (i.e., CD7+CD44+CD1a-, CD7+CD44-CD1a-, and CD7+CD44-CD1a+) as a function of the number of input cells per well (top). Limited dilution growth assays measuring the frequency of growth for each sorted T-cell population when transferred into T-cell or myeloid culture conditions (bottom). For all plots, each dot is coloured according to the media condition that cells were sorted into. A minimum of 8 replicates were performed for each condition and cell concentration.

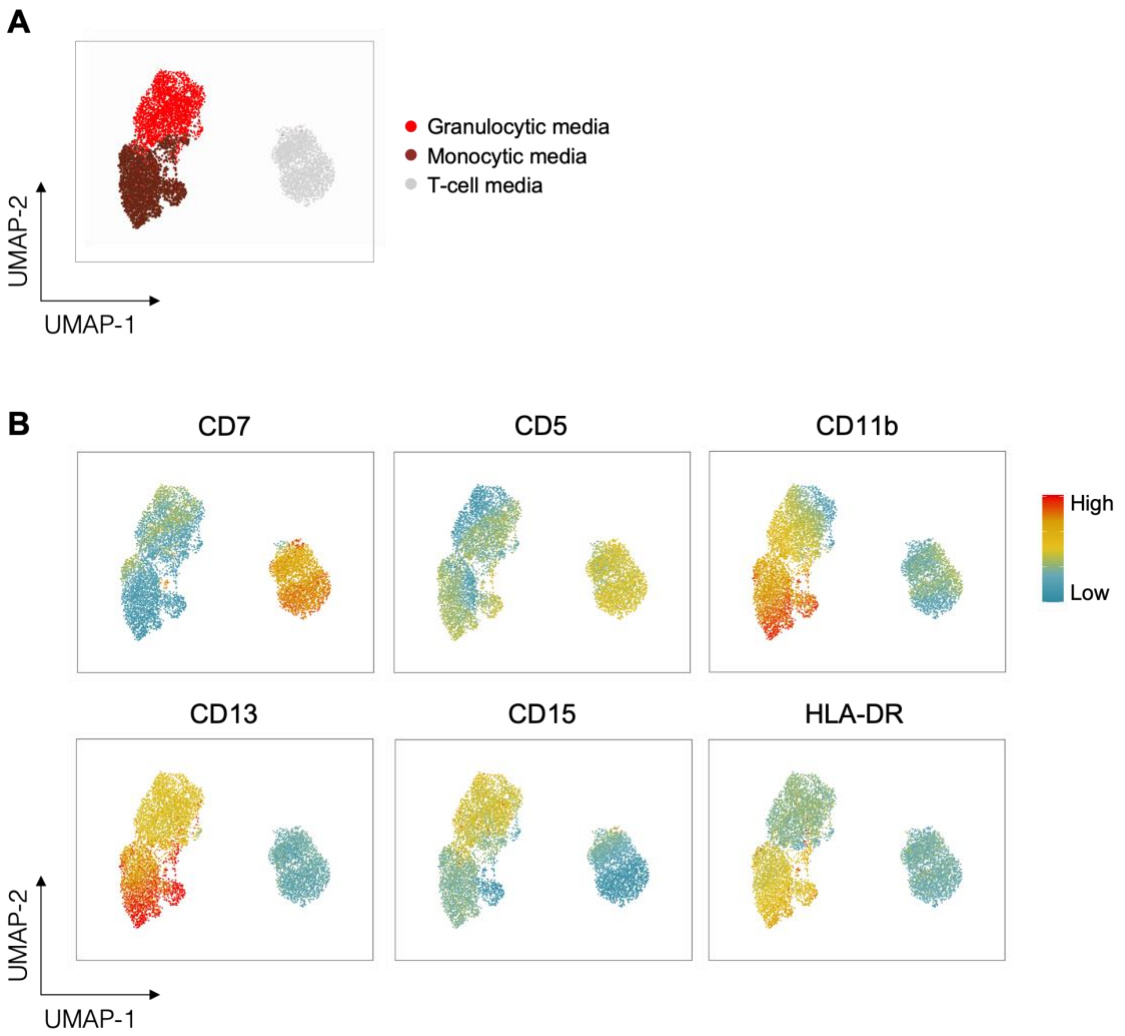


Figure 3.6 14-parameter UMAP distribution of flow cytometric immunophenotypes of T-lineage restricted cells after transfer into T-cell or myeloid growth conditions.

A Cells were cultured in T-cell conditions for 14 days and subsequently transferred into granulocytic, monocytic or T-cell conditions for an additional 10 days. The UMAP distribution indicates cells harvested after 10 days of growth in the indicated media conditions (red, burgundy, or grey).
B The expression of select surface markers which discriminate between the T-cell or myeloid lineage are shown and are coloured according to relative staining intensity.

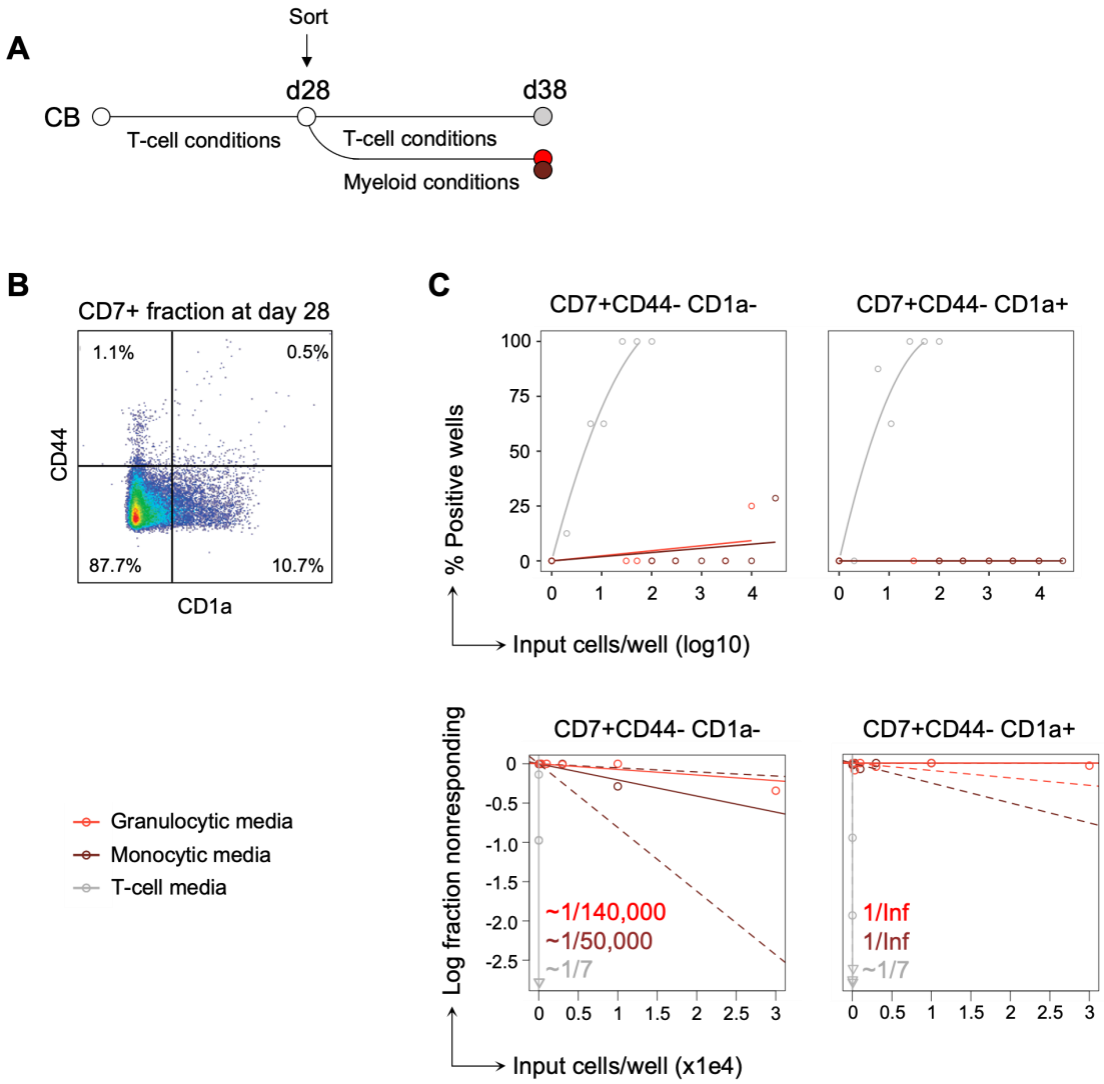


Figure 3.7 T-lineage restriction in *in vitro*-derived T-cell populations after 28 days of culture.

A General experimental design of media transfer experiments. T-cells were derived *in vitro* from CD34+ cord blood (CB) and cultured for 28 days in T-cell growth conditions. CD7+ cells were sorted upon the basis of CD44 and CD1a expression into conditions supportive of T-cell (grey) or granulocytic/monocytic lineages (red and burgundy). Cells were scored for growth and harvested for flow cytometry at day 38.
B Representative flow cytometry plot indicating the levels of CD44 and CD1a expression within the CD7+ subset after 28 days of culture.
C Top: The percent of positive wells for each phenotypic population as a function of the number of input cells per well. Bottom: Limited dilution growth assays measuring the frequency of growth in sorted T-cell populations when transferred into T-cell or myeloid culture conditions. For all plots, each dot is coloured according to the media condition that cells were sorted into. A minimum of 8 replicates was performed for each condition and cell concentration. Inf = infinity.

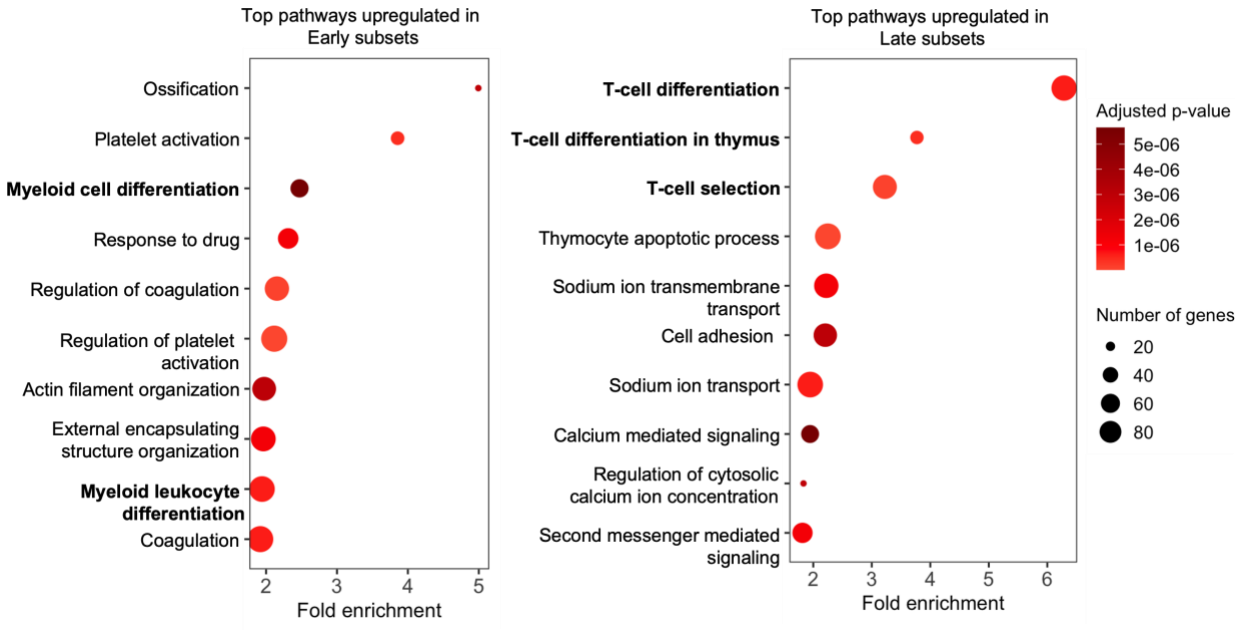


Figure 3.8 Pathways associated with myeloid differentiation are upregulated in early T-cell subsets.

Pathway analysis was performed using genes differentially expressed between Early (i.e., day 7 and 10) and Late (i.e., day 14, 36 and 42) T-cell subsets. The top 10 significantly enriched pathways are shown for genes upregulated in Early or in Late subsets. The size of each dot indicates the number of genes that overlap with each pathway.

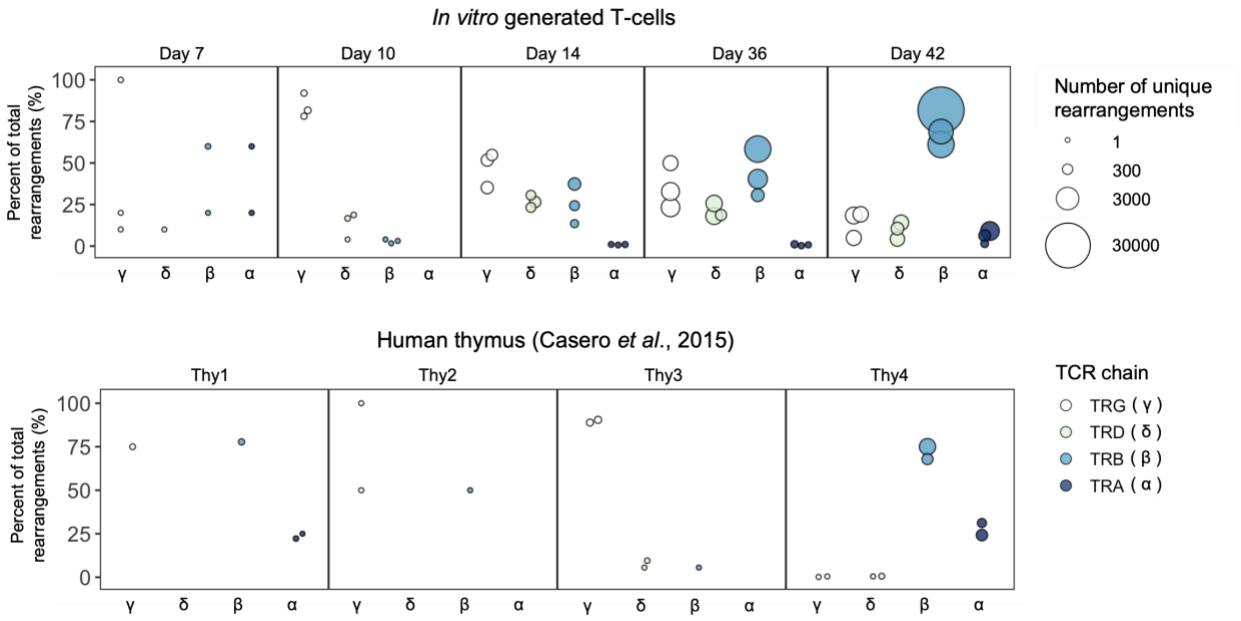


Figure 3.9 Comparison of T-cell receptor rearrangement analysis of *in vitro*-derived T-cell subsets and human intrathymic populations.

TCR rearrangements were inferred from RNA-seq data using MiXCR. Each dot indicates a sample replicate and is coloured according to the TCR chain. The size of each dot indicates the number of unique rearrangements as determined from the *CDR3* regions.

3.3.4 Interrogation of DNA methylation patterns and gene expression profiles of T-cell populations throughout differentiation

Due in part to the relative paucity of DNA methylation analyses in normal T-cell populations, we next performed whole genome bisulfite sequencing (WGBS) to gain a better understanding of the antecedent DNA methylation changes that occur as cells establish and embellish upon the T-cell identity. The dataset described here was collected in conjunction with the RNA-seq dataset described in Section 3.3.2, however the WGBS data includes only one replicate for each time point surveyed, and thus the following methylation analyses are strictly descriptive (Figure 3.10A). To complement this dataset, DN and DP subsets from human thymus were also collected for WGBS analysis as described in Materials and Methods (Section 3.2.3.2, Appendix C).

I first performed a genome-wide analysis of DNA methylation patterns across each of the *in vitro* T-cell subsets (day 7 – day 36) and two human intrathymic populations (DN and DP). When parsed into different

genomic features, DNA methylation levels showed expected patterns such as hypomethylation of gene promoters and CpG islands (CGIs), and comparatively higher levels of methylation within gene bodies (Figure 3.10B). Across the various T-cell subsets, global genome-wide methylation levels remained consistent, suggesting that T-cell differentiation does not coincide with major changes in methylation (Figure 3.10B). This is congruent with previously published reports in human thymic populations, which show that methylation of the CpGs does not vary greatly throughout development[196].

Although the majority of CpGs do not show drastic differences in methylation during development, principal component analysis of the top 8000 variable methylated CpGs revealed that samples segregate both by cell type (*in vitro*-generated T-cells vs human thymocytes) and differentiation stage, suggesting that methylation differences exist between and within cell types, albeit in a minority of CpGs (Figure 3.10C). Motivated by these results, I sought to determine if differences in DNA methylation correspond to changes in gene expression. As such, I focused on methylation within the promoter region. Using the RNA-seq data generated from each of the *in* vitro-derived T-cell subsets, I pared down analysis to the subset of genes that are differentially expressed between pre-committed T-cell subsets ("Early", day 7 and 10) and T-lineage restricted subsets ("Late", day 14, 36 and 42), as described previously (Figure 3.11A; differentially expressed genes were defined by an absolute $\log_2$ fold change $\geq$ 1, BH-adjusted p-value < 0.05). set of ~4000 differentially expressed genes could be loosely grouped based on expression increase vs decrease during development (Figure 3.11B, red vs blue bars). Inspection of promoter methylation of these two groups (increasing vs decreasing RNA expression during development) revealed that decreases in gene expression were concomitant with increases in promoter methylation, although the opposite trend was not observed among genes that increased in expression throughout development (Figure 3.11B). To determine which genes had coordinated expression and methylation patterns, RNA-seq and WGBS data were integrated (Figure 3.11C, highlighted points indicate average absolute methylation difference $\geq$ 10% and absolute $\log_2$ fold-change $\geq$ 1 between "Early" and "Late" subgroups from WGBS and RNA-seq, respectively). This approach revealed several genes important for myeloid/B-lineage development, such as *TBXA2R, CD179B, CLNK*, and *SPI1*, which showed decreased gene expression and increased promoter methylation among early T-cell subsets (Figure 3.11C, Q2 and Figure 3.11D, middle panel). Although genes which increased in expression throughout development were not consistently associated with decreased

promoter methylation, we identified several exceptions including the critical T-cell genes *RAG2, PTCRA,* and *CD1A* (Figure 3.11C, Q3 and Figure 3.11D, top panel). The association of gene expression and methylation in some of these genes, including *RAG2* and *PDCD1*, are consistent with previous studies[196]. Taken together, these results may suggest that promoter methylation can help coordinate the expression and repression of genes associated with T-cell maturation and alternative lineages, respectively, and may implicate a role of DNA methylation in development.

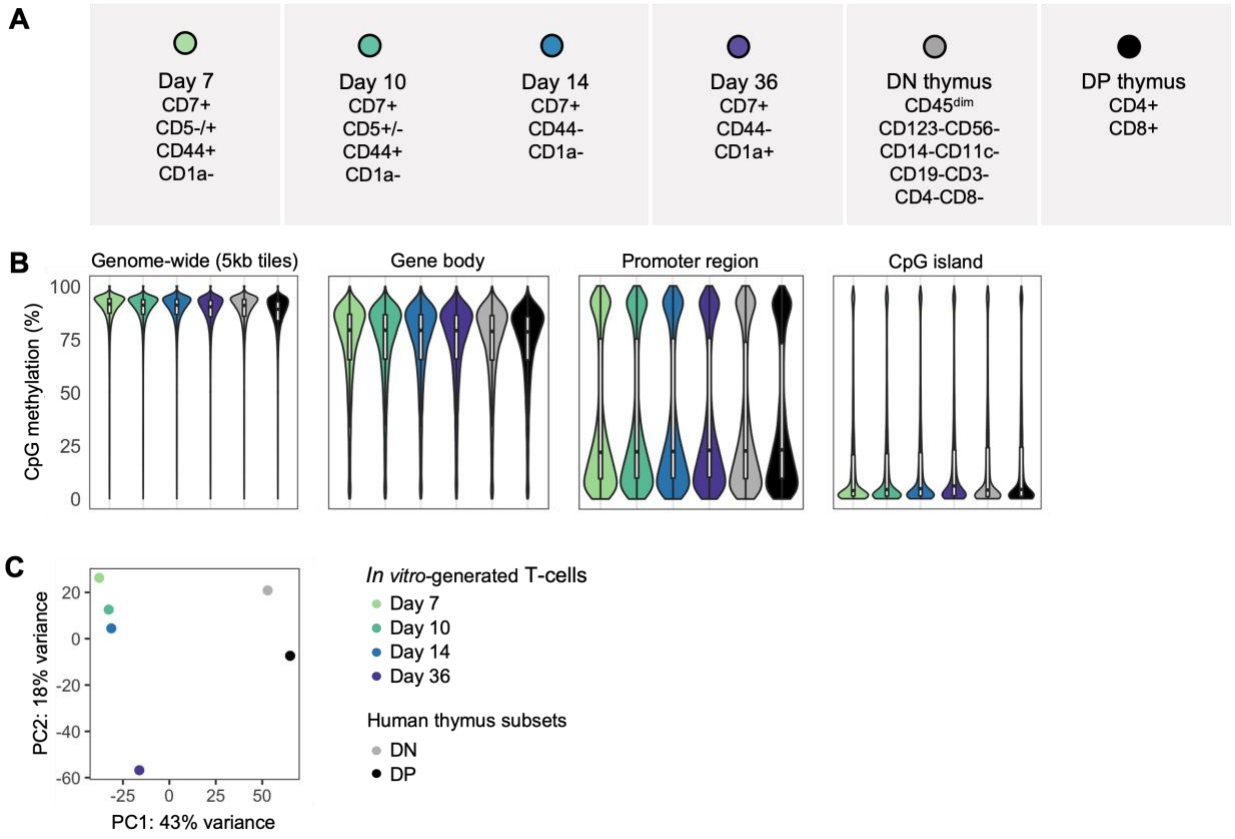


Figure 3.10 Whole genome bisulfite sequencing of T-cell subsets.

A Overview of samples collected for DNA methylation analysis by WGBS (n = 1 for each subset). DN = double negative, DP = double positive.
B CpG methylation analysis measured by WGBS for various genomic features in *in vitro*-derived T-cell subsets collected at various time points (shades of green, blue, and purple) or human thymic subsets (grey and black).
C Principal component analysis of the top 8000 variably methylated CpGs across all samples.

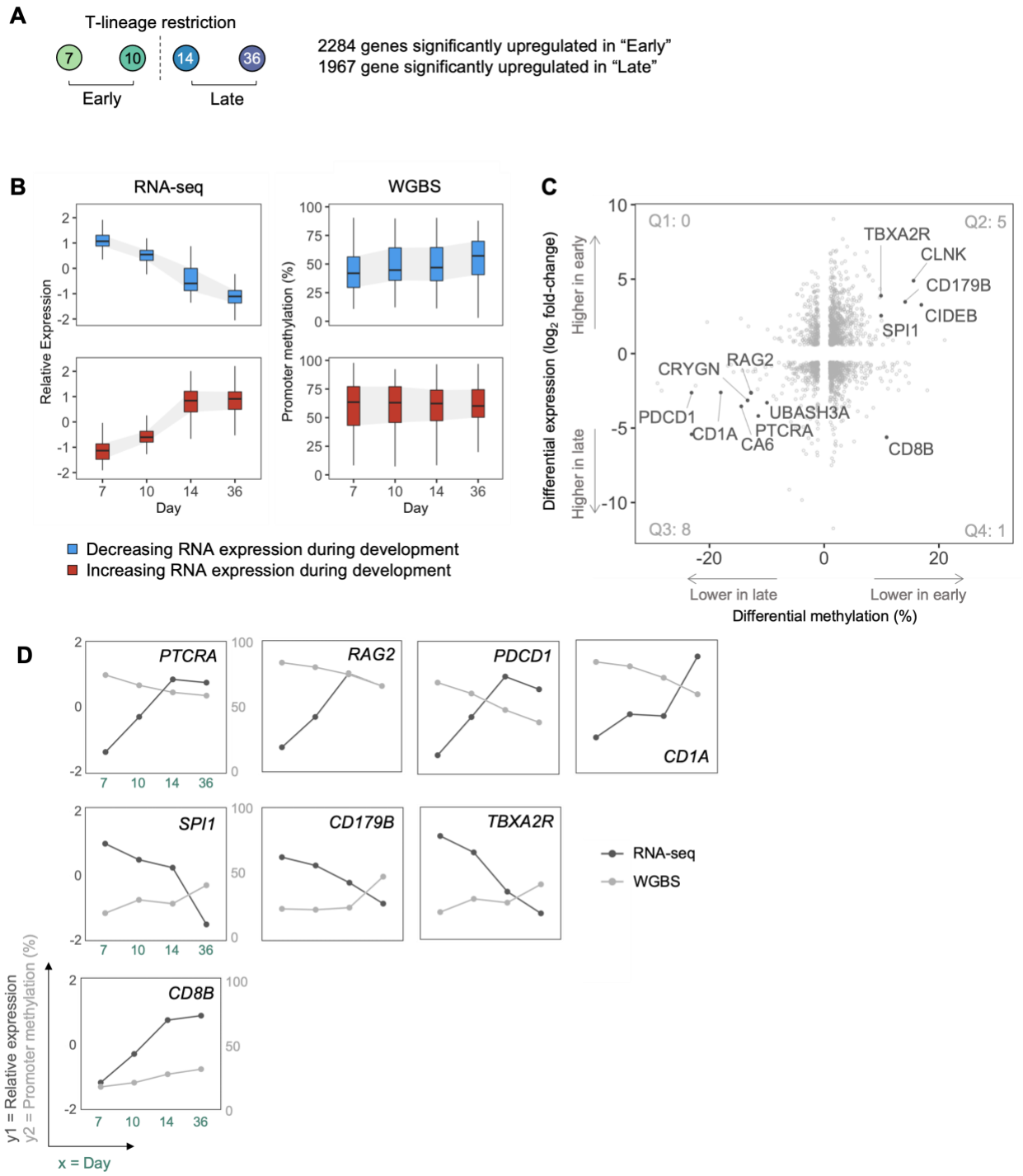


Figure 3.11 Comparison of CpG methylation by WGBS and gene expression by RNA-seq in the in vitro-derived T-cell subsets.

A Overview of sample comparisons. Subsets collected at day 7 and 10 (Early) or day 14 and 36 (Late) were grouped. Differential gene expression analysis was performed on the two groups (Early vs Late).

B Relative expression and promoter methylation of genes significantly upregulated in Early T-cell subsets (blue) or significantly upregulated in Late T-cell subsets (red).
C Comparison of CpG methylation (x-axis) and gene expression (y-axis). Only significantly differentially expressed genes (absolute log2 fold-change $\geq$ 1, adjusted p-value < 0.05) or promoter regions with an absolute CpG methylation difference $\geq$ 10% are shown. Each quadrant is defined as follows: Q1, genes associated with increased expression and higher CpG methylation in Early; Q2, genes associated with increased expression and lower CpG methylation in Early; Q3, genes associated with increased expression and lower CpG methylation in Late, Q4, genes associated with increased expression and higher CpG methylation in Late.
D Exemplary gene expression and CpG methylation patterns of select genes in each quadrant. Top: Q3, Increasing expression / decreasing promoter methylation; middle: Q2, Decreasing expression / increasing promoter methylation; bottom: Q4, Increasing expression / increasing promoter methylation. RNA-seq data is indicated by the dark grey lines and plotted on the y1-axis. WGBS data is indicated by the light grey lines and plotted on the y2-axis.

3.4 Discussion

Understanding normal T-cell biology can provide context for what may go awry during disease. Experimentally tractable systems that model human T-cell differentiation provide essential tools for studying context-specific effects of various genetic alterations. The benefits of using *in vitro*-derived cells to model elements of T-cell development are threefold: (1) These cultures yield large numbers of T-cells which can be harvested at any stage during differentiation, (2) T-cells are generated in a reproducible manner based on RNA-seq expression, and (3) cells obtained from this culture system are amenable to genetic manipulation. The first and third points are particularly salient when considering alternative T-cell sources, such as human thymocytes, in which immature T-cell subsets can be difficult to capture in appreciable number and are more difficult to manipulate *in vitro*. In this chapter, I employed the use of flow cytometric profiling and molecular profiling (i.e., by RNA-seq and whole-genome bisulfite sequencing) to characterize an established feeder-free system for modeling the development of T-cells. Unlike the “gold standard” OP9-DL1 culture system, the conditions described here are chemically defined and do not require the use of stromal feeder cells, which may increase inter-experimental reproducibility. When compared to human thymocyte populations, cells derived from the feeder-free system recapitulated aspects of T-cell development based on gene expression profiles and on functional evidence such as TCR rearrangement and restriction of the myeloid lineage.

Based on our RNA-seq analysis, the *in vitro*-derived T-cell subsets undergo TCR rearrangement in a predictable manner, marked by the emergence of *TRG* rearrangement in early T-cell subsets and

eventual *TRA* rearrangements arising in $CD4^+CD8^+$ double positive cells. The *in vitro*-derived T-cell subsets also showed similar gene expression trends compared to human intrathymic populations, displaying a predictable increase in expression of genes important for T-cell function among mature T-cell subsets, such as *RAG2*, and *PTCRA*, and a concomitant decrease in genes associated with stem/progenitor cells or alternative lineages, such as *LMO2, LYL1, MEF2C, FLT3, SPI1, MPO,* and *ELANE.* Altogether this suggested that, in these cultures, differentiation is associated with initiation of a T-cell program and downregulation of a progenitor program. When we performed pathway analysis of genes significantly upregulated in immature T-cell subsets (i.e., subsets collected after 7 and 10 days of *in vitro* culture), we noted enrichment of several pathways that pertain to myeloid cell differentiation. This may suggest that myeloid and T-cell developmental programs are at one point simultaneously available in immature T-cells, but myeloid programs become deactivated or repressed as cells later upregulate genes important for specialized T-cell function. This has been long understood by several groups who have focused on lineage potential in human and murine T-cell progenitors[43,97,101,107,325,326] (reviewed in[104]). Yet, the mechanisms underlying repression of the myeloid lineage in developing T-cells remain poorly understood and require further investigation.

DNA methylation is a major mechanism of gene repression, but few studies have examined the incipient DNA methylation changes in early human T-cell differentiation[196]. To interrogate these changes, we performed WGBS on the same T-cell subsets as those collected for RNA sequencing. Our exploratory analysis suggested that although the majority of CpGs do not change in DNA methylation across each subset, promoter methylation most prominently increases among a selection of genes whose expression is significantly downregulated during development. Among these genes include those associated with alternative lineages, such as *TBXA2R, CD179B, CLNK,* and *SPI1*. When focusing on genes that were significantly upregulated during development, an inverse correlation between gene expression and promoter methylation less clear, which could suggest that DNA demethylation may not play as large a role in gene activation. This is consistent with the general finding across several tissue types that DNA methylation is primarily – though not strictly – associated with gene repression, and only to a lesser extent with gene activation[147,131–138]. These results present a model in which DNA methylation may be among the repressive mechanisms which enforce T-lineage restriction through inhibition of non-T-lineage identities. In

line with this, previous studies have indicated that inhibition of DNA methyltransferases in various murine T-cell subsets through the use of 5-aza-2'-deoxycitadine result in a modest increase in myeloid, as opposed to T-lineage, output[192]. This outcome was most evident in immature T-cell stages (i.e., DN1 and DN2 subsets), but not in the mature DN3 subset. This could suggest that DNA methylation may play an important role in maintaining T-lineage fidelity.

To provide functional evidence that the molecular profiles that we observed correlated with a loss of stem/progenitor features, I harvested T-cells populations at different timepoints throughout differentiation on the basis of CD7, CD44 and CD1a expression. Each T-cell subset was then transferred into conditions supportive of myeloid-lineage growth. The results from these experiments revealed a reduction in monocytic and granulocytic output which became most pronounced upon loss of $CD44^{dim}$ expression, which in human thymocytes has been previously identified as a proximal marker of T-lineage commitment[95,107]. The results from this experiment suggest that myeloid lineage potential is reduced upon $CD44^{dim}$ loss. Whether this also correlates with T-cell commitment in these cultures remains to be determined; T-cell commitment implies the exclusion of all alternative fates, including B-cell, erythroid, NK, and DC lineages, but our experiments assess only myeloid lineage potential. At least in murine and human T-cell development, divergence of B-lymphoid, dendritic, and erythroid lineages is thought to occur prior to the divergence of myeloid and NK fates[107] (reviewed in[104,324]), which by extension might imply that B, erythroid, and DC fates have already been excluded in these cultures by the time of $CD44^{dim}$ loss. Nonetheless, a more comprehensive analysis of alternative (non-T) lineage potential will be required in order to directly assess T-cell commitment and restriction.

Overall, the findings discussed here provide a blueprint of normal T-cell development modelled *in vitro* and serve as a framework within which to dissect the effect of genetic perturbations on T-cell differentiation and lineage specification. From the results of our expression profiling and DNA methylation analyses, we observed that increased methylation throughout development correlated with decreased expression of genes operative in alternative lineage programs, indicating that DNA methylation might be an important regulator of the T-cell lineage. This is congruent with previous studies which suggest that the level of DNA methylation could play a role in regulating T-lineage vs myeloid lineage development[191,192]. The main family of enzymes which are involved in establishing and maintaining DNA methylation are the

DNA methyltransferases (DNMTs), and include DNMT3A, DNMT3B, and DNMT1. Accordingly, the focus of the next chapter is to examine the effect of genetic perturbation of *DNTM3A*, one of two *de novo* DNA methyltransferases, on T-cell development and lineage regulation.

Chapter 4: Characterization of the functional consequences of *DNMT3A* loss in normal development and in T-ALL

4.1 Introduction

The factors that underlie T-lineage restriction and differentiation are highly complex, and hinge on the interplay of epigenetic regulators and transcription factors to orchestrate the sequential elimination of alternative fates (reviewed in[327]). In normal development, the role of these elements can be highly context-specific and may act to coordinate repression of factors associated with stem/progenitor cells and activation of lineage-specific programs. Failure to properly shut off developmental programs operative in immature T-cells can manifest in diseases such as T-ALL, where alterations in epigenetic patterning or aberrant transcription factor expression may co-operate to circumvent otherwise highly regulated cellular processes such as proliferation, apoptosis, and differentiation.

DNA methylation is an epigenetic modification that is often associated with gene repression[147,132–138]. Furthermore, mutation of the *de novo* DNA methyltransferase, *DNMT3A*, is frequently observed in hematological malignancies, including T-ALL. The role of DNMT3A in T-cell development and progression to disease, however, is not wholly understood. Systematic profiling of disease-associated *DNMT3A* mutations suggest that the majority lead to protein loss-of-function[267]. Among T-ALL cases, *DNMT3A* mutations are associated with older age[257] and are enriched in ETP-ALL[243–245], a disease subtype commonly characterized by unique co-expression of lymphoid and myeloid/progenitor markers, and which has been previously associated with poor prognosis[226]. Although the frequency of purported loss-of-function mutations would suggest that *DNMT3A* plays tumor suppressor roles in pathogenesis, the functional implications of *DNMT3A* loss in T-ALL remains mostly unexplored.

At least in the context of constitutively active *Notch1*, conditional *Dnmt3a* knockout in murine HSCs has been previously associated with accelerated T-ALL development in mice and hypomethylation of enhancers relative to *Dnmt3a*-wildtype disease[195,268]. Although early T-cell arrest is only rarely observed upon *Dnmt3a* knockout, leukemic blasts display upregulated gene expression signatures associated with the myeloid lineage program, including *Elane, Mpo, Csf2rb and Gfi1b*[195,269]. This set of observations might suggest that loss of *DNMT3A* is associated with reactivation of the myeloid program. Indeed, a connection

between T-cell leukemia, myeloid leukemia, and *Dnmt3a* loss has been previously reported; in these studies, the degree of *Dnmt3a* loss (i.e., heterozygous vs complete knockout) in murine HSCs influenced the progression to myeloid vs T-cell leukemia in the context of FLT3-ITD mutation, where heterozygous *Dnmt3a* loss was associated with a higher frequency of myeloid leukemia, and complete *Dnmt3a* loss resulted in both T-cell and myeloid leukemia[269]. These results indicate that modulating the level of *Dnmt3a* expression can guide disease outcome (i.e., AML vs T-ALL) in the context of murine disease models.

In this chapter, I embellish and extrapolate upon these previous studies to further understand the role of DNMT3A in the context of normal human T-cell development. The aforementioned observations might posit a role of DNMT3A in influencing T-lymphoid and myeloid lineage decisions. Nonetheless, direct characterization of DNMT3A as a regulator of lineage choice in human T-cells remains to be seen. To this end, I investigate the role of *DNMT3A* in T-lineage restriction and exclusion of myeloid potential through knockdown and knockout experiments in *in vitro*-derived T-cells. I also explore the role of *DNMT3A* in the context of established leukemias through knockdown experiments in T-ALL cell lines. The results of these experiments provide a new perspective of the functional consequence of *DNMT3A* loss in T-cell development and may lend insight into novel pathogenic mechanisms in *DNMT3A*-mutated T-ALL.

4.2 Materials and methods

4.2.1 Cell culture and conditions

4.2.1.1 Primary cells

Human umbilical cord blood samples were obtained from Stem Cell Assay (Terry Fox Laboratory, BC Cancer Research Centre, Vancouver, British Columbia, Canada) with informed consent from donors. These samples contained pools of cord blood samples from multiple donors (2-10 individual donors). A list of cord blood pools accessed, and the experiments performed for each cord blood pool, can be found in Appendix A.

4.2.1.2 Cell lines

HEK293T cells were obtained from John Aster (Brigham & Women's Hospital, Boston) and cultured in DMEM (Thermo Fisher) supplemented with 10% heat-inactivated FBS (Sigma Aldrich) incubated at 37°C

and 5% CO_2. Cells were passaged at 80-90% confluency. The RPMI-8402 human T-ALL cell line was cultured in RPMI 1640 Media (Thermo Fisher) supplemented with Penicillin/Streptomycin (Stem Cell Technologies), 1mM sodium pyruvate (Thermo Fisher), and 10% heat-inactivated FBS (Sigma Aldrich). All cell lines were authenticated by STR genotyping and confirmed mycoplasma negative.

4.2.1.3 In vitro development of T-cells

Cells were cultured *in vitro* using the StemSpan T cell Generation Kit (STEMCELL Technologies) incubated at 37°C and 5% CO_2. Briefly, non-treated culture dishes were coated with StemSpan Lymphoid Differentiation Coating Material (100X), diluted 1:100 in PBS (Manufacturer) and left overnight at 4˚C, or at room temperature for 2 hours, before the coating material was removed and plates were rinsed with PBS. Cord blood samples were resuspended in StemSpan Lymphoid Progenitor Expansion Supplement (10X), diluted 1:10 in StemSpan Serum-Free Expansion Medium II (SFEM II, STEMCELL Technologies), or in StemSpan T cell Progenitor Maturation Supplement (10X), diluted 1:10 in SFEM II (STEMCELL Technologies), according to manufacturer's protocol. For the media transfer experiments, T-cells were sorted into non-treated culture dishes containing StemSpan Myeloid Expansion Supplement (100X) or StemSpan Myeloid Expansion Supplement II (100X), diluted 1:100 in StemSpan SFEM II (STEMCELL Technologies).

4.2.2 Cell proliferation assays

24h prior to cytokine stimulation, RPMI 8402 cells were cultured in reduced serum conditions (0.5% FBS). Proliferation was assessed by sorting GFP+ cells into 96 well plates containing RPMI 1640 Media supplemented with antibiotics and 100ng/μL rhGM-CSF, 100ng/μL rhG-CSF, or 100ng/μL rhM-CSF (Stem Cell Technologies). After 10 days of culture, GFP+ cell counts were determined by the addition of Accuchek Counting Beads (Thermo Fisher) at sample acquisition.

4.2.3 Media transfer experiments

Cells were sorted at various cell concentrations according to the phenotypes indicated into non-treated 96-well round bottom plates (Falcon). For all experiments, 3-8 replicates were performed for each

shRNA/sgRNA construct, media condition, and cell concentration. For populations cultured in conditions supportive of granulocytic cell expansion, cells were sorted into wells containing StemSpan Myeloid Expansion Supplement (100X), diluted 1:100 in StemSpan SFEM II (STEMCELL Technologies). For populations cultured in conditions promoting monocyte cell expansion, cells were sorted into wells containing StemSpan Myeloid Expansion Supplement II (100X), diluted 1:100 in StemSpan SFEM II (STEMCELL Technologies). For populations cultured in conditions supportive of T-cell differentiation, cells were sorted into wells coated with StemSpan Lymphoid Differentiation Coating Material (100X), diluted 1:100 in PBS, containing StemSpan Lymphoid Progenitor Expansion Supplement (10X), diluted 1:10 in StemSpan SFEM II (STEMCELL Technologies). Data analysis was performed using R (version 4.1.1) using the elda package (version 1.4.36) using default parameters except for "observed", which was set to "TRUE".

4.2.4 Lentiviral vectors and transduction

Lentiviral constructs

The TRC2-pLKO.5-puro vectors containing shRNA targeting *DNMT3A* (shDNMT3A-54, shDNMT3A-55, shDNMT3A-56, and shDNMT3A-58; Appendix E) were obtained from The RNAi Consortium (Broad Institute) and were distributed to us by the Center for High-Throughput Biology (University of British Columbia, Vancouver, Canada). The non-silencing control shRNA was gifted from David Sabatini (Addgene #1864)[328]. The PGK-Puro^R cassette was removed from all constructs and replaced with a PGK-GFP cassette, and the cloned constructs were verified by Sanger sequencing.

Lentiviral transduction

Cells were transduced with high-titre replication-incompetent lentivirus generated by transient co-transfection of HEK 293T cells using polyethyleneimine HCl MAX (Polysciences) with second-generation packaging and envelope vectors (pCMVΔR8.74, Addgene #22036; pCMV-VSV-G, Addgene #8454; and pRSV-Rev, Addgene #12253) as previously described[329], along with TRC2-pLKO.5-puro expression vectors containing shRNA targeting a scrambled non-silencing control or targeting *DNMT3A.* Supernatants from transfected HEK 293T cells were collected 48h post-transfection and concentrated by ultracentrifugation at 25,000 rpm for 90 minutes at 4°C in a Beckman SW32Ti rotor. *In vitro*-derived T-cells

and T-ALL cell lines were transduced by adding lentiviral supernatant and 4μg/mL polybrene and by centrifuging cells at 300 x *g* at 32°C for 2 hours.

4.2.5 CRISPR knockout experiments

Guide RNAs targeting the catalytic loop of *DNMT3A* (15R, 54R, 59R; Appendix E) or the *EMX* locus[330] were designed using CRISPOR[331] and were obtained from Integrated DNA Technologies (IDT). Recombinant *S. pyogenes* Cas9 nuclease (Alt-R® S.p. HiFi cas9 Nuclease V3) was obtained from IDT. Assembled Cas9 RNP complexes were electroporated into cells using the Neon Transfection System (Thermo Fisher Scientific), at 1600V with 3 x 10ms pulses. The frequency of insertions and deletions (indels) was determined by PCR amplification of the region targeted by the *DNMT3A* guides followed by Sanger sequencing as described below.

4.2.6 Analysis of CRISPR knockout cells

Genomic DNA was extracted from cell pellets by lysing in 1M Tris-HCL, 0.5M EDTA, 10% SDS, 5M NaCl, and 1mg/mL Proteinase K. Cells were lysed overnight at 60°C. DNA was precipitated with the addition of ice-cold EtOH/NaCl followed by incubation for 30 minutes. Samples were then centrifuged at 5,000 x *g* for 20 minutes. Pellets were washed twice with cold 70% EtOH and air dried. PCR amplification of the DNMT3A locus for Sanger sequencing was performed using Q5 High-Fidelity 2X Master Mix (NEB) and PCR primers (final concentration = 0.5μM) as listed in Appendix E and according to manufacturer's protocol. To determine the indel frequency, Sanger sequencing traces were deconvoluted using Inference of CRISPR Edits (ICE)[332].

4.2.7 Flow cytometry and cell sorting

4.2.7.1 Intracellular Phospho-flow cytometry

RPMI-8402 cells were cultured in RPMI 1640 media (Thermo Fisher) as indicated above. Prior to stimulation with cytokines, samples were starved for 24h in culture media without FBS. Samples were stimulated with 100ng/μL rhGM-CSF, 100ng/μL rhG-CSF, or 100ng/μL rhM-CSF (Stem Cell Technologies)

in RPMI 1640 media for 15 minutes in a 37°C water bath. Cells were stained with LIVE/DEAD Fixable Yellow Dead Cell Stain (Thermo Fisher) for 15 minutes and subsequently fixed with 1.5% paraformaldehyde (PFA) for 15 minutes at room temperature, followed by permeabilization with pre-chilled 100% methanol for 10 minutes at -20°C. In separate panels, cells were stained with AlexaFluor647-conjugated anti-Phospho-Erk1/2 (Thr202/Tyr204) antibodies (clone 19G2, Cell Signaling Technologies), AlexaFluor647-conjugated anti-Phospho-Stat3 (pY05) antibodies (clone 4/P-STAT3, BD Biosciences), AlexaFluor647-conjugated anti-Phospho-Stat5 (pY694) antibodies (clone 47/Stat5, BD Biosciences), in addition to BV421-conjugated anti-Human CD115, BV421-conjugated anti-Human CD116, and BV421-conjugated anti-Human CD114 (BD Biosciences). Equivalent amounts of isotype control antibodies were used. Samples were acquired on a LSRFortessa cytometer (Becton Dickinson). Data was analyzed using FlowJo software (Tree Star) and in R (version 4.1.1). Data was visualized using ggplot2 (version 3.3.6).

4.2.7.2 Flow cytometry

Samples were stained with fluorochrome-conjugated antibodies as listed in Appendix B. Data was acquired on the BD LSRFortessa Cell Analyzer (BD Biosciences) and the FACSymphony (BD Biosciences). To obtain cell numbers for each sample, AccuCheck counting beads (Invitrogen/ThermoFisher Scientific) were added prior to sample acquisition. In experiments where samples were acquired over several weeks, per-channel normalization was performed by staining BD CompBead Compensation Particles (BD Biosciences) with the antibody mastermix.

4.2.7.3 Cell sorting

Samples were stained with fluorochrome-conjugated antibodies as listed in Appendix B. Cell sorting was performed on ice using the BD FACSAria II (BD Biosciences) and the BD FACSAria Fusion (BD Biosciences).

4.2.7.4 Flow cytometry data analysis

Data was analyzed using FlowJo v10 software (Tree Star) and in R (version 4.1.1) using custom scripts. Single-cell data collected from 16 dimensions (FSC-A, SSC-A, CD4, CD8, CD7, CD5, CD1a, CD3,

CD13, CD33, HLADR, CD15, CD11b, CD44, CD34, and CD38) were reduced to 2 dimensions using the R package 'Monocle3' (version 1.0.0)[318] or 'umap' (version 0.2.28). Scale data was exported from FlowJo and was subsequently transformed using the arcsinh function from the 'flowCore' package (version 2.6.0)[319] in R (version 4.1.1) using the cofactor set to 150. For experiments where data was acquired on separate days, per-channel normalization was performed prior to archsinh transformation. Sample normalization was performed by collecting a bead standard stained with the antibody master mix at each timepoint. For generation of UMAP projections, data were generated using the predict function in 'umap' (version 0.2.28).

4.2.8 Whole-cell lysate extraction

Cell pellets from RPMI 8402 cells were resuspended in RIPA buffer supplemented with activated Na_3VO_4, PMSF Protease inhibitor, and Protease Inhibitor Cocktail (Calbiochem). Pellets were solubilized for 10 minutes on ice and centrifuged at 15,000 x *g* for 10 minutes at 4°C. Cleared lysates were quantified using the Pierce BCA Protein Assay Kit (Thermo Fisher). Lysates were heated at 70°C for 10 minutes prior to separation by gel electrophoresis.

4.2.9 Western blotting

Whole cell protein lysates were separated by gel electrophoresis on SDS-PAGE gels (5%/8% acrylamide stacking/separating layers) at 80V for ~30 minutes followed by 100 V for ~1 hour. Proteins were transferred to a methanol-activated PVDF membrane for 90 minutes at 70V. Membranes were blocked with 5%(w/v) skim milk in TBS with 0.1% Tween 20 for 1 hour at room temperature. Membranes were probed with anti-DNMT3A antibodies (1:1000 dilution, clone D23G1, Cell Signaling Technologies), and anti-β-actin antibodies (1:10000 dilution, AC-74, Sigma Aldrich) followed by incubation with secondary antibodies for one hour at room temperature. Chemiluminescent detection of the HRP-conjugated secondary antibodies was performed using ECL Western Blotting Substrate (Thermo Fisher).

4.3 Results

4.3.1 *DNMT3A* loss enhances the outgrowth of myeloid-like cells

We hypothesized that *DNMT3A* knockdown might delay or exclude T-lineage restriction in developing T-cells. This was motivated by three main observations: first, that inhibition of DNA methylation in murine progenitors results in myeloid skewing at the expense of lymphoid output[192]; second, that lower global DNA methylation levels are associated with myeloid commitment as opposed to T-lineage commitment[191,192]; and third, that *DNMT3A* is frequently mutated in a subtype of T-ALL with myeloid characteristics[245]. Our previous experiments discussed in Chapter 3 suggested that, at least in a genetically unperturbed state, *in vitro*-derived $CD7^+CD44^-$ T-cells cultured for 14 or 28 days exhibit a reduced frequency of cells capable of initiating growth in conditions supportive of granulocytic or monocytic lineages (referred to throughout as “myeloid”) but not in conditions supportive of the T-lineage (Figure 3.5, Figure 3.7). In contrast, no difference in initiation frequency was observed in $CD7^+CD44^+$ T-cells obtained from the same cultures when cultured in myeloid conditions (Figure 3.5). These results, consistent with previously reported findings[95,100], suggest that CD44 loss also correlates with lineage restriction in *in vitro* developed T-cells.

To assess if *DNMT3A* loss might alter this restriction point, I performed short hairpin RNA (shRNA)-mediated knockdown of *DNMT3A* in *in vitro* differentiating T-cells by lentiviral transduction of 4 different shRNAs tagged with green fluorescent protein (GFP), compared to a scrambled control shRNA. shRNAs targeting *DNMT3A* showed a reduction of DNMT3A protein by western blot analysis in a human T-ALL cell line (Figure 4.1A). Transduced T-cells were cultured for a total period of 28 days in culture, at which point $CD7^+CD44^-GFP^+$ cells were harvested and sorted into media supportive of either T-cell, granulocytic, or monocytic lineages (Figure 4.1B,C). Growth was assessed after 10 days of culture and visually inspected under a microscope, and wells were classified as positive or negative upon the basis of cell growth. To verify that classification by visual inspection was accurate, in one experiment, all wells classified as positive or negative were harvested and no viable events were detected by flow cytometry in wells classified as negative by visual inspection (data not shown).

Consistent across 4 independent trials, I observed that knockdown of *DNMT3A* conferred the ability of $CD7^+CD44^-GFP^+$ cells to grow in media supportive of myeloid lineages (Figure 4.1D), with the proportion of positive wells increasing as a function of the number of input cells. No positive wells were observed in

conditions transduced with scrambled shRNA, overall suggesting that *DNTM3A* knockout correlated with a greater proportion of "T-lineage restricted" $CD7^+CD44^-$ cells capable of proliferating under myeloid growth conditions.

We next considered if growth of $CD7^+CD44^-GFP^+$ cells in myeloid conditions also corresponded to a change in surface marker expression. The immunophenotype of positive wells was assessed using a 14-colour panel containing T- or myeloid-lineage markers, including CD4, CD8, CD7, CD5, CD1a, CD13, CD33, HLADR, CD15, CD11b, CD44, CD34, and CD38. UMAP analysis revealed two distinct groups that clustered according to the culture conditions (i.e., monocytic, granulocytic, or T cell media; Figure 4.2). When immunophenotypes were parsed in each cluster, we observed that cells cultured in T-cell conditions for the entirety of the experiment maintained expression of T-cell markers such as CD7 and CD5, whereas *DNMT3A* knockdown cells cultured in myeloid conditions expressed myeloid markers such as CD13 and CD33, and had relatively lower expression of CD7 and CD5, thus appearing "myeloid-like" in immunophenotype. Further segregation of the myeloid-like cell cluster could be achieved on the basis of CD15 expression, which resolved the monocytic and granulocytic lineages.

We wondered how the myeloid-like T-cells might compare to *bona fide* myeloid cells. To this end, we derived monocytic and granulocytic populations *in vitro* from $CD34^+$ cord blood and visualized them together with the myeloid-like T-cell populations (Figure 4.3A). Based on UMAP visualization, we observed that *DNMT3A* knockdown cells cultured in myeloid conditions clustered together with *bona fide* myeloid cells (Figure 4.3B). The myeloid-like populations expressed high levels of markers associated with the myeloid lineage, such as CD13 and CD33 (Figure 4.3C). Nonetheless, these populations still expressed higher levels of T-cell markers relative to *bona fide* myeloid populations, including CD4, CD3, CD5, CD7, and CD1a. Hierarchical clustering placed the myeloid-like cells at a state intermediate between *bona fide* T-cell and myeloid populations, possibly suggesting that the myeloid-like populations retain vestiges of a former T-cell phenotype (Figure 4.3D). Altogether, these results suggest that *DNMT3A* knockdown in "T-lineage restricted" $CD7^+CD44^-$ cells may increase the proportion of cells capable of proliferating and differentiating in myeloid conditions.

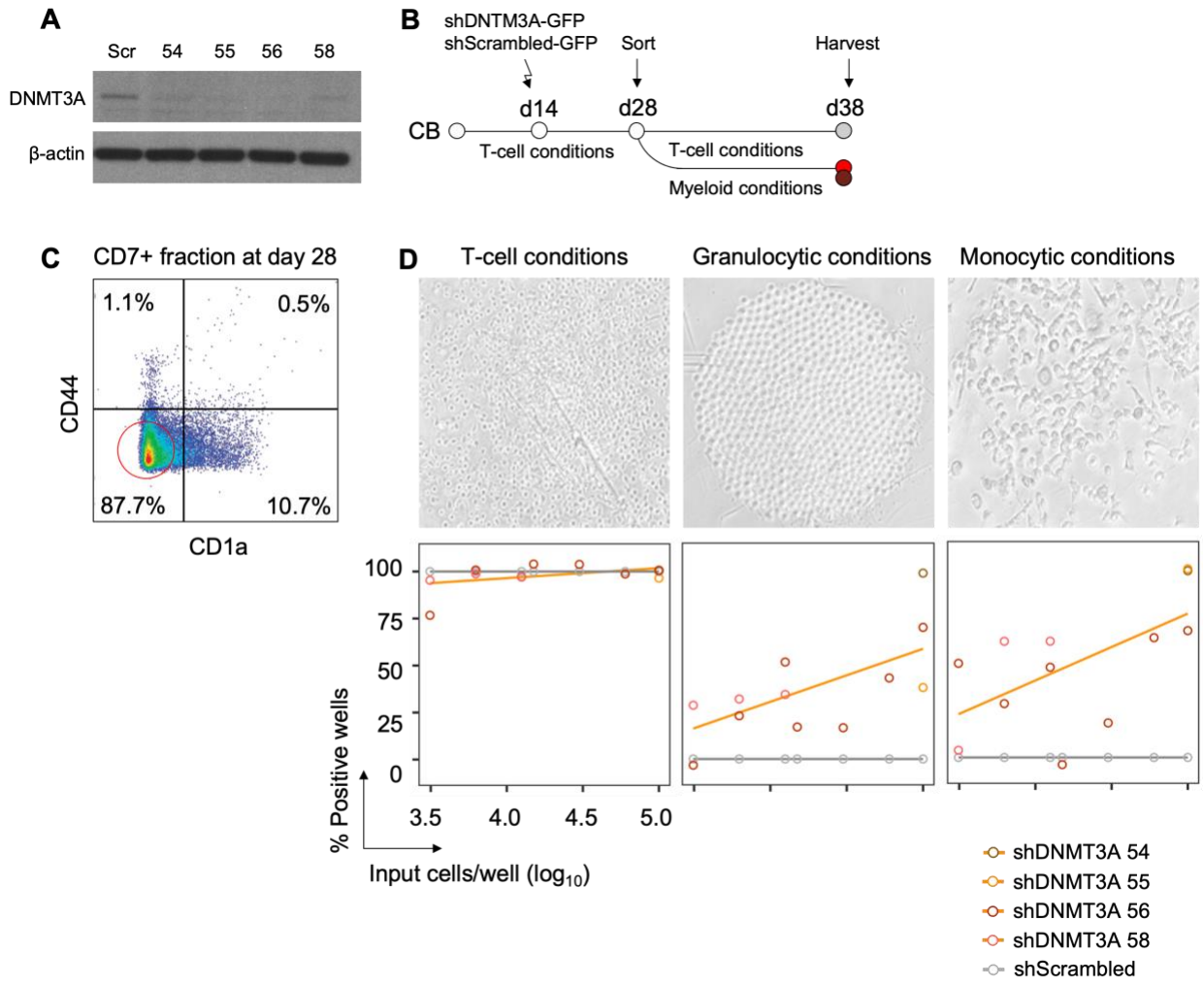


Figure 4.1 shRNA-mediated knockdown of *DNMT3A* in T-cells correlates with increased growth in myeloid culture conditions.

A Western blot analysis of DNMT3A protein levels upon shRNA-mediated *DNMT3A* knockdown (54, 55, 56, 58) compared to a scrambled control (Scr). B-actin was used as a loading control.
B General experimental design of media transfer experiments. T-cells were derived *in vitro* from CD34+ cord blood (CB) and lentivirally transduced at day 14. CD7+CD44-CD1a- populations were sorted into conditions supportive of T-cell or granulocytic/monocytic lineages at day 28. Cells were scored for growth and harvested for flow cytometry at day 38.
C Representative flow cytometry plot indicating the levels of CD44 and CD1a expression within the CD7+ subset after 28 days of culture. The CD44-CD1a- population (circled in red) was sorted for media transfer experiments at day 28.
D Top: Representative images of cells after culture in the indicated conditions at time of harvest. Bottom: The percent of positive wells for each experiment as a function of the number of input cells per well. Each dot indicates a different experiment and is coloured according to the shRNA. The yellow regression line summarizes all constructs targeting *DNMT3A*. The grey regression line summarizes the scrambled control. A minimum of 3 replicates were collected for each shRNA, media condition, and cell dilution.

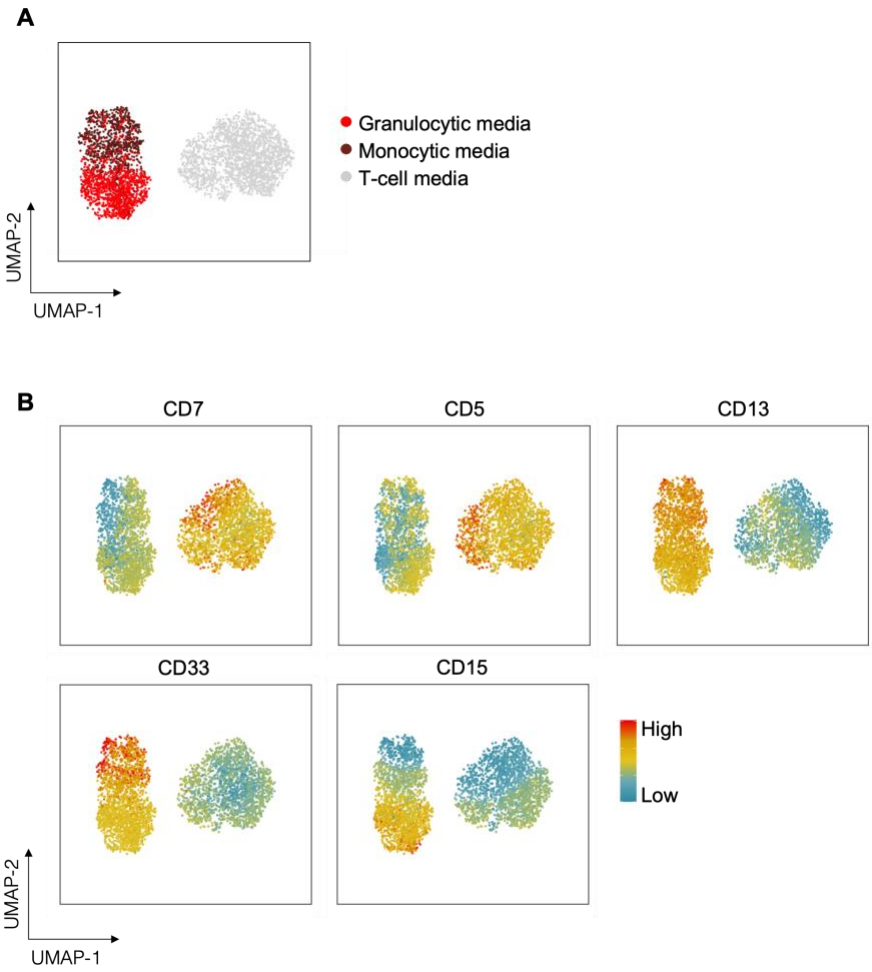


Figure 4.2 14-parameter UMAP distribution of flow cytometric immunophenotypes of *DNTM3A* knockdown T-lineage restricted cells after transfer into T-cell or myeloid growth conditions.

A Cells were cultured in T-cell conditions for 28 days and subsequently transferred into granulocytic, monocytic or T-cell conditions for an additional 10 days. The UMAP distribution indicates cells harvested after 10 days of growth in the indicated media conditions (red, burgundy, or grey).
B The expression of select markers which discriminate between the T-cell or myeloid lineage are shown and are coloured according to relative staining intensity.

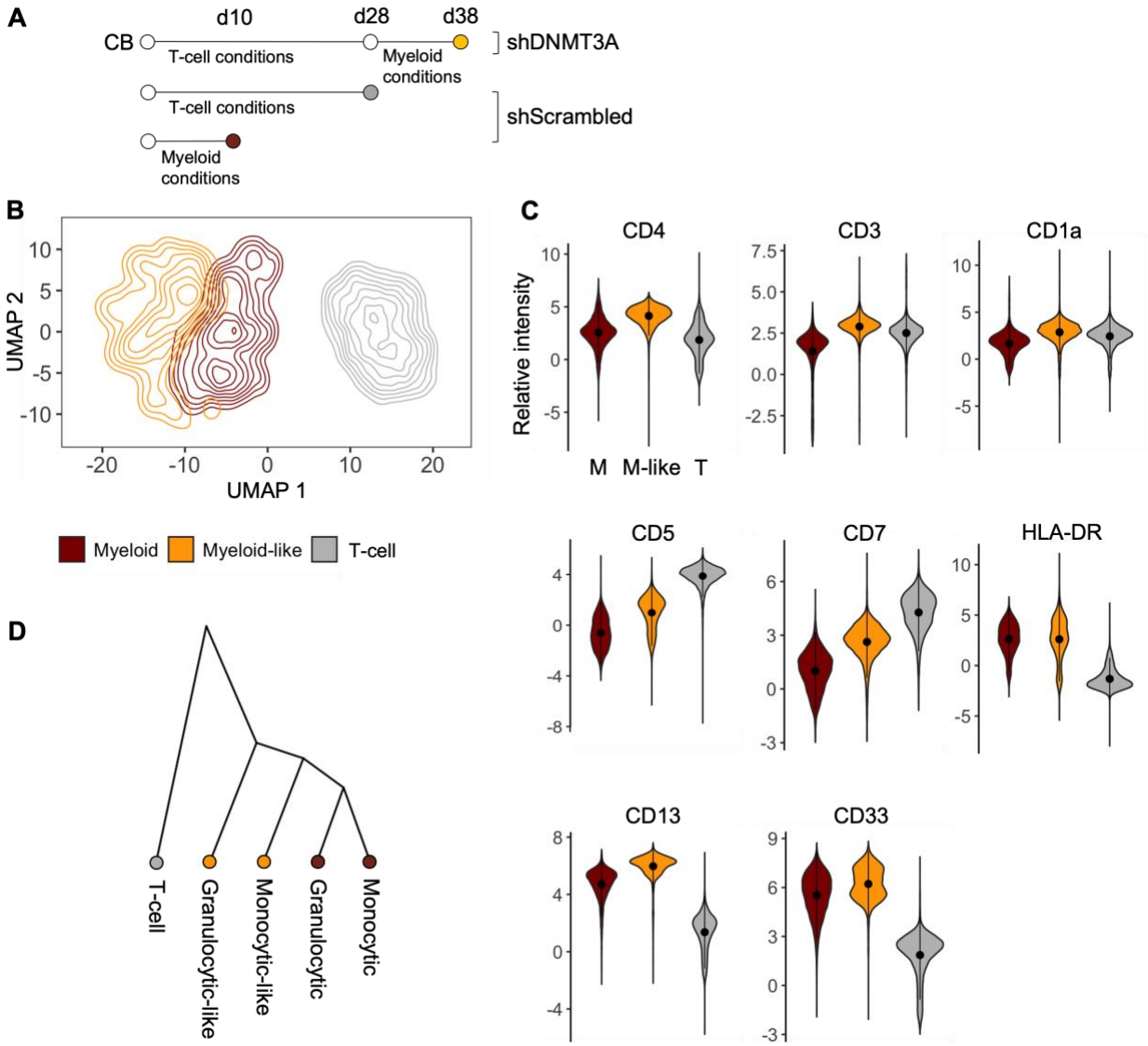


Figure 4.3 T-lineage restricted cells transferred into myeloid growth conditions adopt an intermediate T-lymphoid/myeloid phenotype.

A General experimental design pertaining to analyzed populations. Unmanipulated CD34+ cord blood (CB) cells were cultured in T-cell conditions and harvested at day 28 ("T-cell"; grey), or cultured in myeloid conditions and harvested at day 10 ("Myeloid"; burgundy). *DNMT3A* knockdown T-cells were cultured in T-cell conditions for 28 days, at which point the T-lineage restricted population was sorted into myeloid growth conditions for an additional 10 days ("Myeloid-like"; orange).
B 14-parameter UMAP of flow cytometric immunophenotypes displayed as density distribution of CD34+ cord blood derived T-cells (grey) or CD34+ cord blood derived myeloid cells (dark red), together with *DNMT3A* knockdown T-cells transferred into myeloid growth conditions ("myeloid-like"; orange).
C Violin plots indicating the relative staining intensity of select lineage-discriminating surface markers. M = myeloid, M-like = myeloid like, T = T-cell.

D Hierarchical clustering of myeloid, myeloid-like and T-cell populations based on the median expression of 14 surface markers. Myeloid growth conditions were segregated into granulocytic or monocytic culture conditions.

4.3.1.1 The timing of *DNMT3A* knockdown influences the frequency of myeloid-like cell growth

The developmental context in which *DNMT3A* is lost might influence the frequency of T-cells that retain myeloid potential. That is, knockdown of *DNMT3A* in an early stage of development may have a different outcome than *DNMT3A* knockdown in a later stage of T-cell development. This supposition is based on the observation that *de novo* DNA methylation changes primarily occur during T-cell commitment[196], which possibly indicates that methylation is critical at this stage in T-cell development.

To address this possibility, *in vitro*-derived T-cells were transduced with shRNA targeting *DNMT3A* after 7 or 14 days of culture. After a total of 28 days, the $CD7^{+}CD44^{-}GFP^{+}$ population was sorted at various cell dilutions into media supportive of T-cell, granulocytic or monocytic lineages, as before. Among cells cultured in myeloid conditions, a significant increase in well-initiating frequency was observed in cells transduced after 7 days of culture in comparison to cells transduced after 14 days in culture (Figure 4.4). The difference in well-initiating frequency was not apparent in cells that were maintained in T-cell conditions throughout the duration of the experiment, altogether suggesting that the developmental context of *DNMT3A* loss may influence the frequency of $CD7^{+}CD44^{-}$ T-cells that are capable of proliferating in myeloid conditions.

Altogether, our results may suggest that *DNMT3A* knockdown delays or disrupts T-cell lineage restriction and confers the ability of T-cells to proliferate and differentiate under myeloid growth conditions. This effect was most pronounced when knockdown was performed after 7 days of *in vitro* differentiation, as opposed to 14 days, which could suggest that expression of *DNMT3A* is crucial for enforcement of the T-cell lineage during the first 2 weeks of *in vitro* differentiation. We considered that this may be through its role as a *de novo* DNA methyltransferase; our analysis of CpG methylation changes during *in vitro* T-cell differentiation (discussed in Chapter 2) suggested that promoter methylation notably increases in regulatory regions of the myeloid "master regulator" *SPI1* beginning around 10 days in culture (Figure 3.12). We suspect that *DNMT3A* knockdown at day 7 may prevent these methylation events from occurring, however further experiments which directly link DNTM3A with *de novo* methylation over this region are needed.

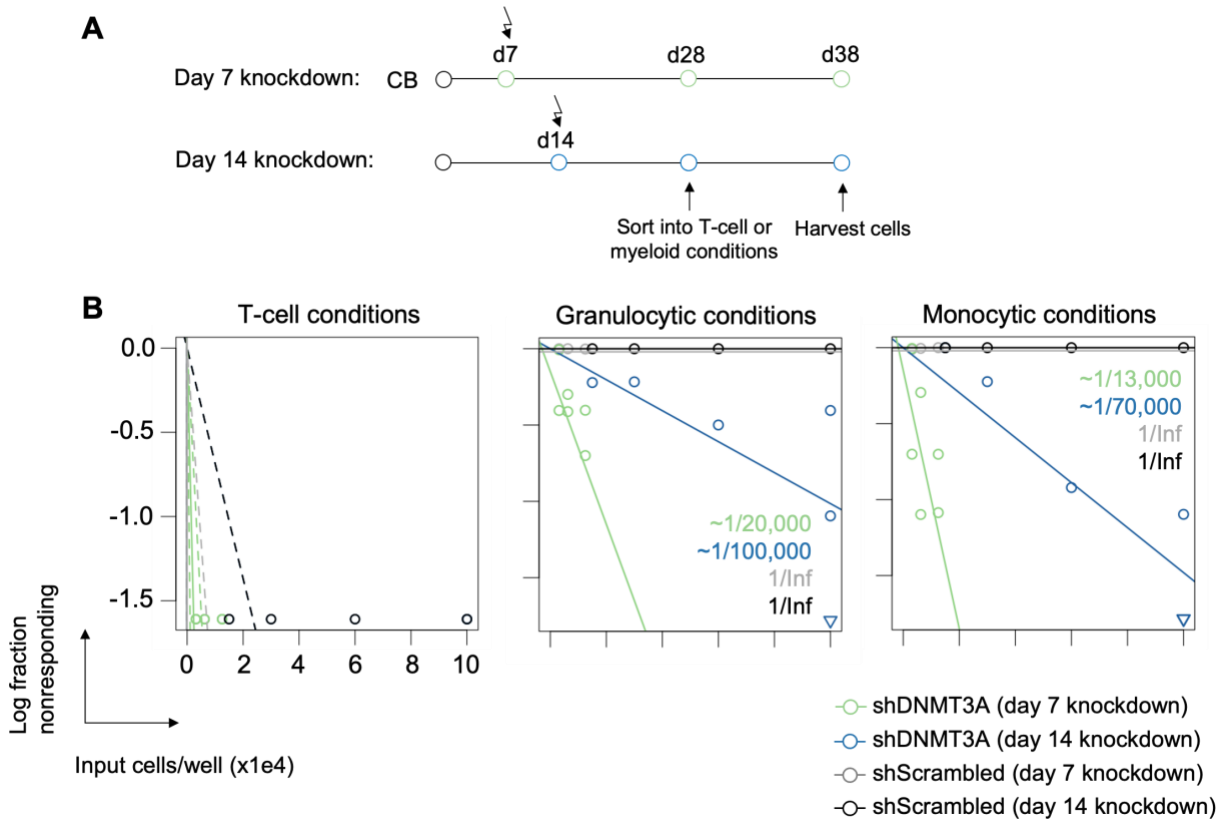


Figure 4.4 Stage-specific *DNMT3A* knockdown influences the proportion of T-lineage restricted myeloid-like cells.

A Experimental design of media transfer experiments. The jagged arrow indicates the timepoint of lentiviral transduction. Cells were maintained in T-cell media conditions for a total of 28 days prior to sorting into myeloid growth conditions. All wells were harvested for flow cytometry at day 38.
B Limited dilution growth assays measuring the frequency of growth in sorted T-cell populations in the indicated culture conditions. Each dot indicates the proportion of positive wells with a minimum of 3 replicates for each tested shRNA, media condition, and cell concentration. Inf = infinity.

4.3.1.2 Selection for *DNMT3A* loss occurs in myeloid-like populations

To further validate our findings, we knocked out *DNMT3A* in developing T-cells by electroporating Cas9 protein complexed with guide RNA designed to target the catalytic loop of DNMT3A (referred to throughout as sgDNMT3A 59R, 54R and 15R), or a guide designed to target the control gene, *EMX1*[330] (Appendix D). Cells were electroporated at an early timepoint in *in vitro* differentiation, and as in previous experiments, the $CD7^{+}CD44^{-}$ cells were sorted into pro-myeloid media after a total of 28 days of culture in

media supportive of T-cell differentiation (Figure 4.5). Among the $CD7^{+}CD44^{-}$ population, cells expressing CD13 or CD33 were removed during cell sorting to exclude the hematopoietic progenitor or myeloid compartments[333]. At the end of the experiment, wells were scored as positive or negative based on cell growth, as described for previous experiments. To determine the indel frequency at the *DNMT3A* locus, we also harvested genomic DNA for Sanger sequencing from each of the positive wells. Genomic DNA from bulk cultures was also collected at various timepoints before cells were transferred into myeloid growth conditions (day 28, and for some cultures, day 7) to compare changes in indel frequency over time. Overall, we observed a gradual increase in indel frequency between day 7 and day 28, which pertains to the period in which cells were cultured in T-cell differentiation media (indicated by filled circles) (Figure 4.5). This was punctuated by a rapid increase in indel frequency upon transfer into myeloid conditions (indicated by open circles). The majority of indels (~70%) were predicted to result in a frameshift mutation. In instances where the indel frequency at day 28 was already high (i.e., sgDNMT3A 59R), the increase in indel frequency upon transfer to myeloid conditions was less apparent; this is likely due to the indel frequency approaching the upper limit of detection. Consistent with our previous knockdown experiments, cells electroporated with the control guide did not grow in myeloid conditions.

Altogether, these results suggest that *DNMT3A* loss may generally enhance cell proliferation based on a gradual increase in indel frequency when cultured in T-cell differentiation conditions. There is, however, a much stronger selection pressure for *DNMT3A* loss among the T-lineage restricted population that is capable of growth in myeloid conditions. These findings, which are in accordance with our previous shRNA knockdown experiments, provide further evidence that *DNMT3A* loss may alter myeloid lineage restriction in T-cells.

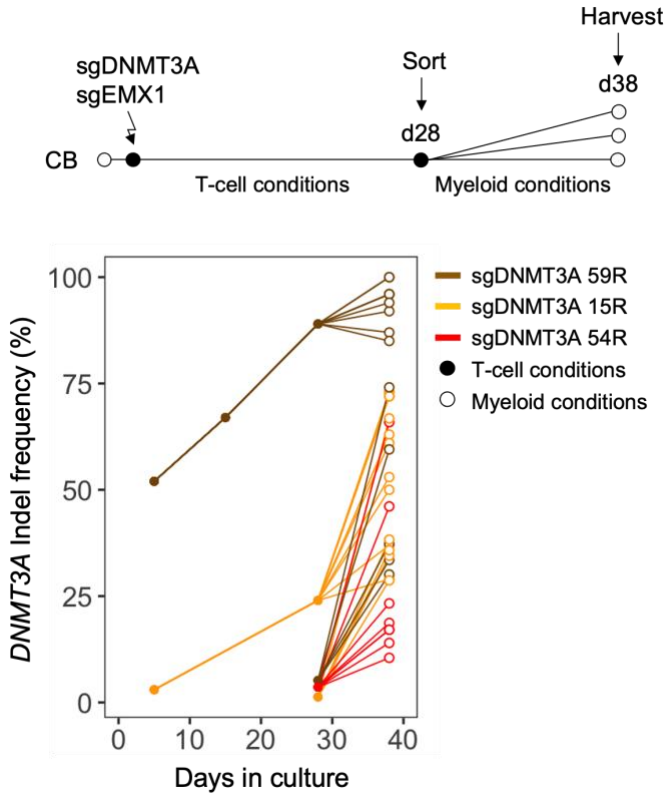


Figure 4.5 Selection for *DNMT3A* loss occurs in CRISPR/Cas9 edited populations upon transfer to myeloid conditions

Cells were electroporated with RNPs targeting *DNMT3A* or *EMX1* (control) in bulk and the T-lineage restricted population was sorted into multiple wells containing myeloid growth media at day 28. Wells were harvested at day 38 for flow immunophenotypic profiling and genotyping to assess *DNMT3A* indel frequency. Filled dots indicate the *DNMT3A* indel frequency of bulk cultures cultured in T-cell conditions. Open circles indicate the *DNMT3A* indel frequency of individual wells after culture in myeloid growth conditions.

4.3.2 *DNMT3A* knockout does not demonstrably alter T-cell differentiation

We considered that the apparent T-cell commitment defect induced by *DNMT3A* knockdown and knockout could instead be rationalized by a differentiation delay in *DNMT3A*-depleted cells. To address this concern, cells were electroporated with a guide targeting *DNMT3A* (sgDNMT3A 59R) or the control gene *EMX1*[330], and were subsequently single-cell sorted (Figure 4.6A; referred to as "*DNMT3A*^{WT}"). Once a clonally expanded population was achieved, genomic DNA was collected for Sanger sequencing at day 28

for each single-cell-derived clone. A portion of each clonal population was also harvested for flow cytometric profiling.

We then compared the phenotypes of clonally derived populations to unmanipulated, bulk *in vitro*-derived T-cells, which were collected at the same timepoint (day 28). As multiple markers were used to interrogate the stage of T-cell development (including CD5, CD7, CD38, CD44, CD1a, CD4, CD8, and CD3), dimensionality reduction was first performed using UMAP to visualize the unmanipulated T-cells in 2D space. *k*-means clustering was performed to classify cells based on phenotype. Accordingly, each cell was grouped into one of four major clusters (Figure 4.6B). Each cluster could be roughly ascribed to developmental stages throughout T-cell differentiation based upon the pattern of CD marker expression (Figure 4.6C,D).

The UMAP clusters generated from unmanipulated cells (i.e., *DNMT3A* wildtype) were then used as a developmental schema upon which to project the single-cell-expanded CRISPR clones (Figure 4.6E). When comparing the cluster composition from each clonal population grouped by genotype (i.e., *DNMT3A* knockout, heterozygous, or wildtype), we did not observe a significant difference in the percent of cells in each cluster (Figure 4.6F); *DNMT3A*^{KO} and *DNMT3A*^{HET} cells displayed a spectrum of differentiation states that did not differ demonstrably from control cells, suggesting that, on a per-cell basis, there is no clear differentiation defect upon *DNMT3A* loss. These results are in accordance with previous studies, in which *Dnmt3a* knockout in murine HSCs did not demonstrate a consistent T-cell differentiation arrest[195].

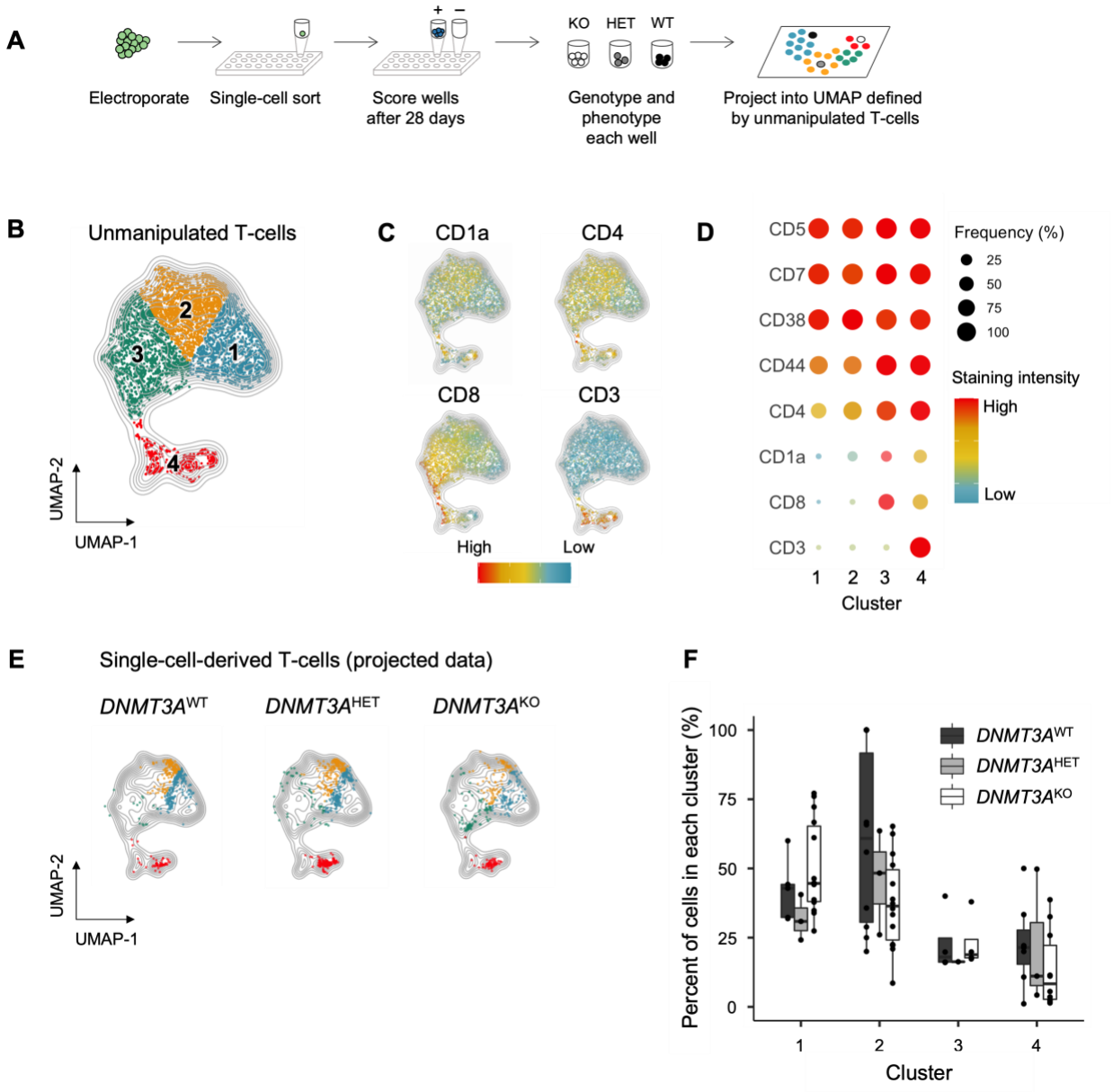


Figure 4.6 T-cell differentiation in single-cell-derived CRISPR clones.

A General experimental design and analysis. Cells were electroporated with RNPs targeting *DNMT3A* or *EMX1* (control) and were subsequently single-cell sorted. Wells were scored as positive (cell growth was observed) or negative (no cell growth was observed) after 28 days of culture. Positive wells were harvested for genotyping and flow cytometric profiling. Flow cytometric data was then projected into UMAP space defined by unmanipulated T-cells cultured in bulk for 28 days.
B 8-parameter UMAP distribution of flow cytometric immunophenotypes of unmanipulated cells cultured in T-cell growth conditions for 28 days. *k*-means clustering was performed and resulted in 4 phenotypic clusters (1-4).
C Expression of select CD markers.

D Dot plot indicating the expression of CD markers for each UMAP cluster. Frequency indicates the percent of cells positive for each CD marker based on manual gating and is displayed by the size of each dot. **E** Single-cell-derived clones projected into the UMAP distribution defined by unmanipulated T-cells and segregated according to genotype (WT, wildtype; HET, *DNMT3A*+/-, KO, *DNMT3A*-/-).
F Boxplot indicating the percent of genotyped cells in each *k*-means cluster. No significant differences were observed (One-way ANOVA).

4.3.3 Myeloid-like cell potential is retained in the G-CSF, M-CSF and/or GM-CSF receptor expressing population

Although *DNMT3A* loss in the T-lineage restricted population correlated with enhanced growth in myeloid conditions, not every cell displayed this capability upon *DNMT3A* knockdown or knockout; the frequency of this population was quite rare (down to 1/100,000 cells) indicating an opportunity to further enrich for this subset of cells. We hypothesized that our population of interest may be enriched among cells expressing surface receptors that bind to cytokines present in the myeloid growth conditions. Most notably, this includes the granulocyte-, macrophage-, and granulocyte-macrophage colony stimulating factors (G-CSF, M-CSF, and GM-CSF, respectively), which promote the proliferation and differentiation of the myeloid lineage[334].

To determine if the expression of G-, M-, or GM-CSF receptors correlates with enrichment of our population of interest, cells were electroporated with RNPs which target *DNTM3A*, or the control guide which targets *EMX1.* After culture for 28 days in T-cell differentiation media, $CD7^{+}CD44^{-}CD13^{-}CD33^{-}$ cells were sorted into myeloid conditions according to G-, M-, or GM-CSF receptor expression (Figure 4.7A). Among the receptor-expressing cells, I observed enrichment of T-lineage restricted cells that were able to grow in myeloid conditions, with a frequency of 1/300 – 1/3000 cells (Figure 4.7C). This number likely represents an underestimate of the true frequency, since the genotypes were not assessed prior to sort, and we expect that many input cells were unedited (i.e., *DNMT3A* wildtype) and therefore unlikely to grow. When we focused on the population lacking expression of the G-, M-, or GM-CSF receptors, we found a relative depletion of myeloid-competent cells (40-80 wells tested per sgRNA) (Figure 4.7B). Consistent with our previous experiments, growth was not observed in cultures electroporated with the EMX1 control. Altogether, our results suggest that the expression of the G-, M-, or GM-CSF receptor can enrich for T-lineage restricted cells capable of growing in myeloid media upon *DNMT3A* knockout.

A breadth of research suggests that there is an inverse correlation between DNA methylation and gene expression, and altogether implicate DNA methylation in gene repression[131–138]. We therefore hypothesized that *DNMT3A* loss during early T-cell development might result in impaired repression of genes encoding for the G-, M-, and GM-CSF receptors, ultimately leading to their increased surface expression. To determine if *DNMT3A* loss might increase the expression of these receptors at the protein level, we performed *DNMT3A* knockdown with shRNA targeting *DNMT3A,* or a scrambled control (Figure 4.8A, orange vs grey). Compared to control, no clear difference was observed in the level of expression of the G-, M-, or GM-CSF receptors, even when comparing the expression levels at the 99th percentile (Figure 4.8B, orange vs grey), suggesting altogether that *DNMT3A* knockdown does not demonstrably alter the expression levels of these receptors, nor the proportion of cells that express the receptors. These results were also consistent upon DNMT3A knockout with sgRNA (data not shown). I also examined the G-, M-, and GM-CSF receptor expression of *DNMT3A* knockout myeloid-like cells, which had been transferred into myeloid media at day 28, as indicated in blue (Figure 4.8A). These cells expressed markers consistent with the myeloid lineage, including CD13 and CD33, and displayed increased expression of all three cytokine receptors (Figure 4.8B, blue histograms). This result may indicate that, for a T-lineage restricted cell to grow in myeloid media conditions, there is selection for cells with high levels of receptor expression or for cells capable of eventually upregulating the G-, M-, or GM-CSF receptors upon stimulation.

Consistent with our observation that *DNMT3A* knockout does not alter the expression of the G-, M-, or GM-CSF receptors at the cell surface, we found no increase in *de novo* DNA methylation within the promoter regions of genes encoding for the G- or M-CSF receptor throughout T-cell differentiation (Figure 4.9). Taken together, our results suggest that DNMT3A may not regulate the expression of the G-, M-, or GM-CSF receptors, and may instead act downstream of receptor activation.

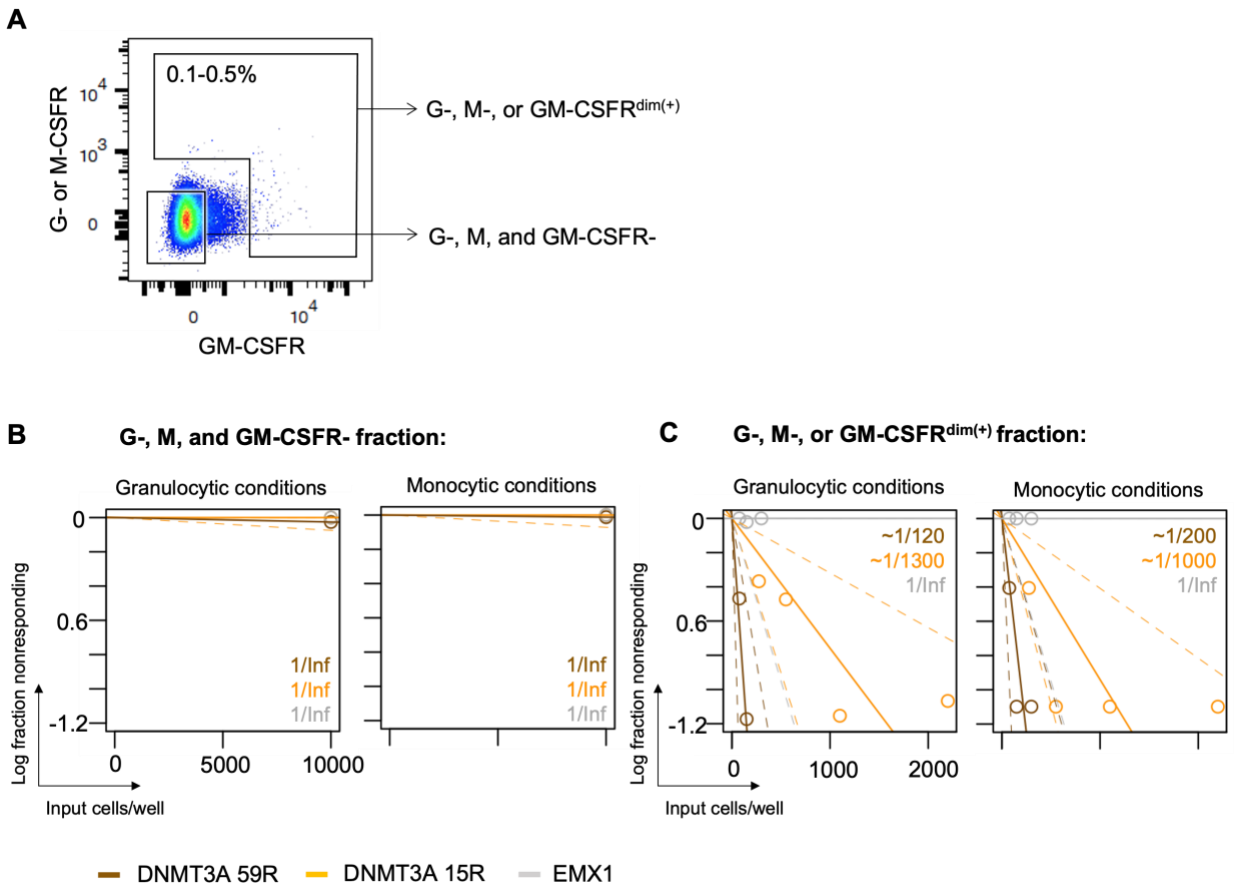


Figure 4.7 Myeloid-like cells are enriched in the G-CSF, M-CSF, and GM-CSF receptor-expressing populations within T-lineage restricted compartment.

A Representative flow plot indicating the G-CSF, M-CSF, and GM-CSF receptor expression in CD7+CD44-CD13-CD33- T-cells after 28 days of culture.
B Limited dilution growth assays measuring the frequency of growth in the receptor negative population (G-, M-, and GM-CSFR-) in granulocytic or monocytic culture conditions after electroporation with RNPs targeting *DNMT3A* or *EMX1* (control).
C Limited dilution growth assays measuring the frequency of growth in the receptor dim-positive population (G-, M-, and GM-CSFR$^{dim/+}$) in granulocytic or monocytic culture conditions after electroporation with RNPs targeting *DNMT3A* or *EMX1* (control). Inf = infinity.

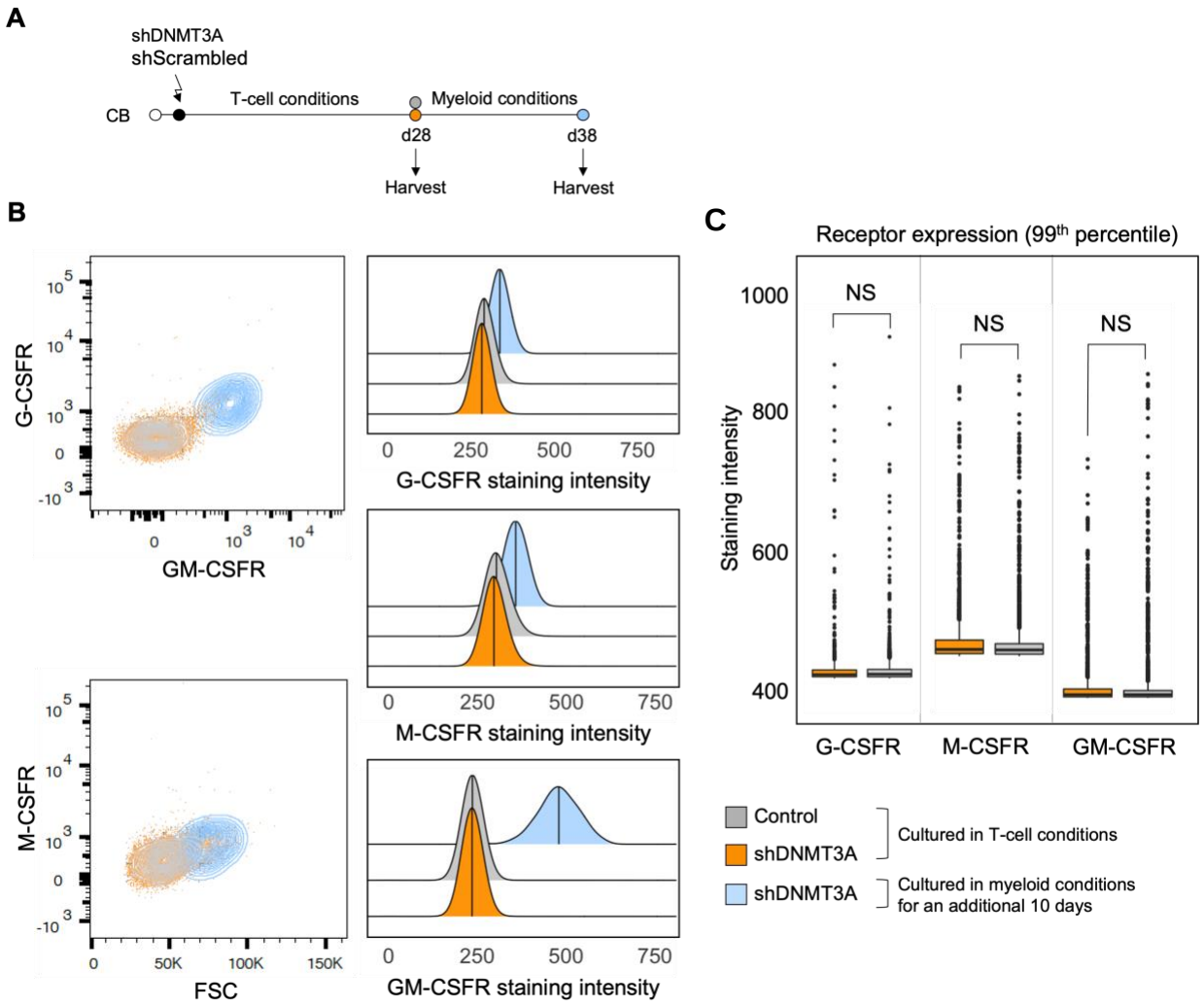


Figure 4.8 Analysis of the expression of G-CSF, M-CSF, and GM-CSF receptors by flow cytometry.

A Experimental design and sample collection. *DNMT3A* knockdown or control T-cells were harvested for immunophenotypic profiling by flow cytometry after 28 days of culture in T-cell growth conditions (shDNMT3A, orange; control, grey) or for an additional 10 days in myeloid media conditions (blue).
B Left: Contour plots indicating the expression of G-CSFR, M-CSFR and GM-CSFR by flow cytometric analysis; Right: Expression data plotted as histograms. The black line indicates the sample median.
C Boxplots indicating the level of expression of G-CSFR, M-CSFR and GM-CSFR by flow cytometric analysis. The staining intensity is shown for populations in the 99th percentile of expression. NS = not significant.

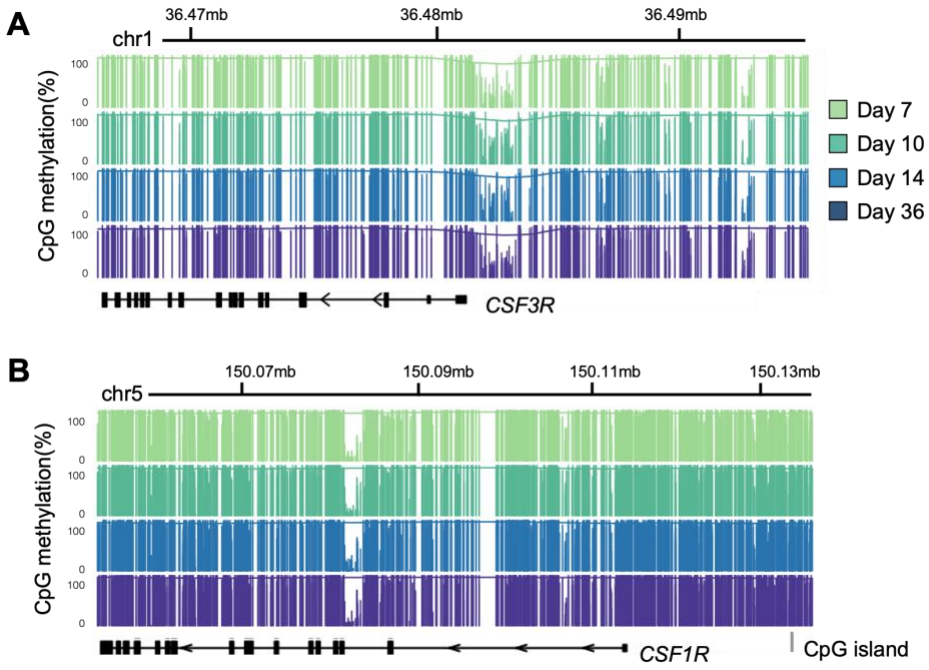


Figure 4.9 Levels of CpG methylation over the *CSF3R* and *CSF1R* loci.

WGBS data indicating the levels of DNA methylation across the promoter region and gene body of *CSF3R* and *CSF1R* (encoding G-CSFR and M-CSFR, respectively). Tracks are coloured according to the sample collection time point. Within each track, the height of each bar indicates the level of CpG methylation for individual CpGs between 0 to 100%. CpG islands are indicated in grey boxes. The RefSeq gene annotation for each gene is displayed at the bottom of the tracks.

4.3.4 Knockdown of *DNMT3A* induces proliferation in response to G-, M-, and GM-CSF in human T-ALL cell lines

It was perhaps surprising that a subset of lineage-restricted T-cells expressed the G-, M-, or GM-CSF receptors. To exclude the possibility that this population was an artifact of *in vitro* culture, we examined receptor expression in additional T-cell contexts including human post-natal thymus (PNT) and T-ALL cell lines. To limit possible contamination of progenitor populations within the PNT samples, analysis was restricted to $CD45^{+}CD3^{+}lin^{-}$ mature DP or CD4/CD8 SP T-cells. Across 2 PNT donors and 4 T-ALL cell lines analyzed, we observed a population of cells with dim positive expression of the G-, M-, or GM-CSF receptors, suggesting that this population is present in multiple T-cell contexts (Figure 4.10).

Motivated by this finding, we wondered if *DNMT3A* knockdown in a T-ALL context might alter cellular response in myeloid growth conditions, as was observed in our previous experiments in *in vitro*-

derived T-cells. To this end, the T-ALL cell line RPMI-8402 was lentivirally transduced with two GFP-tagged shRNAs targeting *DNMT3A,* or a scrambled control. Cells were cultured for a period of 2 weeks to allow for DNA methylation changes to occur before sorting the GFP+, G-, M-, GM-CSFR$^{dim(+)}$ population into base media, or media supplemented with G-CSF, M-CSF, or GM-CSF (referred to throughout as "myeloid cytokines"; Figure 4.11A). After 10 days, all wells were harvested and the number of GFP+ cells were counted. Compared to the scrambled control, *DNTM3A* knockdown in RPMI-8402 resulted in a significant increase in cell number when cultured in media supplemented with G- and GM-CSF, or M- and GM-CSF (Figure 4.11B). This was only evident among the G-, M-, or GM-CSFR$^{dim(+)}$ population. These results suggest that *DNMT3A* loss may enhance cellular response to myeloid cytokines in T-ALL cell lines, as measured by increased proliferation.

As the mechanisms underlying enhanced proliferation in response to G-, M-, and GM-CSF in *DNMT3A* knockdown populations were unclear to us, we considered the possibility that *DNMT3A* knockdown might result in differential intracellular signaling response downstream of the G-, M-, or GM-CSF receptors. At least in the context of myeloid cells, activation of these receptors can lead to increased phosphorylation STAT3, STAT5, and ERK[335–338]. To determine if this was also true in *DNMT3A* knockdown T-ALL cells, we starved *DNMT3A* knockdown and control RPMI-8402 cells in serum-free media for 24 hours, then stimulated the cells with a cocktail containing G-, M-, and GM-CSF, or with serum-free media as an unstimulated control (Figure 4.12A). We then harvested cells to analyze the levels of phosphorylated STAT3 (pSTAT3), pSTAT5, and pERK.

In response to cytokine stimulation, RPMI-8402 cells did not show consistent phosphorylation of ERK or STAT3, even when parsing cells according to receptor expression (Figure 4.12C,D). This suggests that ERK and STAT3 are not activated downstream of G-, M-, and GM-CSF stimulation in a T-ALL context. To our surprise, we did observe an increase in pSTAT5 expression, which was limited to the G-, M-, or GM-CSF receptor expressing population (Figure 4.12E). Notably, there were no discernable differences in STAT5 activation between *DNMT3A* knockdown and scrambled populations when comparing each population to the unstimulated control. Together, these results may indicate that cellular proliferation in response to G-, M-, and GM-CSF upon *DNMT3A* knockdown may be regulated instead at the level of transcriptional control, although it will also be of interest to determine if global levels of STAT5 protein are

altered upon *DNMT3A* loss. We suspect that *DNMT3A* knockdown might render certain loci more susceptible to transcriptional activation, perhaps as a result of DNA methylation changes which increase the accessibility of key target genes downstream of STAT5 activation.

Although further experiments are required to address the possibility that *DNMT3A* regulates loci accessibility of downstream target genes, we hypothesize that a key downstream target of signal activation is *SPI1*, which encodes the myeloid "master regulator" PU.1. In some contexts, *SPI1* upregulation has been observed downstream of the GM-CSF and M-CSF receptors, and PU.1 itself has also been linked to upregulation of the G-, M-, and GM-CSF receptors[339–343], which may indicate a positive regulatory loop. In the context of certain hematopoietic progenitor types, PU.1 is important for cell proliferation and growth[344,345], and in some T-ALL contexts may be essential for disease development[346]. Intriguingly, when we analyzed the expression of *SPI1* in an RNA-seq dataset containing 129 T-ALL cases, we observed significantly higher expression of *SPI1* among the T-ALL cases in which *DNMT3A* was mutated (Figure 4.13). Many of the mutations in *DNMT3A* were predicted to result in protein loss-of-function[222]. This may indicate that a regulatory mechanism exists between *SPI1* expression and DNMT3A activity, although further experiments will be required to address these possibilities.

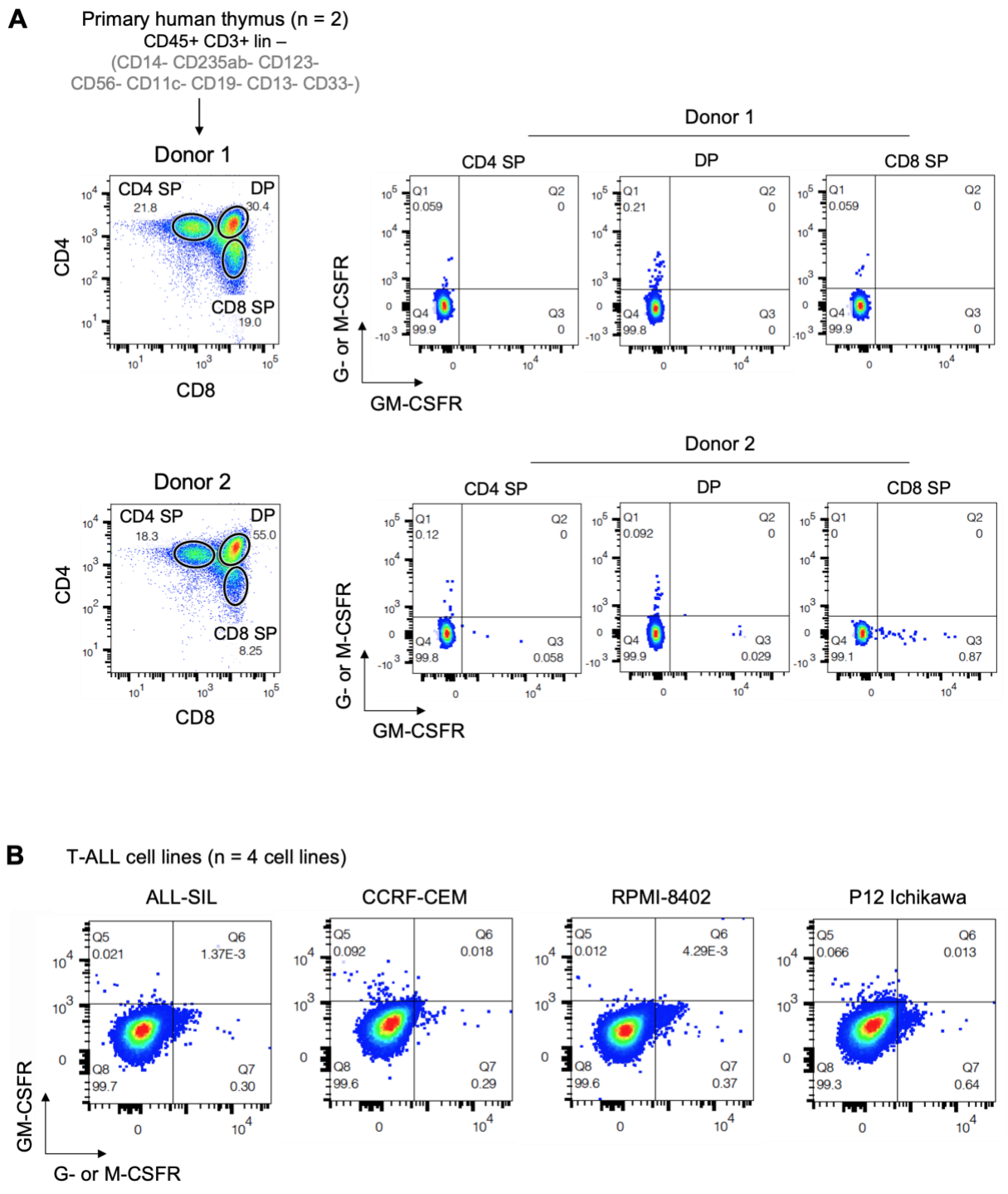


Figure 4.10 G-CSF, M-CSF, and GM-CSF receptor-expressing populations are present in human postnatal thymus and T-ALL cell lines.

A Primary human thymus samples are gated on CD45+CD3+lin- populations as indicated. The expression of G-CSF, M-CSF and GM-CSF receptors are plotted for the CD4+ single positive (CD4 SP), CD8+ single positive (CD8 SP) and CD4+CD8+ double positive (DP) populations

B Expression of G-CSF, M-CSF, and GM-CSF receptors in T-ALL cell lines.

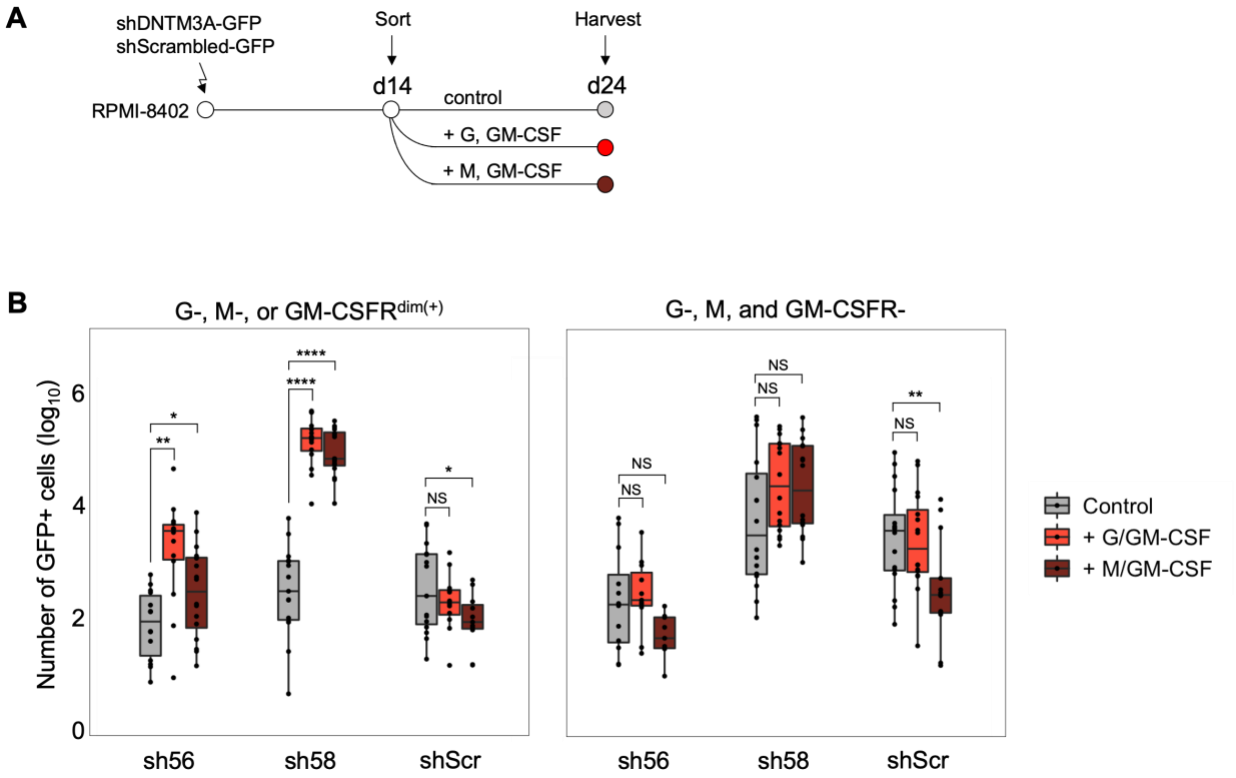


Figure 4.11 Populations expressing the G-CSF, M-CSF or GM-CSF receptors expand in response to the addition of cognate ligands upon DNMT3A knockdown in T-ALL cells.

A General experimental design. RPMI-8402 T-ALL cells were lentivirally transduced and cultured for 14 days prior to sort into conditions supplemented with G-CSF and GM-CSF (red) or M-CSF and GM-CSF (burgundy). Cells were harvested to obtain cell counts after 10 days of growth.
B Numbers of transduced GFP+ cells among the receptor dim (left) or receptor negative (right) populations. (*, P $\leq$ 0.05; ** P $\leq$ 0.01; *** P $\leq$ 0.001; **** P $\leq$ 0.0001; NS, P > 0.05; Dunn's post-hoc test, two-sided, BH multiple test correction). Each dot indicates an individual well.

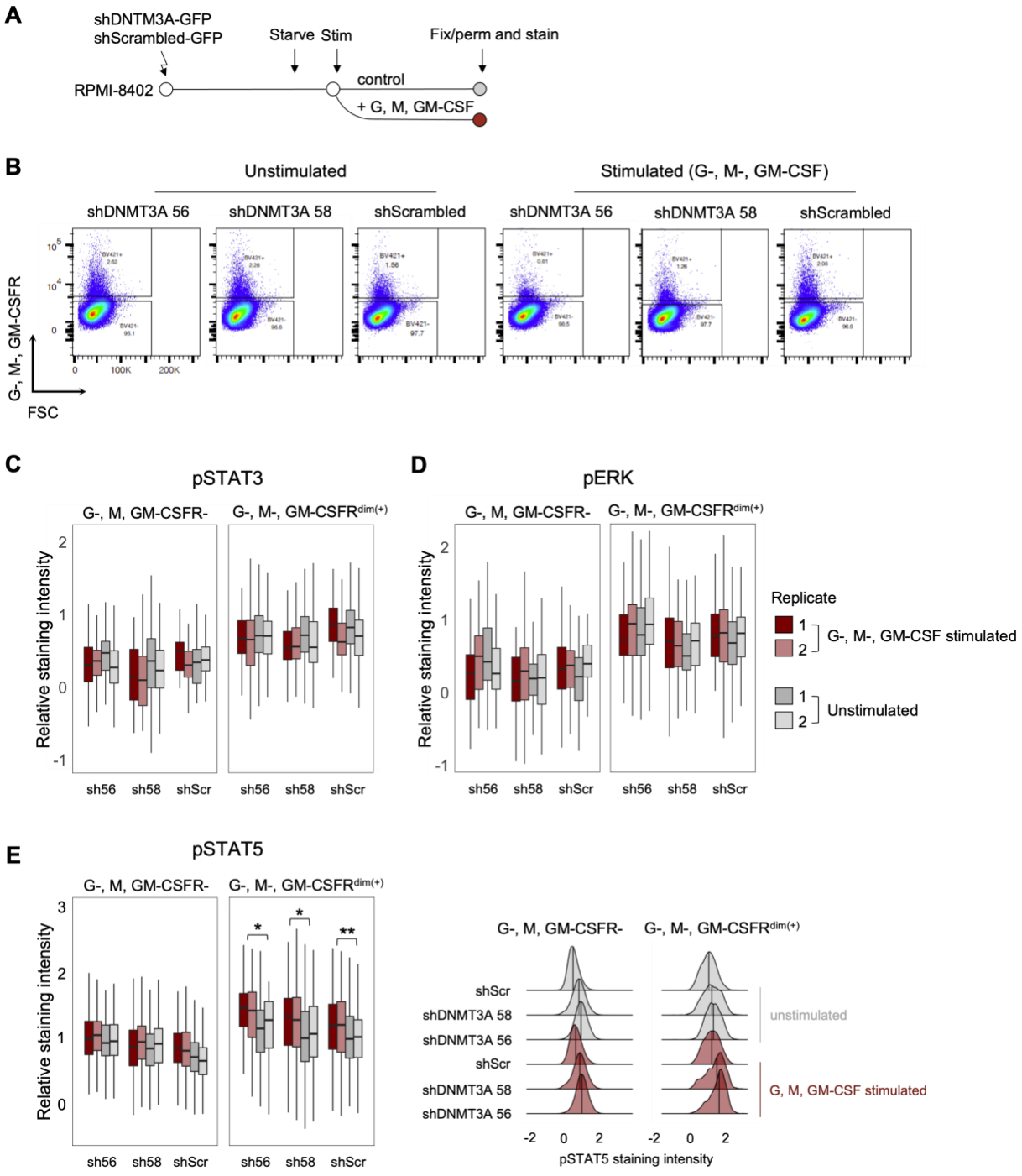


Figure 4.12 STAT5 phosphorylation increases upon stimulation of T-ALL cells with G-CSF, M-CSF, and GM-CSF.

A General experimental design. RPMI-8402 T-ALL cells were lentivirally transduced and cultured for 14 days prior to serum starvation for 24 hours. Cells were stimulated with a cocktail of G-CSF, M-CSF, and GM-CSF. The activation of phosphorylated intermediates was assessed by flow cytometry.

B Expression of the G-CSF, M-CSF, and GM-CSF receptors in unstimulated conditions (left) or upon stimulation with cytokines (right) for each shRNA condition. A representative plot of 2 replicates is shown.
C Phosphorylation of pSTAT3 in the receptor negative (left) or dim/positive (right) populations for each replicate upon stimulation (red) or unstimulated (grey).
D Phosphorylation of pERK.
E Phosphorylation of pSTAT5 shown as a boxplot (left) or as histograms (right). The histograms are representative of 2 replicates. (All significant differences shown; *, $P \leq 0.05$; ** $P \leq 0.01$; Student's t-test, two-sided).

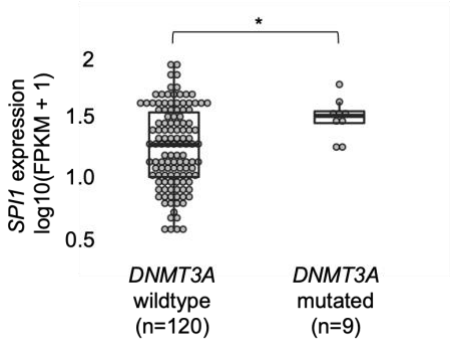


Figure 4.13 *SPI1* is highly expressed in *DNMT3A* mutated T-ALLs.

Expression of *SPI1* (encoding PU.1) in *DNMT3A* wildtype and mutated T-ALL samples. (*, $P \leq 0.05$, Student's t-test, two-sided).

4.4 Discussion

In this chapter, we describe the use of an *in vitro* model of T-cell differentiation to explore the role of DNMT3A in human T-cell development and lineage restriction and extrapolate our findings to human T-ALL. In the context of normal T-cell development, *DNMT3A* loss enhanced cell proliferation and myeloid-like differentiation in the presence of G-, M-, and GM-CSF, which was only observed among the "T-lineage restricted" (i.e., $CD13^-CD33^-CD7^+CD44^-$) T-cell subset with dim positive expression of the G-, M- and GM-CSF receptors. In human T-ALL cells and PNT, we similarly identified a population of T-cells with dim positive receptor expression, and *DNMT3A* knockdown in the context of T-ALL cells resulted in enhanced proliferation among this subset. The proliferation and defective lineage-restriction phenotypes that we observed upon *DNMT3A* loss could not be explained by alterations in levels of surface expression of the G-, M-, or GM-CSF receptors, nor by differential phosphorylation of intracellular signaling factors (i.e., ERK, STAT3 and STAT5), because a discernible difference in *DNMT3A* knockdown vs control cells was not

observed upon analysis of receptor expression and protein phosphorylation. Given the role of DNMT3A in *de novo* DNA methylation, it thus remains to be elucidated if these phenotypes are instead a consequence of differential loci accessibility leading to aberrant transcriptional activation of a proliferative and/or alternative lineage program upon *DNMT3A* loss.

Our findings suggest that, in the context of normal T-cell differentiation, the developmental stage in which *DNMT3A* loss occurs has critical consequences on T-cell lineage restriction; knockdown of *DNMT3A* in $CD7^{+}CD44^{-}$ T-cells at day 14 in culture reduced the overall capacity to give rise to myeloid-like cells compared to knockdown of *DNMT3A* at day 7 in culture. When taken together with our WGBS data collected from developing T-cells, this presents a model in which DNA methylation accrues over select genes as T-cells differentiate, and interruption of this process (e.g., by *DNMT3A* knockdown or knockout) results in defective or incomplete myeloid lineage exclusion. This idea is consistent with lineage specification in embryonic stem cells, in which DNA methylation accumulates over genes associated with pluripotency or alternative tissue types but diminishes over tissue-specific genes, thus acting to restrain alternative developmental potentials and allow for activation of cell-type specific programs[347–349]. Although further studies are required to directly link *DNMT3A* knockdown or knockout with DNA methylation changes in our cultures, we suspect that *de novo* methylation by DNMT3A may repress the expression of genes important for myeloid lineage programs, such as PU.1. At least in the context of murine T-cell development, PU.1 overexpression has been previously shown to re-direct T-cells to myeloid or dendritic cell fate upon cessation of Notch signaling[109]. Though it remains to be determined if the myeloid-like phenotypes that we observe upon *DNMT3A* loss are a result of PU.1 upregulation, our WGBS and RNA-seq results indicate that methylation of the PU.1 promoter increases, and gene expression decreases, throughout normal T-cell differentiation. This might suggest that repression of PU.1, perhaps through DNA methylation, may be an important early event during T-lineage restriction.

Importantly, there remains the possibility that *DNMT3A* loss does not alter lineage commitment but rather has subtle effects on differentiation, and that the phenotypes we observe may be a consequence of uncommitted progenitor populations remaining in the cultures. Although populations were sorted based on $CD7^{+}$, $CD13/33^{-}$, $CD44^{-}$, and G-, M-, or $GM\text{-}CSF^{+(dim)}$ marker expression, we cannot exclude that there is heterogeneity between our *DNMT3A* knockout and wildtype populations within this sorted subset. Although

phenotypic analysis of clonally-derived *DNMT3A* knockout or wildtype populations harvested after 4 weeks in culture did not reveal discernible differences in marker expression, suggesting that differentiation is not grossly altered upon *DNMT3A* loss, addressing this possibility will require higher resolution methods. This may be achieved by expanding our panel of flow cytometric markers or through higher-dimensional single-cell analyses such as single-cell RNA sequencing. Alternatively, the use of other markers of T-cell commitment such as TCRβ rearrangement, which creates an indelible mark in a T-cell's differentiation history, may provide more compelling evidence that the myeloid-like cells we observe are originating from committed T-cell populations.

In the context of clonal and malignant hematopoiesis, the prevalence of *DNMT3A* mutations have been described extensively (reviewed in[198,350]). Indeed, functions in clonal expansion have become increasingly evident due to the dominance of *DNMT3A* mutations (>50%) among the hematopoietic compartment of aging individuals who otherwise display no hematological malignancy[351–353]. The prevalence of *DNMT3A* mutation in mature blood subsets might suggest that mutation arises in an early progenitor[354,355]. This is corroborated by studies in mice, in which *Dnmt3a*-null HSCs display enhanced self-renewal compared to normal counterparts[173,194,356]. Collectively, these studies posit that *DNMT3A* loss provides a selective advantage among the HSC compartment, and malignant transformation may occur upon acquisition of a secondary hit within the clonally-expanded pool. Nonetheless, the functional implication of *DNMT3A* loss among downstream progeny and mature blood subsets remains poorly understood. Our findings here may suggest a novel role of DNMT3A in restricting proliferative signals in response to myeloid cytokines which appears to hold true in normal and malignant T-cell contexts. In the context of T-ALL, where the majority of leukemias have mutations that result in constitutively active Notch and may thus be T-lineage restricted, *DNMT3A* knockdown can provide a selective advantage by enhancing cell expansion in the presence of these same cytokines, thus possibly conferring plasticity in signaling response. We speculate that this may be occurring through de-repression of programs operative in alternative lineages.

Overall, the results presented here provide a model in which DNMT3A constricts lineage-specific responses in T-cells. This is of particular importance during normal development, in which repression of factors associated with stem/progenitor programs must be coordinated with activation of lineage-specific

function. During malignancy, however, failure to properly downregulate early developmental programs may provide growth advantages through inappropriate reactivation of progenitor-associated genes. Thus, in the context of T-ALL, *DNMT3A* loss may provide a selective advantage among a subset of cells which have exited the nurturing thymic microenvironment and have infiltrated into areas replete with factors that, under normal circumstances, do not explicitly support T-cell proliferation and expansion.

Chapter 5: Conclusions

The central objective of the work presented in this thesis was to garner a greater understanding of the molecular events in early human T-cell differentiation to allow for dissection of particular genetic events that occur in specific developmental contexts in T-ALL. This was approached in two ways: First, through a comprehensive expression analysis of T-ALL in comparison to human intrathymic populations, and second, through genetic perturbation and functional characterization of *in vitro*-differentiating T-cell populations. The findings revealed (1) an improved classification of T-ALL which better captures the developmental context in which specific transcription factors operate, and (2) novel insights into a possible role of DNMT3A in restricting lineage-specific signal responses in T-cell development and T-ALL. The results of these approaches and the insight they provide in the greater context of T-cell development and T-ALL, will be highlighted here.

5.1 Incorporation of RNA-seq based developmental profiling improves resolution of phenotypic subsets in human T-ALL

T-ALL has long been characterized as a malignancy of developmentally arrested T-cells. As a proof-of-concept, I extracted gene signatures of normal T-cell development from a publicly available RNA-seq dataset of human intrathymic populations generated by the Crooks lab[286], as described in Chapter 2. This gene signature was used to superimpose two RNA-seq datasets of T-ALL samples upon the spectrum of normal T-cell development, which resulted in a number of sample clusters that correlated with the developmental stage. Although the application of gene signatures from normal T-cells to T-ALL samples has been previously met with success by other groups[226], few have extended this characterization to cover the entire spectrum of T-cell development. The data presented in this chapter thus presents a more comprehensive understanding of T-ALL parsed according to developmental stage.

A key finding presented in this chapter is that immunophenotypic profiling of T-ALL blasts is limited in dimensionality (i.e., only a handful of CD markers may be assessed in parallel), and thus classification, particularly of the ETP-ALL subgroup, lacks sensitivity. This might suggest that immunophenotypic classification – the current standard for developmental classification of T-ALL – does not fully appreciate the true number of ETP-ALL cases, and thus advocates for the use of more comprehensive and inclusive

classification methods. Accurate subgroup classification has demonstrated consequences on treatment approaches in other cancers[357], and precise identification of prognostically poor cases may help guide more prophylactic treatment regimens in T-ALL.

Our approach also revealed biological insight into T-ALL pathogenesis; when parsing the sample clusters by TF subgroups, I observed that many TF subgroups correlated with certain stages of T-cell development. While the synchrony between TF overexpression and developmental context may indicate co-operation between oncogenic TFs and developmental context, perhaps as a result of TF-induced differentiation arrest, it may instead more accurately reflect normal molecular features of the stage of differentiation arrest. In other words, certain TFs may be highly expressed primarily as a consequence of specific developmental contexts but may not be oncogenic effectors themselves. The findings presented in Chapter 2 suggest that both mechanisms may be operative; for example, *LMO2* overexpression is observed in both immature (i.e., *LMO2/LYL*) and cortical (i.e., *TAL/LMO2*) T-ALL, which lie at opposing ends of the differentiation spectrum. *LMO2* overexpression in the former subgroup, which is often observed coincident with *LYL1* overexpression, is likely a consequence of developmental arrest at an immature stage of T-cell development in which *LMO2* and *LYL1* are already highly expressed. In the latter instance, *LMO2* overexpression is often a consequence of translocation, and is typically observed in the context of *TAL1* overexpression. This might suggest the possibility that, in the absence of translocation, high expression of *LMO2* more accurately reflects the developmental context and may not indicate pathogenic mechanism.

With the advent of emerging personalized therapies, care must be taken to ensure segregation of oncogenic effectors from the prevailing stage of differentiation arrest. Thus, accurate classification of the developmental state may pose an important step in understanding disease mechanism. On one hand, precise characterization of the developmental context in which a cell is arrested may provide a clean background against which to compare the effect of gene overexpression. On the other hand, this information may be leveraged to target specific developmental contexts; the TFs which do correlate with particular stages of development might indicate a unique dependence on a particular developmental context, which itself is a product of epigenetic, genetic, and proteomic circumstance. Targeting the specific developmental context upon which oncogenic TFs rely, as opposed to the TFs themselves, may therefore broaden the list of possible candidates for targeted therapy in a subtype-specific manner. Unlike other

disease contexts, targeted therapies are not yet integrated in treatment of T-ALL, although many have been investigated in clinical trials (reviewed in[280]). Successful application of targeted therapy to T-ALL treatment will require dissection of the molecular mechanisms underlying T-ALL pathogenesis, combined with careful consideration of the developmental contexts in which oncogenic TFs operate. The results presented in this chapter provide a small step towards developing a more comprehensive understanding of T-ALL pathogenesis in the context of normal development.

5.2 Identification of DNMT3A as a regulator of lineage-specific responses in developing human T-cells

Specific transcription factors are often operative in a limited range of developmental contexts in T-ALL, but the underlying constraints which may be enforced – by epigenetic factors, for example – are poorly characterized. To better understand how *DNTM3A* loss might be operating in specific developmental stages of T-cell differentiation, I first characterized a commercially available, chemically-defined system to model *in vitro* T-cell development. By gene expression profiling, the *in vitro*-derived subsets recapitulated expression patterns observed in human intrathymic populations. The results presented in Chapter 3 provide a glimpse of these expression profiles, as well as the accompanying DNA methylation patterns. Although an interrogation of early stages of human T-cell development (either through the use of *in vitro* models or by accessing primary thymic tissue), has been accomplished by a few other groups through the use of RNA-seq[196,286,314], and more recently with scRNA-seq[107,358], there are very few studies which have in parallel interrogated DNA methylation changes[196]. The data presented in Chapter 3 represents, to my knowledge, the first RNA-seq and WGBS dataset that has placed more emphasis on characterization of the early stages of human T-cell development that occur prior to CD1a acquisition.

Having established the expression and methylation landscape of *in vitro*-derived T-cells, I then used this system to model the role of *DNMT3A* loss in early T-cell development, particularly in relation to regulating T-lineage restriction and differentiation. The experiments presented in Chapter 4 were motivated by emerging evidence that: (1) DNA methylation patterning is distinct in lymphoid and myeloid lineages[191,192]; (2) that mutation of the *de novo* methyltransferase *DNMT3A* occurs frequently in lymphoid and myeloid malignancies, and is particularly enriched among ETP-ALL[244,245], a subset of T-ALL which

displays myeloid characteristics and which may have poorer prognosis[226]; (3) that the majority of mutations in *DNMT3A* are purported to result in protein loss-of-function based on functional studies[267]; (4) that the degree of *DNMT3A* loss (heterozygous vs complete knockout) can influence the progression to myeloid vs T-cell leukemia[269]; and (5) that loss of the myeloid fate is among the last to occur prior to complete T-cell commitment during T-cell development[107] (reviewed in[104,324]). Taken together, I hypothesized that loss of *DNMT3A* during T-cell development impairs restriction of the myeloid fate, and that understanding the consequence of *DNMT3A* loss might yield important insight into T-ALL biology, and in particular ETP-ALL.

The results presented in Chapter 4 revealed that *DNMT3A* loss correlates with loss of T-lineage restriction associated with an outgrowth of myeloid-like cells under specific experimental conditions, which was most evident when *DNMT3A* knockdown occurred during earlier stages of T-cell differentiation. This is not likely occurring as a result of a differentiation defect at an early stage of T-cell development, where the myeloid lineage may still remain accessible, because no demonstrable phenotypic difference was observed between clonally-derived *DNMT3A*^{WT} or *DNMT3A*^{KO} T-cells. Based on these results, it is tempting to hypothesize that *de novo* DNA methylation patterns are established during early T-cell development and are critical for exclusion of alternative lineages, such as the myeloid lineage, and that perturbation of this process (i.e., by altering *DNMT3A* expression during early stages of development), might impede lineage restriction. Nevertheless, the data presented in this thesis is limited because we have not directly assessed changes in DNA methylation upon *DNMT3A* knockdown; therefore, at present these data cannot directly correlate DNA methylation changes with *DNMT3A* loss. Regardless, our WGBS analysis of developing T-cells in an unperturbed state suggest that the majority of gene promoters do not undergo *de novo* DNA methylation throughout T-cell development, which is consistent with reports in human thymic subsets by others[196]. In the context of the findings reported in Chapter 4, this might indicate that whilst the majority of loci exhibit relatively stable methylation levels throughout differentiation, methylation of key genes associated with alternative (non-T) lineages may occur.

Our WGBS and expression profiling data point to PU.1 as a potential candidate due to increased promoter methylation and decreased gene expression of PU.1 throughout normal T-cell differentiation. Additionally, in the context of murine T-cells, PU.1 overexpression correlates with acquisition of a myeloid-like or dendritic cell-like phenotype and expression of genes associated with the myeloid lineage, which

occurs at the expense of T-cell development[109,359,360]. To elaborate upon the possibility of reprogramming by PU.1, further studies which directly link DNA methylation changes over PU.1 regulatory elements or PU.1 target genes as a consequence of *DNMT3A* loss are required. Of note, reprogramming of murine T-cells by PU.1 overexpression is only evident upon cessation of Notch signaling[109], perhaps suggesting that Notch may also play important roles in T-cell lineage restriction. It may therefore be of future interest to explore the possible role of DNMT3A in regulating accessibility of Notch and PU.1 to target loci.

To our surprise, the myeloid-like phenotype observed in lineage-restricted T-cells upon *DNMT3A* knockdown was enriched among a population of cells that express receptors typically associated with myeloid-lineage differentiation (i.e., the G-CSF, M-CSF, and GM-CSF receptors). Expression of these receptors was also observed in mature DP and SP $CD4^+/CD8^+$ human thymic subsets and in human T-ALL cell lines, suggesting that expression of these receptors normally occurs among a minority of T-cells. The presence of a G-, M-, and/or GM-CSF receptor expressing population in these T-cell contexts was not expected – historically, activation of the G-, M-, and/or GM-CSF receptors by binding cognate ligands has been associated with the proliferation and differentiation of myeloid populations[334]. Nonetheless, G-CSF receptor expression has been previously observed by others in mature $CD4^+$ and $CD8^+$ SP T-cell subsets, and treatment with G-CSF resulted in increased *STAT5* expression[361]. In the context of mature T-cells, the expression of the G-CSF receptor may be important for controlling immune responses[361]. The relevance of this subpopulation in immature T-cells and in T-ALL, however, was only realized upon genetic perturbation. This may point to a general regulatory mechanism in which DNMT3A restricts promiscuous signaling activation prior to T-cell commitment to maintain lineage-appropriate cellular responses. Nonetheless, whether *DNMT3A* loss alters signaling response in additional populations (i.e., not limited to the G-, M-, and/or GM-CSF receptor-expressing cells), or in response to a different assortment of cytokines, remains an unexplored possibility.

In the context of T-ALL, expression of the G-, M-, or GM-CSF receptors bestowed a proliferative advantage in response to the addition of cognate cytokines, which was only evident upon *DNMT3A* knockdown. This suggests that, under certain circumstances, *DNMT3A* knockdown may also confer plasticity in signaling response in T-ALL cells. We hypothesize that the proliferative phenotype that we observe in *DNMT3A* knockdown populations is occurring through increased accessibility of key

downstream target genes, since we did not observe an appreciable difference in surface expression of the G-, M-, or GM-CSF receptors, nor did we observe a difference in activation of downstream signaling factors (i.e., phosphorylation of ERK, STAT3, and STAT5), upon *DNMT3A* loss. Previously, we pointed to PU.1 as a possible downstream target gene given the pattern of methylation and expression during normal T-cell development. In the context of T-ALL, we noted that *SPI1,* which encodes PU.1, is highly expressed among *DNMT3A*-mutated cases. Of note, Spi1 has been previously implicated in T-ALL pathogenesis in a *Pten*-null mouse model, and conditional deletion of *Spi1* resulted in increased disease latency. In a subset of murine blasts and human T-ALL cell lines with low *SPI1* expression, treatment with the pan-DNMT inhibitor, 5-azacytidine, restored levels of *SPI1* expression. Furthermore, treatment of human T-ALL cell lines with 5-azacytidine accelerated development of disease compared to untreated cells upon injection into a mouse[346]. These results may highlight the relevance of PU.1 in T-ALL, and furthermore point to DNA methylation as a possible regulatory mechanism controlling *SPI1* expression.

Taken together, our results suggest that in the context of normal T-cell development and in T-ALL, DNMT3A may act to restrict T-lineage responses. In a normal T-cell context, loss of *DNMT3A* correlates with a loss of myeloid lineage restriction. In the context of T-ALL, loss of *DNMT3A* is associated with a selective growth advantage through enhanced cellular proliferation in response to a more diverse array of signaling molecules. While we hypothesize that this mechanism may be occurring through regulation of key target genes (e.g., PU.1), further experiments are required to determine the molecular mechanism underlying the phenotypes we observe. For example, to determine if this regulatory axis is operating in the context of developing T-cells, we could interrogate CpG methylation within the *SPI1* promoter in DNMT3A knockout/knockdown vs control cells and, in parallel, analyze *SPI1* expression changes. Nonetheless, the results presented in this chapter provide novel insight regarding the role of DNMT3A in maintaining lineage-specific responses in the context of T-cells.

5.3 Limitations of the work and future directions

Although a few studies have interrogated the role of DNMT3A in the context of murine hematopoietic progenitors[173,194,356] and in mouse models of T-ALL[195,268,269], there is a relative paucity of data in a human T-cell context perhaps related to limited accessibility of primary thymic tissue and difficulty

with manipulation of these tissues. To circumvent these challenges, our experiments were based upon the use T-cells differentiated from $CD34^+$ enriched CB progenitors. Interpretation of our results thus requires careful consideration of the model system, and extrapolation of our findings to human thymocyte populations must be proceeded with caution. Limitations of this research pertaining to the use of $CD34^+$ enriched CB progenitors as a source of T-cells include: (1) the effect of the $CD34^+$ progenitor source (i.e., cord blood, bone marrow, or mobilized peripheral blood); and (2) differences between $CD34^+$ CB progenitors and thymus seeding progenitors. Previous studies have indicated that different sources of $CD34^+$ progenitors display different capacities for T-lymphoid and myeloid differentiation[362], which may suggest that tissue source can introduce biases related to developmental capacity. Furthermore, because not every progenitor cell is capable of homing to the thymus, it may be of interest to identify a population of progenitor cells which better approximate the cells which enter the thymus, and which thus initiate T-cell development, in order to more accurately model T-lineage restriction processes. Although several groups have postulated which cells are capable of entering the thymus, identification of the thymus seeding progenitor population remains contentious[4–12,6,100] (reviewed in[3,99]). Nonetheless, identification of a population which more closely resembles thymus seeding progenitors may represent a better progenitor source than cord blood for accurate modeling of T-cell development and lineage-restriction processes.

There are several limitations to drawing meaningful conclusions pertaining to the T-lineage restriction (i.e., myeloid exclusion) experiments discussed throughout Chapter 3 and 4. In these experiments, the $CD7^+CD44^-CD13^-CD33^-$(G-, M-, GM-CSFR)$^{+(dim)}$ population was sorted into conditions supportive of T- or myeloid-lineage development. This population was chosen because loss of $CD44^{dim}$ expression was previously identified to be among the first phenotypic changes that mark T-lineage commitment[95]. Nonetheless, we cannot entirely rule out the possibility that there remain uncommitted T-cell populations within the $CD7^+CD44^-CD13^-CD33^-$(G-, M- ,GM-CSFR)$^{+(dim)}$ *in vitro-derived* T-cell population after 28 days in culture, and that it is the uncommitted T-cell fraction which is responsible for the myeloid-like phenotypes we observe. Although the myeloid-like cells still displayed some vestiges of the T-cell lineage (i.e., higher expression of CD7, CD5, CD3, and CD1a than granulocytic or monocytic cells), interpretation of the studies presented here would greatly benefit from the use of a trackable, irrevocable marker of T-cell commitment. For example, indication of TCRβ rearrangement in our myeloid-like

populations may provide more compelling evidence that these cells had initiated T-cell commitment, and that *DNMT3A* loss is able to divert these committed T-cells to adopt myeloid phenotypes.

Furthermore, a much greater issue is that the series of lineage potential assays described throughout Chapter 3 and Chapter 4 of this thesis do not necessarily reflect developmental processes that occur *in vivo,* but rather only describe events which can occur under highly contrived experimental contexts. It thus remains to be determined if there is any physiological relevance of the findings highlighted in our *in vitro* lineage potential experiments.

The experiments demonstrating that expression of the G-, M-, and/or GM-CSF receptor may provide a proliferative advantage in T-ALL cells upon *DNMT3A* knockdown leave many questions unanswered. For one, the cell line examined displays phenotypic features of an immature T-cell[363], and it may be interesting to see if the growth advantage that we observe can be extrapolated to additional immature T-ALLs and ETP-ALLs, or to T-ALL generally. Additionally, we did not assess DNA methylation pattern changes upon *DNTM3A* loss, nor did we investigate particular loci that may be key in driving proliferation upon stimulation with G-, M-, and GM-CSF. As we did not detect a demonstrable difference in activation of signaling intermediates downstream of the G-, M-, and GM-CSF receptors, the proliferative advantage exhibited by *DNMT3A* knockdown cells upon cytokine stimulation may be instead related to changes in DNA methylation over genes that contribute to a proliferative phenotype. These possible targets, however, remain uncharacterized.

Lastly, the relevance of G-, M-, and GM-CSF receptor expression in normal human intrathymic populations and in T-ALL is yet to be realized. Although the results of these experiments suggest that this population displays a proliferative advantage in a specific context (i.e., upon stimulation with G-, M-, and GM-CSF), whether this relates to plasticity in signaling response generally (e.g., to a wider array of cytokines), remains to be determined. At least in the context of T-ALL, the heterogeneity in response to cytokine stimulation upon *DNMT3A* knockdown may indicate a survival mechanism in which selective advantages are only realized when T-cells encounter different, perhaps less T-cell supportive, environments such as what might be found outside of the thymic microenvironment. Extrapolation of these findings to other T-ALL models, such as additional cell lines or *de novo* synthetic leukemias that we have

previously described[294], and in the context of different receptor/cytokine combinations, may be of future interest.

5.4 Concluding remarks

It is long known that T-ALL represents abated T-cell development, but this knowledge is yet to be leveraged into improved therapies. This will require improved classification of the developmental stage of arrest and further understanding of the molecular drivers of different disease subtypes in specific developmental contexts. In this thesis, I have approached these two issues by (1) incorporating high resolution methods (i.e., RNA-sequencing) to developmental-stage classification of T-ALL, and by (2) interrogating the functional consequences of *DNMT3A* in a defined developmental context. The results of this thesis have made a small step towards improved understanding of the developmental framework in which specific genetic alterations operate, as well as the functional implications of these alterations. Better understanding of the molecular underpinnings of T-ALL by improved classification and with the use of defined developmental models will hopefully accelerate the advancement of targeted therapies for this disease.

References

1. Staal, F. J. T., Wiekmeijer, A.-S., Brugman, M. H. & Pike-Overzet, K. The functional relationship between hematopoietic stem cells and developing T lymphocytes. *Ann. N. Y. Acad. Sci.* **1370**, 36–44 (2016).
2. van der Weerd, K. *et al.* Combined TCRG and TCRA TREC analysis reveals increased peripheral T-lymphocyte but constant intra-thymic proliferative history upon ageing. *Mol. Immunol.* **53**, 302–312 (2013).
3. Bhandoola, A., Boehmer, H. von, Petrie, H. T. & Zúñiga-Pflücker, J. C. Commitment and Developmental Potential of Extrathymic and Intrathymic T Cell Precursors: Plenty to Choose from. *Immunity* **26**, 678–689 (2007).
4. Wu, L. *et al.* CD4 expressed on earliest T-lineage precursor cells in the adult murine thymus. *Nature* **349**, 71–74 (1991).
5. Allman, D. *et al.* Thymopoiesis independent of common lymphoid progenitors. *Nat. Immunol.* **4**, 168–174 (2003).
6. Benz, C., Martins, V. C., Radtke, F. & Bleul, C. C. The stream of precursors that colonizes the thymus proceeds selectively through the early T lineage precursor stage of T cell development. *J. Exp. Med.* **205**, 1187–1199 (2008).
7. Martin, C. H. *et al.* Efficient thymic immigration of B220+ lymphoid-restricted bone marrow cells with T precursor potential. *Nat. Immunol.* **4**, 866–873 (2003).
8. Perry, S. S., Welner, R. S., Kouro, T., Kincade, P. W. & Sun, X.-H. Primitive lymphoid progenitors in bone marrow with T lineage reconstituting potential. *J. Immunol. Baltim. Md 1950* **177**, 2880–2887 (2006).
9. Porritt, H. E. *et al.* Heterogeneity among DN1 prothymocytes reveals multiple progenitors with different capacities to generate T cell and non-T cell lineages. *Immunity* **20**, 735–745 (2004).
10. Benz, C. & Bleul, C. C. A multipotent precursor in the thymus maps to the branching point of the T versus B lineage decision. *J. Exp. Med.* **202**, 21–31 (2005).
11. Krueger, A. & von Boehmer, H. Identification of a T lineage committed progenitor in adult blood. *Immunity* **26**, 105–116 (2007).

12. Lai, A. Y. & Kondo, M. Identification of a bone marrow precursor of the earliest thymocytes in adult mouse. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 6311–6316 (2007).

13. Rossi, F. M. V. *et al.* Recruitment of adult thymic progenitors is regulated by P-selectin and its ligand PSGL-1. *Nat. Immunol.* **6**, 626–634 (2005).

14. Liu, C. *et al.* Coordination between CCR7- and CCR9-mediated chemokine signals in prevascular fetal thymus colonization. *Blood* **108**, 2531–2539 (2006).

15. Zlotoff, D. A. *et al.* CCR7 and CCR9 together recruit hematopoietic progenitors to the adult thymus. *Blood* **115**, 1897–1905 (2010).

16. Hernández-López, C. *et al.* Stromal cell–derived factor 1/CXCR4 signaling is critical for early human T-cell development. *Blood* **99**, 546–554 (2002).

17. Plotkin, J., Prockop, S. E., Lepique, A. & Petrie, H. T. Critical Role for CXCR4 Signaling in Progenitor Localization and T Cell Differentiation in the Postnatal Thymus. *J. Immunol.* **171**, 4521–4527 (2003).

18. Anderson, G. & Takahama, Y. Thymic epithelial cells: working class heroes for T cell development and repertoire selection. *Trends Immunol.* **33**, 256–263 (2012).

19. Wu, L., Kincade, P. W. & Shortman, K. The CD44 expressed on the earliest intrathymic precursor population functions as a thymus homing molecule but does not bind to hyaluronate. *Immunol. Lett.* **38**, 69–75 (1993).

20. Prockop, S. E. *et al.* Stromal cells provide the matrix for migration of early lymphoid progenitors through the thymic cortex. *J. Immunol. Baltim. Md 1950* **169**, 4354–4361 (2002).

21. Matheny, H. E., Deem, T. L. & Cook-Mills, J. M. Lymphocyte migration through monolayers of endothelial cell lines involves VCAM-1 signaling via endothelial cell NADPH oxidase. *J. Immunol. Baltim. Md 1950* **164**, 6550–6559 (2000).

22. Lind, E. F., Prockop, S. E., Porritt, H. E. & Petrie, H. T. Mapping Precursor Movement through the Postnatal Thymus Reveals Specific Microenvironments Supporting Defined Stages of Early Lymphoid Development. *J. Exp. Med.* **194**, 127–134 (2001).

23. Porritt, H. E., Gordon, K. & Petrie, H. T. Kinetics of Steady-state Differentiation and Mapping of Intrathymic-signaling Environments by Stem Cell Transplantation in Nonirradiated Mice. *J. Exp. Med.* **198**, 957–962 (2003).

24. Petrie, H. T. & Zúñiga-Pflücker, J. C. Zoned Out: Functional Mapping of Stromal Signaling Microenvironments in the Thymus. *Annu. Rev. Immunol.* **25**, 649–679 (2007).

25. Blom, B. & Spits, H. Development of human lymphoid cells. *Annu. Rev. Immunol.* **24**, 287–320 (2006).

26. Kumar, B. V., Connors, T. & Farber, D. L. Human T cell development, localization, and function throughout life. *Immunity* **48**, 202–213 (2018).

27. Radtke, F. *et al.* Deficient T cell fate specification in mice with an induced inactivation of Notch1. *Immunity* **10**, 547–558 (1999).

28. Mohtashami, M., Shah, D. K., Kianizad, K., Awong, G. & Zúñiga-Pflücker, J. C. Induction of T-cell development by Delta-like 4-expressing fibroblasts. *Int. Immunol.* **25**, 601–611 (2013).

29. La Motte-Mohs, R. N., Herer, E. & Zúñiga-Pflücker, J. C. Induction of T-cell development from human cord blood hematopoietic stem cells by Delta-like 1 in vitro. *Blood* **105**, 1431–1439 (2005).

30. Koch, U. *et al.* Delta-like 4 is the essential, nonredundant ligand for Notch1 during thymic T cell lineage commitment. *J. Exp. Med.* **205**, 2515–2523 (2008).

31. Hozumi, K. *et al.* Delta-like 4 is indispensable in thymic environment specific for T cell development. *J. Exp. Med.* **205**, 2507–2513 (2008).

32. Heinzel, K., Benz, C., Martins, V. C., Haidl, I. D. & Bleul, C. C. Bone marrow-derived hemopoietic precursors commit to the T cell lineage only after arrival in the thymic microenvironment. *J. Immunol. Baltim. Md 1950* **178**, 858–868 (2007).

33. García-León, M. J., Fuentes, P., de la Pompa, J. L. & Toribio, M. L. Dynamic regulation of NOTCH1 activation and Notch ligand expression in human thymus development. *Dev. Camb. Engl.* **145**, dev165597 (2018).

34. Egerton, M., Shortman, K. & Scollay, R. The kinetics of immature murine thymocyte development in vivo. *Int. Immunol.* **2**, 501–507 (1990).

35. Pearse, G. Normal Structure, Function and Histology of the Thymus. *Toxicol. Pathol.* **34**, 504–514 (2006).

36. Petrie, H. T. Cell migration and the control of post-natal T-cell lymphopoiesis in the thymus. *Nat. Rev. Immunol.* **3**, 859–866 (2003).

37. Godfrey, D. I., Kennedy, J., Suda, T. & Zlotnik, A. A developmental pathway involving four phenotypically and functionally distinct subsets of CD3-CD4-CD8- triple-negative adult mouse thymocytes defined by CD44 and CD25 expression. *J. Immunol. Baltim. Md 1950* **150**, 4244–4252 (1993).

38. Boyd, R. L. *et al.* The thymic microenvironment. *Immunol. Today* **14**, 445–459 (1993).

39. Han, J. & Zúñiga-Pflücker, J. C. A 2020 View of Thymus Stromal Cells in T Cell Development. *J. Immunol. Baltim. Md 1950* **206**, 249–256 (2021).

40. Shi, J., Fallahi, M., Luo, J.-L. & Petrie, H. T. Nonoverlapping functions for Notch1 and Notch3 during murine steady-state thymic lymphopoiesis. *Blood* **118**, 2511–2519 (2011).

41. D'souza, B., Miyamoto, A. & Weinmaster, G. The many facets of Notch ligands. *Oncogene* **27**, 5148–5167 (2008).

42. Li, X. & von Boehmer, H. Notch Signaling in T-Cell Development and T-ALL. *ISRN Hematol.* **2011**, 921706 (2011).

43. Weerkamp, F. *et al.* Human thymus contains multipotent progenitors with T/B lymphoid, myeloid, and erythroid lineage potential. *Blood* **107**, 3131–3137 (2006).

44. García-Peydró, M., Yébenes, V. G. de & Toribio, M. L. Notch1 and IL-7 Receptor Interplay Maintains Proliferation of Human Thymic Progenitors while Suppressing Non-T Cell Fates. *J. Immunol.* **177**, 3711–3720 (2006).

45. Spits, H. Development of αβ T cells in the human thymus. *Nat. Rev. Immunol.* **2**, 760–772 (2002).

46. Huang, E. Y., Gallegos, A. M., Richards, S. M., Lehar, S. M. & Bevan, M. J. Surface expression of Notch1 on thymocytes: correlation with the double-negative to double-positive transition. *J. Immunol. Baltim. Md 1950* **171**, 2296–2304 (2003).

47. Ciofani, M. & Zúñiga-Pflücker, J. C. Notch promotes survival of pre-T cells at the beta-selection checkpoint by regulating cellular metabolism. *Nat. Immunol.* **6**, 881–888 (2005).

48. Wang, H., Pierce, L. J. & Spangrude, G. J. Distinct roles of IL-7 and stem cell factor in the OP9-DL1 T cell differentiation culture system. *Exp. Hematol.* **34**, 1730–1740 (2006).

49. Alves, N. L. *et al.* Characterization of the thymic IL-7 niche in vivo. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 1512–1517 (2009).

50. Shimosato, Y., Mukai, K. & Matsuno, Y. *Tumors of the Mediastinum*. (1997).

51. Young, B., O'Dowd, G. & Woodford, P. *Wheater's Functional Histology*. (2013).

52. Lopes, N., Sergé, A., Ferrier, P. & Irla, M. Thymic Crosstalk Coordinates Medulla Organization and T-Cell Tolerance Induction. *Front. Immunol.* **6**, (2015).

53. Kadouri, N., Nevo, S., Goldfarb, Y. & Abramson, J. Thymic epithelial cell heterogeneity: TEC by TEC. *Nat. Rev. Immunol.* **20**, 239–253 (2020).

54. Buono, M. *et al.* A dynamic niche provides Kit ligand in a stage-specific manner to the earliest thymocyte progenitors. *Nat. Cell Biol.* **18**, 157–167 (2016).

55. Ribeiro, A. R., Rodrigues, P. M., Meireles, C., Di Santo, J. P. & Alves, N. L. Thymocyte selection regulates the homeostasis of IL-7-expressing thymic cortical epithelial cells in vivo. *J. Immunol. Baltim. Md 1950* **191**, 1200–1209 (2013).

56. Brunk, F., Augustin, I., Meister, M., Boutros, M. & Kyewski, B. Thymic Epithelial Cells Are a Nonredundant Source of Wnt Ligands for Thymus Development. *J. Immunol.* **195**, 5261–5271 (2015).

57. Ohigashi, I., Kozai, M. & Takahama, Y. Development and developmental potential of cortical thymic epithelial cells. *Immunol. Rev.* **271**, 10–22 (2016).

58. Akashi, K., Kondo, M., Freeden-Jeffry, U. von, Murray, R. & Weissman, I. L. Bcl-2 Rescues T Lymphopoiesis in Interleukin-7 Receptor–Deficient Mice. *Cell* **89**, 1033–1041 (1997).

59. Maraskovsky, E. *et al.* Bcl-2 can rescue T lymphocyte development in interleukin-7 receptor-deficient mice but not in mutant rag-1-/- mice. *Cell* **89**, 1011–1019 (1997).

60. Foxwell, B. M., Beadling, C., Guschin, D., Kerr, I. & Cantrell, D. Interleukin-7 can induce the activation of Jak 1, Jak 3 and STAT 5 proteins in murine T cells. *Eur. J. Immunol.* **25**, 3041–3046 (1995).

61. Kim, K., Lee, C. K., Sayers, T. J., Muegge, K. & Durum, S. K. The trophic action of IL-7 on pro-T cells: inhibition of apoptosis of pro-T1, -T2, and -T3 cells correlates with Bcl-2 and Bax levels and is independent of Fas and p53 pathways. *J. Immunol. Baltim. Md 1950* **160**, 5735–5741 (1998).

62. Muegge, K., Vila, M. P. & Durum, S. K. Interleukin-7: a Cofactor For V(D)J Rearrangement of the T Cell Receptor β Gene. *Science* **261**, 93–95 (1993).

63. Schlissel, M. S., Durum, S. D. & Muegge, K. The Interleukin 7 Receptor Is Required for T Cell Receptor γ Locus Accessibility to the V(D)j Recombinase. *J. Exp. Med.* **191**, 1045–1050 (2000).

64. Ripen, A. M., Nitta, T., Murata, S., Tanaka, K. & Takahama, Y. Ontogeny of thymic cortical epithelial cells expressing the thymoproteasome subunit β5t. *Eur. J. Immunol.* **41**, 1278–1287 (2011).

65. Murata, S. *et al.* Regulation of CD8+ T cell development by thymus-specific proteasomes. *Science* **316**, 1349–1353 (2007).

66. Gommeaux, J. *et al.* Thymus-specific serine protease regulates positive selection of a subset of CD4+ thymocytes. *Eur. J. Immunol.* **39**, 956–964 (2009).

67. Starr, T. K., Jameson, S. C. & Hogquist, K. A. Positive and negative selection of T cells. *Annu. Rev. Immunol.* **21**, 139–176 (2003).

68. Yu, Q. *et al.* Cytokine signal transduction is suppressed in preselection double-positive thymocytes and restored by positive selection. *J. Exp. Med.* **203**, 165–175 (2006).

69. Tanigaki, K. *et al.* Regulation of αβ/γδ T Cell Lineage Commitment and Peripheral T Cell Responses by Notch/RBP-J Signaling. *Immunity* **20**, 611–622 (2004).

70. Kwan, J. & Killeen, N. CCR7 directs the migration of thymocytes into the thymic medulla. *J. Immunol. Baltim. Md 1950* **172**, 3999–4007 (2004).

71. Heino, M. *et al.* Autoimmune regulator is expressed in the cells regulating immune tolerance in thymus medulla. *Biochem. Biophys. Res. Commun.* **257**, 821–825 (1999).

72. Derbinski, J., Schulte, A., Kyewski, B. & Klein, L. Promiscuous gene expression in medullary thymic epithelial cells mirrors the peripheral self. *Nat. Immunol.* **2**, 1032–1039 (2001).

73. Anderson, M. S. & Su, M. A. Aire and T cell Development. *Curr. Opin. Immunol.* **23**, 198–206 (2011).

74. Scollay, R. G., Butcher, E. C. & Weissman, I. L. Thymus cell migration. Quantitative aspects of cellular traffic from the thymus to the periphery in mice. *Eur. J. Immunol.* **10**, 210–218 (1980).

75. Charles A Janeway, J., Travers, P., Walport, M. & Shlomchik, M. J. Generation of lymphocytes in bone marrow and thymus. *Immunobiol. Immune Syst. Health Dis. 5th Ed.* (2001).

76. Wolfer, A. *et al.* Inactivation of Notch 1 in immature thymocytes does not perturb CD4 or CD8T cell development. *Nat. Immunol.* **2**, 235–241 (2001).

77. Radtke, F., Fasnacht, N. & MacDonald, H. R. Notch Signaling in the Immune System. *Immunity* **32**, 14–27 (2010).

78. Eagar, T. N. *et al.* Notch 1 Signaling Regulates Peripheral T Cell Activation. *Immunity* **20**, 407–415 (2004).

79. Repass, J. F. *et al.* IL7-hCD25 and IL7-Cre BAC transgenic mouse lines: new tools for analysis of IL-7 expressing cells. *Genes. N. Y. N 2000* **47**, 281–287 (2009).

80. Shalapour, S. *et al.* Commensal microflora and interferon-gamma promote steady-state interleukin-7 production in vivo. *Eur. J. Immunol.* **40**, 2391–2400 (2010).

81. Hong, C., Luckey, M. & Park, J.-H. Intrathymic IL-7: The where, when, and why of IL-7 signaling during T cell development. *Semin. Immunol.* **24**, 151–158 (2012).

82. Janeway, C., Travers, P., Mark Walport & Shlomchik, M. J. *Immunobiology: The Immune System in Health and Disease*. (Garland Science, 2001).

83. Kalyan, S. & Kabelitz, D. Defining the nature of human γδ T cells: a biographical sketch of the highly empathetic. *Cell. Mol. Immunol.* **10**, 21–29 (2013).

84. Kang, J., Volkmann, A. & Raulet, D. H. Evidence that gammadelta versus alphabeta T cell fate determination is initiated independently of T cell receptor signaling. *J. Exp. Med.* **193**, 689–698 (2001).

85. Hayes, S. M., Laird, R. M. & Love, P. E. Beyond alphabeta/gammadelta lineage commitment: TCR signal strength regulates gammadelta T cell maturation and effector fate. *Semin. Immunol.* **22**, 247–251 (2010).

86. Kreslavsky, T., Gleimer, M., Garbe, A. I. & von Boehmer, H. αβ versus γδ fate choice: counting the T-cell lineages at the branch point. *Immunol. Rev.* **238**, 169–181 (2010).

87. Dik, W. A. *et al.* New insights on human T cell development by quantitative T cell receptor gene rearrangement studies and gene expression profiling. *J. Exp. Med.* **201**, 1715–1723 (2005).

88. Blom, B. *et al.* TCR gene rearrangements and expression of the pre-T cell receptor complex during human T-cell differentiation. *Blood* **93**, 3033–3043 (1999).

89. Asnafi, V. *et al.* Analysis of TCR, pT alpha, and RAG-1 in T-acute lymphoblastic leukemias improves understanding of early human T-lymphoid lineage commitment. *Blood* **101**, 2693–2703 (2003).

90. Schatz, D. G., Oettinger, M. A. & Baltimore, D. The V(D)J recombination activating gene, RAG-1. *Cell* **59**, 1035–1048 (1989).

91. Wilson, A., Held, W. & MacDonald, H. R. Two waves of recombinase gene expression in developing thymocytes. *J. Exp. Med.* **179**, 1355–1360 (1994).

92. Monroe, R. J. *et al.* RAG2:GFP knockin mice reveal novel aspects of RAG2 expression in primary and peripheral lymphoid tissues. *Immunity* **11**, 201–212 (1999).

93. Krangel, M. S. Mechanics of T cell receptor gene rearrangement. *Curr. Opin. Immunol.* **21**, 133–139 (2009).

94. Langerak, A. W., Wolvers-Tettero, I. L., van Gastel-Mol, E. J., Oud, M. E. & van Dongen, J. J. Basic helix-loop-helix proteins E2A and HEB induce immature T-cell receptor rearrangements in nonlymphoid cells. *Blood* **98**, 2456–2465 (2001).

95. Canté-Barrett, K. *et al.* Loss of CD44dim Expression from Early Progenitor Cells Marks T-Cell Lineage Commitment in the Human Thymus. *Front. Immunol.* **8**, 32 (2017).

96. Schmitt, T. M., Ciofani, M., Petrie, H. T. & Zúñiga-Pflücker, J. C. Maintenance of T Cell Specification and Differentiation Requires Recurrent Notch Receptor–Ligand Interactions. *J. Exp. Med.* **200**, 469–479 (2004).

97. Bell, J. J. & Bhandoola, A. The earliest thymic progenitors for T cells possess myeloid lineage potential. *Nature* **452**, 764–767 (2008).

98. King, A. G., Kondo, M., Scherer, D. C. & Weissman, I. L. Lineage infidelity in myeloid cells with TCR gene rearrangement: A latent developmental potential of proT cells revealed by ectopic cytokine receptor signaling. *Proc. Natl. Acad. Sci.* **99**, 4508–4513 (2002).

99. Petrie, H. T. & Kincade, P. W. Many roads, one destination for T cell progenitors. *J. Exp. Med.* **202**, 11–13 (2005).

100. Lavaert, M. *et al.* Integrated scRNA-Seq Identifies Human Postnatal Thymus Seeding Progenitors and Regulatory Dynamics of Differentiating Immature Thymocytes. *Immunity* **52**, 1088-1104.e6 (2020).

101. Hao, Q.-L. *et al.* Human intrathymic lineage commitment is marked by differential CD7 expression: identification of CD7− lympho-myeloid thymic progenitors. *Blood* **111**, 1318–1326 (2008).

102. Wu, L., Antica, M., Johnson, G. R., Scollay, R. & Shortman, K. Developmental potential of the earliest precursor cells from the adult mouse thymus. *J. Exp. Med.* **174**, 1617–1627 (1991).

103. Kondo, M. *et al.* Cell-fate conversion of lymphoid-committed progenitors by instructive actions of cytokines. *Nature* **407**, 383–386 (2000).

104. Rothenberg, E. V. T Cell Lineage Commitment: Identity and Renunciation. *J. Immunol.* **186**, 6649–6655 (2011).

105. Laiosa, C. V., Stadtfeld, M. & Graf, T. Determinants of lymphoid-myeloid lineage diversification. *Annu. Rev. Immunol.* **24**, 705–738 (2006).

106. Rothenberg, E. V., Zhang, J. & Li, L. Multilayered specification of the T-cell lineage fate. *Immunol. Rev.* **238**, 150–168 (2010).

107. Le, J. *et al.* Single-Cell RNA-Seq Mapping of Human Thymopoiesis Reveals Lineage Specification Trajectories and a Commitment Spectrum in T Cell Development. *Immunity* **52**, 1105-1118.e9 (2020).

108. Luc, S. *et al.* The earliest thymic T cell progenitors sustain B cell and myeloid lineage potential. *Nat. Immunol.* **13**, 412–419 (2012).

109. Franco, C. B. *et al.* Notch/Delta signaling constrains reengineering of pro-T cells by PU.1. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 11993–11998 (2006).

110. Reik, W. Stability and flexibility of epigenetic gene regulation in mammalian development. *Nature* **447**, 425–432 (2007).

111. Jackson, M. *et al.* Severe global DNA hypomethylation blocks differentiation and induces histone hyperacetylation in embryonic stem cells. *Mol. Cell. Biol.* **24**, 8862–8871 (2004).

112. Chen, T., Ueda, Y., Dodge, J. E., Wang, Z. & Li, E. Establishment and maintenance of genomic methylation patterns in mouse embryonic stem cells by Dnmt3a and Dnmt3b. *Mol. Cell. Biol.* **23**, 5594–5605 (2003).

113. Suelves, M., Carrió, E., Núñez-Álvarez, Y. & Peinado, M. A. DNA methylation dynamics in cellular commitment and differentiation. *Brief. Funct. Genomics* **15**, 443–453 (2016).

114. Okano, M., Bell, D. W., Haber, D. A. & Li, E. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell* **99**, 247–257 (1999).

115. Bestor, T. H. Activation of mammalian DNA methyltransferase by cleavage of a Zn binding regulatory domain. *EMBO J.* **11**, 2611–2617 (1992).

116. Leonhardt, H., Page, A. W., Weier, H. U. & Bestor, T. H. A targeting sequence directs DNA methyltransferase to sites of DNA replication in mammalian nuclei. *Cell* **71**, 865–873 (1992).

117. Pradhan, S., Bacolla, A., Wells, R. D. & Roberts, R. J. Recombinant Human DNA (Cytosine-5) Methyltransferase: I. EXPRESSION, PURIFICATION, AND COMPARISON OF DE NOVO AND MAINTENANCE METHYLATION *. *J. Biol. Chem.* **274**, 33002–33010 (1999).

118. Yokochi, T. & Robertson, K. D. Preferential methylation of unmethylated DNA by Mammalian de novo DNA methyltransferase Dnmt3a. *J. Biol. Chem.* **277**, 11735–11745 (2002).

119. Okano, M., Xie, S. & Li, E. Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases. *Nat. Genet.* **19**, 219–220 (1998).

120. Arand, J. *et al.* In vivo control of CpG and non-CpG DNA methylation by DNA methyltransferases. *PLoS Genet.* **8**, e1002750 (2012).

121. Haggerty, C. *et al.* Dnmt1 has de novo activity targeted to transposable elements. *Nat. Struct. Mol. Biol.* **28**, 594–603 (2021).

122. Bird, A. P. DNA methylation and the frequency of CpG in animal DNA. *Nucleic Acids Res.* **8**, 1499–1504 (1980).

123. Venter, J. C. *et al.* The sequence of the human genome. *Science* **291**, 1304–1351 (2001).

124. Lander, E. S. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).

125. McClelland, M. & Ivarie, R. Asymmetrical distribution of CpG in an ‘average’ mammalian gene. *Nucleic Acids Res.* **10**, 7865–7877 (1982).

126. Deaton, A. M. & Bird, A. CpG islands and the regulation of transcription. *Genes Dev.* **25**, 1010–1022 (2011).

127. Gardiner-Garden, M. & Frommer, M. CpG islands in vertebrate genomes. *J. Mol. Biol.* **196**, 261–282 (1987).

128. Klose, R. J. & Bird, A. P. Genomic DNA methylation: the mark and its mediators. *Trends Biochem. Sci.* **31**, 89–97 (2006).

129. Mohandas, T., Sparkes, R. S. & Shapiro, L. J. Reactivation of an inactive human X chromosome: evidence for X inactivation by DNA methylation. *Science* **211**, 393–396 (1981).

130. Rainier, S. & Feinberg, A. P. Genomic imprinting, DNA methylation, and cancer. *J. Natl. Cancer Inst.* **86**, 753–759 (1994).

131. Hansen, R. S. & Gartler, S. M. 5-Azacytidine-induced reactivation of the human X chromosome-linked PGK1 gene is associated with a large region of cytosine demethylation in the 5' CpG island. *Proc. Natl. Acad. Sci. U. S. A.* **87**, 4174–4178 (1990).

132. Boyes, J. & Bird, A. DNA methylation inhibits transcription indirectly via a methyl-CpG binding protein. *Cell* **64**, 1123–1134 (1991).

133. Compere, S. J. & Palmiter, R. D. DNA methylation controls the inducibility of the mouse metallothionein-I gene in lymphoid cells. *Cell* **25**, 233–240 (1981).

134. Kass, S. U., Landsberger, N. & Wolffe, A. P. DNA methylation directs a time-dependent repression of transcription initiation. *Curr. Biol. CB* **7**, 157–165 (1997).

135. Siegfried, Z. *et al.* DNA methylation represses transcription in vivo. *Nat. Genet.* **22**, 203–206 (1999).

136. Walsh, C. P., Chaillet, J. R. & Bestor, T. H. Transcription of IAP endogenous retroviruses is constrained by cytosine methylation. *Nat. Genet.* **20**, 116–117 (1998).

137. Panning, B. & Jaenisch, R. DNA hypomethylation can activate Xist expression and silence X-linked genes. *Genes Dev.* **10**, 1991–2002 (1996).

138. Busslinger, M., Hurst, J. & Flavell, R. A. DNA methylation and the regulation of globin gene expression. *Cell* **34**, 197–206 (1983).

139. Greenberg, M. V. C. & Bourc'his, D. The diverse roles of DNA methylation in mammalian development and disease. *Nat. Rev. Mol. Cell Biol.* **20**, 590–607 (2019).

140. Yin, Y. *et al.* Impact of cytosine methylation on DNA binding specificities of human transcription factors. *Science* **356**, eaaj2239 (2017).

141. Dantas Machado, A. C. *et al.* Evolving insights on how cytosine methylation affects protein-DNA binding. *Brief. Funct. Genomics* **14**, 61–73 (2015).

142. Hu, S. *et al.* DNA methylation presents distinct binding sites for human transcription factors. *eLife* **2**, e00726 (2013).

143. Nan, X., Campoy, F. J. & Bird, A. MeCP2 is a transcriptional repressor with abundant binding sites in genomic chromatin. *Cell* **88**, 471–481 (1997).

144. Deplus, R. *et al.* Dnmt3L is a transcriptional repressor that recruits histone deacetylase. *Nucleic Acids Res.* **30**, 3831–3838 (2002).

145. Fuks, F., Burgers, W. A., Brehm, A., Hughes-Davies, L. & Kouzarides, T. DNA methyltransferase Dnmt1 associates with histone deacetylase activity. *Nat. Genet.* **24**, 88–91 (2000).

146. Ooi, S. K. T. *et al.* DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. *Nature* **448**, 714–717 (2007).

147. Weber, M. *et al.* Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. *Nat. Genet.* **39**, 457–466 (2007).

148. Richards, E. J. & Elgin, S. C. R. Epigenetic Codes for Heterochromatin Formation and Silencing: Rounding up the Usual Suspects. *Cell* **108**, 489–500 (2002).

149. Hervouet, E., Peixoto, P., Delage-Mourroux, R., Boyer-Guittaut, M. & Cartron, P.-F. Specific or not specific recruitment of DNMTs for DNA methylation, an epigenetic dilemma. *Clin. Epigenetics* **10**, 17 (2018).

150. Cheng, X. & Blumenthal, R. M. Mammalian DNA Methyltransferases: A Structural Perspective. *Struct. Lond. Engl. 1993* **16**, 341–350 (2008).

151. Rountree, M. R., Bachman, K. E. & Baylin, S. B. DNMT1 binds HDAC2 and a new co-repressor, DMAP1, to form a complex at replication foci. *Nat. Genet.* **25**, 269–277 (2000).

152. Mortusewicz, O., Schermelleh, L., Walter, J., Cardoso, M. C. & Leonhardt, H. Recruitment of DNA methyltransferase I to DNA repair sites. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 8905–8909 (2005).

153. Chuang, L. S. *et al.* Human DNA-(cytosine-5) methyltransferase-PCNA complex as a target for p21WAF1. *Science* **277**, 1996–2000 (1997).

154. Ishiyama, S. *et al.* Structure of the Dnmt1 Reader Module Complexed with a Unique Two-Mono-Ubiquitin Mark on Histone H3 Reveals the Basis for DNA Methylation Maintenance. *Mol. Cell* **68**, 350-360.e7 (2017).

155. Easwaran, H. P., Schermelleh, L., Leonhardt, H. & Cardoso, M. C. Replication-independent chromatin loading of Dnmt1 during G2 and M phases. *EMBO Rep.* **5**, 1181–1186 (2004).

156. Syeda, F. *et al.* The replication focus targeting sequence (RFTS) domain is a DNA-competitive inhibitor of Dnmt1. *J. Biol. Chem.* **286**, 15344–15351 (2011).

157. Takeshita, K. *et al.* Structural insight into maintenance methylation by mouse DNA methyltransferase 1 (Dnmt1). *Proc. Natl. Acad. Sci. U. S. A.* **108**, 9055–9059 (2011).

158. Song, J., Rechkoblit, O., Bestor, T. H. & Patel, D. J. Structure of DNMT1-DNA complex reveals a role for autoinhibition in maintenance DNA methylation. *Science* **331**, 1036–1040 (2011).

159. Yarychkivska, O., Shahabuddin, Z., Comfort, N., Boulard, M. & Bestor, T. H. BAH domains and a histone-like motif in DNA methyltransferase 1 (DNMT1) regulate de novo and maintenance methylation in vivo. *J. Biol. Chem.* **293**, 19466–19475 (2018).

160. Fatemi, M., Hermann, A., Pradhan, S. & Jeltsch, A. The activity of the murine DNA methyltransferase Dnmt1 is controlled by interaction of the catalytic domain with the N-terminal part of the enzyme leading to an allosteric activation of the enzyme after binding to methylated DNA. *J. Mol. Biol.* **309**, 1189–1199 (2001).

161. Jeltsch, A. On the enzymatic properties of Dnmt1: specificity, processivity, mechanism of linear diffusion and allosteric regulation of the enzyme. *Epigenetics* **1**, 63–66 (2006).

162. Berkyurek, A. C. *et al.* The DNA Methyltransferase Dnmt1 Directly Interacts with the SET and RING Finger-associated (SRA) Domain of the Multifunctional Protein Uhrf1 to Facilitate Accession of the Catalytic Center to Hemi-methylated DNA. *J. Biol. Chem.* **289**, 379–386 (2014).

163. Bashtrykov, P., Jankevicius, G., Jurkowska, R. Z., Ragozin, S. & Jeltsch, A. The UHRF1 Protein Stimulates the Activity and Specificity of the Maintenance DNA Methyltransferase DNMT1 by an Allosteric Mechanism. *J. Biol. Chem.* **289**, 4106–4115 (2014).

164. Jurkowska, R. Z., Jurkowski, T. P. & Jeltsch, A. Structure and function of mammalian DNA methyltransferases. *Chembiochem Eur. J. Chem. Biol.* **12**, 206–222 (2011).

165. Guo, X. *et al.* Structural insight into autoinhibition and histone H3-induced activation of DNMT3A. *Nature* **517**, 640–644 (2015).

166. Zhang, Y. *et al.* Chromatin methylation activity of Dnmt3a and Dnmt3a/3L is guided by interaction of the ADD domain with the histone H3 tail. *Nucleic Acids Res.* **38**, 4246–4253 (2010).

167. Li, B.-Z. *et al.* Histone tails regulate DNA methylation by allosterically activating de novo methyltransferase. *Cell Res.* **21**, 1172–1181 (2011).

168. Ren, W., Gao, L. & Song, J. Structural Basis of DNMT1 and DNMT3A-Mediated DNA Methylation. *Genes* **9**, 620 (2018).

169. Ge, Y.-Z. *et al.* Chromatin targeting of de novo DNA methyltransferases by the PWWP domain. *J. Biol. Chem.* **279**, 25447–25454 (2004).

170. Chen, T., Tsujimoto, N. & Li, E. The PWWP domain of Dnmt3a and Dnmt3b is required for directing DNA methylation to the major satellite repeats at pericentric heterochromatin. *Mol. Cell. Biol.* **24**, 9048–9058 (2004).

171. Neri, F. *et al.* Intragenic DNA methylation prevents spurious transcription initiation. *Nature* **543**, 72–77 (2017).

172. Suetake, I. *et al.* Characterization of DNA-binding activity in the N-terminal domain of the DNA methyltransferase Dnmt3a. *Biochem. J.* **437**, 141–148 (2011).

173. Challen, G. A. *et al.* Dnmt3a and Dnmt3b have Overlapping and Distinct Functions in Hematopoietic Stem Cells. *Cell Stem Cell* **15**, 350–364 (2014).

174. Zhang, Z.-M. *et al.* Structural basis for DNMT3A-mediated de novo DNA methylation. *Nature* **554**, 387–391 (2018).

175. Lin, C.-C., Chen, Y.-P., Yang, W.-Z., Shen, J. C. K. & Yuan, H. S. Structural insights into CpG-specific DNA methylation by human DNA methyltransferase 3B. *Nucleic Acids Res.* **48**, 3949–3961 (2020).

176. Gao, L. *et al.* Comprehensive structure-function characterization of DNMT3B and DNMT3A reveals distinctive de novo DNA methylation mechanisms. *Nat. Commun.* **11**, 3355 (2020).

177. Veland, N. *et al.* DNMT3L facilitates DNA methylation partly by maintaining DNMT3A stability in mouse embryonic stem cells. *Nucleic Acids Res.* **47**, 152–167 (2019).

178. Holz-Schietinger, C., Matje, D. M., Harrison, M. F. & Reich, N. O. Oligomerization of DNMT3A Controls the Mechanism of de Novo DNA Methylation. *J. Biol. Chem.* **286**, 41479–41488 (2011).

179. Jia, D., Jurkowska, R. Z., Zhang, X., Jeltsch, A. & Cheng, X. Structure of Dnmt3a bound to Dnmt3L suggests a model for de novo DNA methylation. *Nature* **449**, 248–251 (2007).

180. Klungland, A. & Robertson, A. B. Oxidized C5-methyl cytosine bases in DNA: 5-Hydroxymethylcytosine; 5-formylcytosine; and 5-carboxycytosine. *Free Radic. Biol. Med.* **107**, 62–68 (2017).

181. Ito, S. *et al.* Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science* **333**, 1300–1303 (2011).

182. Hashimoto, H. *et al.* Recognition and potential mechanisms for replication and erasure of cytosine hydroxymethylation. *Nucleic Acids Res.* **40**, 4841–4849 (2012).

183. Ji, D., Lin, K., Song, J. & Wang, Y. Effects of Tet-induced oxidation products of 5-methylcytosine on Dnmt1- and DNMT3a-mediated cytosine methylation. *Mol. Biosyst.* **10**, 1749–1752 (2014).

184. Valinluck, V. & Sowers, L. C. Endogenous cytosine damage products alter the site selectivity of human DNA maintenance methyltransferase DNMT1. *Cancer Res.* **67**, 946–950 (2007).

185. Kohli, R. M. & Zhang, Y. TET enzymes, TDG and the dynamics of DNA demethylation. *Nature* **502**, 472–479 (2013).

186. He, Y.-F. *et al.* Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science* **333**, 1303–1307 (2011).

187. Maiti, A. & Drohat, A. C. Thymine DNA glycosylase can rapidly excise 5-formylcytosine and 5-carboxylcytosine: potential implications for active demethylation of CpG sites. *J. Biol. Chem.* **286**, 35334–35338 (2011).

188. Moore, L. D., Le, T. & Fan, G. DNA Methylation and Its Basic Function. *Neuropsychopharmacology* **38**, 23–38 (2013).

189. Monk, M., Adams, R. L. & Rinaldi, A. Decrease in DNA methylase activity during preimplantation development in the mouse. *Dev. Camb. Engl.* **112**, 189–192 (1991).

190. Bhutani, N., Burns, D. M. & Blau, H. M. DNA Demethylation Dynamics. *Cell* **146**, 866–872 (2011).

191. Farlik, M. *et al.* DNA Methylation Dynamics of Human Hematopoietic Stem Cell Differentiation. *Cell Stem Cell* **19**, 808–822 (2016).

192. Ji, H. *et al.* A comprehensive methylome map of lineage commitment from hematopoietic progenitors. *Nature* **467**, 338–342 (2010).

193. Accomando, W. P., Wiencke, J. K., Houseman, E. A., Nelson, H. H. & Kelsey, K. T. Quantitative reconstruction of leukocyte subsets using DNA methylation. *Genome Biol.* **15**, R50 (2014).

194. Challen, G. A. *et al.* Dnmt3a is essential for hematopoietic stem cell differentiation. *Nat. Genet.* **44**, 23–31 (2011).

195. Kramer, A. C. *et al.* Dnmt3a Regulates T-cell Development and Suppresses T-ALL Transformation. *Leukemia* **31**, 2479–2490 (2017).

196. Rodriguez, R. M. *et al.* Regulation of the transcriptional program by DNA methylation during human αβ T-cell development. *Nucleic Acids Res.* **43**, 760–774 (2015).

197. Cancer Genome Atlas Research Network *et al.* Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. *N. Engl. J. Med.* **368**, 2059–2074 (2013).

198. Yang, L., Rau, R. & Goodell, M. A. DNMT3A in haematological malignancies. *Nat. Rev. Cancer* **15**, 152–165 (2015).

199. Raetz, E. A. & Teachey, D. T. T-cell acute lymphoblastic leukemia. *Hematol. Am. Soc. Hematol. Educ. Program* **2016**, 580–588 (2016).

200. Pui, C.-H. & Evans, W. E. Treatment of Acute Lymphoblastic Leukemia. *N. Engl. J. Med.* **354**, 166–178 (2006).

201. Uckun, F. M. *et al.* Biology and Treatment of Childhood T-Lineage Acute Lymphoblastic Leukemia. *Blood* **91**, 735–746 (1998).

202. Faderl, S., Kantarjian, H. M., Talpaz, M. & Estrov, Z. Clinical Significance of Cytogenetic Abnormalities in Adult Acute Lymphoblastic Leukemia. *Blood* **91**, 3995–4019 (1998).

203. Bertin, R. *et al.* CDKN2A, CDKN2B, and MTAP gene dosage permits precise characterization of mono- and bi-allelic 9p21 deletions in childhood acute lymphoblastic leukemia. *Genes. Chromosomes Cancer* **37**, 44–57 (2003).

204. Mullighan, C. G. *et al.* Genome-wide analysis of genetic alterations in acute lymphoblastic leukaemia. *Nature* **446**, 758–764 (2007).

205. Sulong, S. *et al.* A comprehensive analysis of the CDKN2A gene in childhood acute lymphoblastic leukemia reveals genomic deletion, copy number neutral loss of heterozygosity, and association with specific cytogenetic subgroups. *Blood* **113**, 100–107 (2009).

206. Ferrando, A. A. *et al.* Gene expression signatures define novel oncogenic pathways in T cell acute lymphoblastic leukemia. *Cancer Cell* **1**, 75–87 (2002).

207. Hebert, J., Cayuela, J. M., Berkeley, J. & Sigaux, F. 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 (1994).

208. Raimondi, S. C. *et al.* Cytogenetics of childhood T-cell leukemia. *Blood* **72**, 1560–1566 (1988).

209. Cauwelier, B. *et al.* Molecular cytogenetic study of 126 unselected T-ALL cases reveals high incidence of TCRβ locus rearrangements and putative new T-cell oncogenes. *Leukemia* **20**, 1238–1244 (2006).

210. Vlierberghe, P. V. & Ferrando, A. The molecular basis of T cell acute lymphoblastic leukemia. *J. Clin. Invest.* **122**, 3398–3406 (2012).

211. Tosello, V. & Ferrando, A. A. The NOTCH signaling pathway: role in the pathogenesis of T-cell acute lymphoblastic leukemia and implication for therapy. *Ther. Adv. Hematol.* **4**, 199–210 (2013).

212. Weng, A. P. *et al.* Activating mutations of NOTCH1 in human T cell acute lymphoblastic leukemia. *Science* **306**, 269–271 (2004).

213. Ellisen, L. W. *et al.* TAN-1, the human homolog of the Drosophila notch gene, is broken by chromosomal translocations in T lymphoblastic neoplasms. *Cell* **66**, 649–661 (1991).

214. O'Neil, J. *et al.* FBW7 mutations in leukemic cells mediate NOTCH pathway activation and resistance to gamma-secretase inhibitors. *J. Exp. Med.* **204**, 1813–1824 (2007).

215. Thompson, B. J. *et al.* The SCFFBW7 ubiquitin ligase complex as a tumor suppressor in T cell leukemia. *J. Exp. Med.* **204**, 1825–1835 (2007).

216. Ferrando, A. A. The role of NOTCH1 signaling in T-ALL. *Hematol. Educ. Program Am. Soc. Hematol. Am. Soc. Hematol. Educ. Program* 353–361 (2009) doi:10.1182/asheducation-2009.1.353.

217. Inaba, H., Greaves, M. & Mullighan, C. G. Acute lymphoblastic leukaemia. *The Lancet* **381**, 1943–1955 (2013).

218. Bene, M. C. *et al.* Proposals for the immunological classification of acute leukemias. European Group for the Immunological Characterization of Leukemias (EGIL). *Leukemia* **9**, 1783–1786 (1995).

219. Homminga, I. *et al.* Integrated transcript and genome analyses reveal NKX2-1 and MEF2C as potential oncogenes in T cell acute lymphoblastic leukemia. *Cancer Cell* **19**, 484–497 (2011).

220. Van Vlierberghe, P. *et al.* The recurrent SET-NUP214 fusion as a new HOXA activation mechanism in pediatric T-cell acute lymphoblastic leukemia. *Blood* **111**, 4668–4680 (2008).

221. Soulier, J. *et al.* HOXA genes are included in genetic and biologic networks defining human acute T-cell leukemia (T-ALL). *Blood* **106**, 274–286 (2005).

222. Chen, B. *et al.* 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 (2018).

223. Liu, Y. *et al.* The genomic landscape of pediatric and young adult T-lineage acute lymphoblastic leukemia. *Nat. Genet.* **49**, 1211–1218 (2017).

224. Zhang, J. *et al.* The genetic basis of early T-cell precursor acute lymphoblastic leukaemia. *Nature* **481**, 157–163 (2012).

225. Girardi, T., Vicente, C., Cools, J. & De Keersmaecker, K. The genetics and molecular biology of T-ALL. *Blood* **129**, 1113–1123 (2017).

226. Coustan-Smith, E. *et al.* Early T-cell precursor leukaemia: a subtype of very high-risk acute lymphoblastic leukaemia. *Lancet Oncol.* **10**, 147–156 (2009).

227. Niehues, T. *et al.* A classification based on T cell selection-related phenotypes identifies a subgroup of childhood T-ALL with favorable outcome in the COALL studies. *Leukemia* **13**, 614–617 (1999).

228. Pui, C. H., Behm, F. G. & Crist, W. M. Clinical and biologic relevance of immunologic marker studies in childhood acute lymphoblastic leukemia. *Blood* **82**, 343–362 (1993).

229. Arber, D. A. *et al.* The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia. *Blood* **127**, 2391–2405 (2016).

230. Inukai, T. *et al.* Clinical significance of early T-cell precursor acute lymphoblastic leukaemia: results of the Tokyo Children's Cancer Study Group Study L99-15. *Br. J. Haematol.* **156**, 358–365 (2012).

231. Zuurbier, L. *et al.* Immature MEF2C-dysregulated T-cell leukemia patients have an early T-cell precursor acute lymphoblastic leukemia gene signature and typically have non-rearranged T-cell receptors. *Haematologica* **99**, 94–102 (2014).

232. Khogeer, H. *et al.* Early T precursor acute lymphoblastic leukaemia/lymphoma shows differential immunophenotypic characteristics including frequent CD33 expression and in vitro response to targeted CD33 therapy. *Br. J. Haematol.* **186**, 538–548 (2019).

233. Patrick, K. *et al.* Outcome for children and young people with Early T-cell precursor acute lymphoblastic leukaemia treated on a contemporary protocol, UKALL 2003. *Br. J. Haematol.* **166**, 421–424 (2014).

234. Wood, B. L. *et al.* T-Lymphoblastic Leukemia (T-ALL) Shows Excellent Outcome, Lack of Significance of the Early Thymic Precursor (ETP) Immunophenotype, and Validation of the Prognostic Value of End-Induction Minimal Residual Disease (MRD) in Children's Oncology Group (COG) Study AALL0434. *Blood* **124**, 1 (2014).

235. Willemse, M. J. *et al.* Detection of minimal residual disease identifies differences in treatment response between T-ALL and precursor B-ALL. *Blood* **99**, 4386–4393 (2002).

236. Szczepański, T., Flohr, T., van der Velden, V. H. J., Bartram, C. R. & van Dongen, J. J. M. Molecular monitoring of residual disease using antigen receptor genes in childhood acute lymphoblastic leukaemia. *Best Pract. Res. Clin. Haematol.* **15**, 37–57 (2002).

237. Seki, M. *et al.* Recurrent SPI1 (PU.1) fusions in high-risk pediatric T cell acute lymphoblastic leukemia. *Nat. Genet.* **49**, 1274–1281 (2017).

238. Belver, L. & Ferrando, A. The genetics and mechanisms of T cell acute lymphoblastic leukaemia. *Nat. Rev. Cancer* **16**, 494–507 (2016).

239. Meijerink, J. P. P. Genetic rearrangements in relation to immunophenotype and outcome in T-cell acute lymphoblastic leukaemia. *Best Pract. Res. Clin. Haematol.* **23**, 307–318 (2010).

240. Van der Meulen, J., Van Roy, N., Van Vlierberghe, P. & Speleman, F. The epigenetic landscape of T-cell acute lymphoblastic leukemia. *Int. J. Biochem. Cell Biol.* **53**, 547–557 (2014).

241. Peirs, S. *et al.* Epigenetics in T-cell acute lymphoblastic leukemia. *Immunol. Rev.* **263**, 50–67 (2015).

242. Huether, R. *et al.* The landscape of somatic mutations in epigenetic regulators across 1,000 paediatric cancer genomes. *Nat. Commun.* **5**, 3630 (2014).

243. Van Vlierberghe, P. *et al.* ETV6 mutations in early immature human T cell leukemias. *J. Exp. Med.* **208**, 2571–2579 (2011).

244. Grossmann, V. *et al.* The molecular profile of adult T-cell acute lymphoblastic leukemia: mutations in RUNX1 and DNMT3A are associated with poor prognosis in T-ALL. *Genes. Chromosomes Cancer* **52**, 410–422 (2013).

245. Neumann, M. *et al.* Whole-exome sequencing in adult ETP-ALL reveals a high rate of DNMT3A mutations. *Blood* **121**, 4749–4752 (2013).

246. Ntziachristos, P. *et al.* Genetic inactivation of the polycomb repressive complex 2 in T cell acute lymphoblastic leukemia. *Nat. Med.* **18**, 298–301 (2012).

247. Van Vlierberghe, P. *et al.* PHF6 mutations in T-cell acute lymphoblastic leukemia. *Nat. Genet.* **42**, 338–342 (2010).

248. Van Vlierberghe, P. *et al.* Prognostic relevance of integrated genetic profiling in adult T-cell acute lymphoblastic leukemia. *Blood* **122**, 74–82 (2013).

249. Simon, C. *et al.* A key role for EZH2 and associated genes in mouse and human adult T-cell acute leukemia. *Genes Dev.* **26**, 651–656 (2012).

250. Danis, E. *et al.* Ezh2 Controls an Early Hematopoietic Program and Growth and Survival Signaling in Early T Cell Precursor Acute Lymphoblastic Leukemia. *Cell Rep.* **14**, 1953–1965 (2016).

251. Margueron, R. & Reinberg, D. The Polycomb complex PRC2 and its mark in life. *Nature* **469**, 343–349 (2011).

252. Van der Meulen, J. *et al.* The H3K27me3 demethylase UTX is a gender-specific tumor suppressor in T-cell acute lymphoblastic leukemia. *Blood* **125**, 13–21 (2015).

253. Ntziachristos, P. *et al.* Contrasting roles of histone 3 lysine 27 demethylases in acute lymphoblastic leukaemia. *Nature* **514**, 513–517 (2014).

254. Wendorff, A. A. *et al.* Phf6 loss enhances HSC self-renewal driving tumor initiation and leukemia stem cell activity in T-ALL. *Cancer Discov.* **9**, 436–451 (2019).

255. Shan, H. *et al.* USP7 deubiquitinates and stabilizes NOTCH1 in T-cell acute lymphoblastic leukemia. *Signal Transduct. Target. Ther.* **3**, 1–10 (2018).

256. Jin, Q. *et al.* USP7 cooperates with NOTCH1 to drive the oncogenic transcriptional program in T cell leukemia. *Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res.* **25**, 222–239 (2019).

257. Bond, J. *et al.* DNMT3A mutation is associated with increased age and adverse outcome in adult T-cell acute lymphoblastic leukemia. *Haematologica* **104**, 1617–1625 (2019).

258. Jang, W. *et al.* CDKN2B downregulation and other genetic characteristics in T-acute lymphoblastic leukemia. *Exp. Mol. Med.* **51**, 1–15 (2019).

259. De Keersmaecker, K. *et al.* Exome sequencing identifies mutation in CNOT3 and ribosomal genes RPL5 and RPL10 in T-cell acute lymphoblastic leukemia. *Nat. Genet.* **45**, 186–190 (2013).

260. Neumann, M. *et al.* Mutational spectrum of adult T-ALL. *Oncotarget* **6**, 2754–2766 (2014).

261. Simonin, M. *et al.* Oncogenetic landscape and clinical impact of IDH1 and IDH2 mutations in T-ALL. *J. Hematol. Oncol.J Hematol Oncol* **14**, 74 (2021).

262. Ward, P. S. *et al.* The common feature of leukemia-associated IDH1 and IDH2 mutations is a neomorphic enzyme activity converting alpha-ketoglutarate to 2-hydroxyglutarate. *Cancer Cell* **17**, 225–234 (2010).

263. Roller, A. *et al.* Landmark analysis of DNMT3A mutations in hematological malignancies. *Leukemia* **27**, 1573–1578 (2013).

264. Ley, T. J. *et al.* DNMT3A mutations in acute myeloid leukemia. *N. Engl. J. Med.* **363**, 2424–2433 (2010).

265. Stegelmann, F. *et al.* DNMT3A mutations in myeloproliferative neoplasms. *Leukemia* **25**, 1217–1219 (2011).

266. Gaidzik, V. I. *et al.* Clinical impact of DNMT3A mutations in younger adult patients with acute myeloid leukemia: results of the AML Study Group (AMLSG). *Blood* **121**, 4769–4777 (2013).

267. Huang, Y.-H. *et al.* Systematic Profiling of DNMT3A Variants Reveals Protein Instability Mediated by the DCAF8 E3 Ubiquitin Ligase Adaptor. *Cancer Discov.* **12**, 220–235 (2022).

268. Mayle, A. *et al.* Dnmt3a loss predisposes murine hematopoietic stem cells to malignant transformation. *Blood* **125**, 629–638 (2015).

269. Yang, L. *et al.* DNMT3A Loss Drives Enhancer Hypomethylation in FLT3-ITD-Associated Leukemias. *Cancer Cell* **29**, 922–934 (2016).

270. Kiyoi, H. *et al.* Internal tandem duplication of the FLT3 gene is a novel modality of elongation mutation which causes constitutive activation of the product. *Leukemia* **12**, 1333–1337 (1998).

271. Griffith, J. *et al.* The structural basis for autoinhibition of FLT3 by the juxtamembrane domain. *Mol. Cell* **13**, 169–178 (2004).

272. Nakao, M. *et al.* Internal tandem duplication of the flt3 gene found in acute myeloid leukemia. *Leukemia* **10**, 1911–1918 (1996).

273. Neumann, M. *et al.* FLT3 Mutations in Early T-Cell Precursor ALL Characterize a Stem Cell Like Leukemia and Imply the Clinical Use of Tyrosine Kinase Inhibitors. *PLoS ONE* **8**, e53190 (2013).

274. Cooper, S. L. & Brown, P. A. Treatment of Pediatric Acute Lymphoblastic Leukemia. *Pediatr. Clin. North Am.* **62**, 61–73 (2015).

275. Roti, G. & Stegmaier, K. New Approaches to Target T-ALL. *Front. Oncol.* **4**, 170 (2014).

276. Deangelo, D. J. *et al.* A phase I clinical trial of the notch inhibitor MK-0752 in patients with T-cell acute lymphoblastic leukemia/lymphoma (T-ALL) and other leukemias. *J. Clin. Oncol.* **24**, 6585–6585 (2006).

277. Takebe, N., Nguyen, D. & Yang, S. X. Targeting Notch signaling pathway in cancer: Clinical development advances and challenges. *Pharmacol. Ther.* **141**, 140–149 (2014).

278. Samon, J. B. *et al.* Preclinical Analysis of the γ-Secretase Inhibitor PF-03084014 in Combination with Glucocorticoids in T-cell Acute Lymphoblastic Leukemia. *Mol. Cancer Ther.* **11**, 1565–1575 (2012).

279. Agnusdei, V. *et al.* Therapeutic antibody targeting of Notch1 in T-acute lymphoblastic leukemia xenografts. *Leukemia* **28**, 278–288 (2014).

280. Cordo', V., van der Zwet, J. C. G., Canté-Barrett, K., Pieters, R. & Meijerink, J. P. P. T-cell Acute Lymphoblastic Leukemia: A Roadmap to Targeted Therapies. *Blood Cancer Discov.* **2**, 19–31 (2021).

281. Chilosi, M. *et al.* Differential expression of cyclin-dependent kinase 6 in cortical thymocytes and T-cell lymphoblastic lymphoma/leukemia. *Am. J. Pathol.* **152**, 209–217 (1998).

282. Sawai, C. M. *et al.* Therapeutic Targeting of the Cyclin D3:CDK4/6 Complex in T Cell Leukemia. *Cancer Cell* **22**, 452–465 (2012).

283. Pikman, Y. *et al.* Synergistic Drug Combinations with a CDK4/6 Inhibitor in T-cell Acute Lymphoblastic Leukemia. *Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res.* **23**, 1012–1024 (2017).

284. Knudsen, E. S. & Witkiewicz, A. K. The Strange Case of CDK4/6 Inhibitors: Mechanisms, Resistance, and Combination Strategies. *Trends Cancer* **3**, 39–55 (2017).

285. Asghar, U., Witkiewicz, A. K., Turner, N. C. & Knudsen, E. S. The history and future of targeting cyclin-dependent kinases in cancer therapy. *Nat. Rev. Drug Discov.* **14**, 130–146 (2015).

286. Casero, D. *et al.* Long non-coding RNA profiling of human lymphoid progenitor cells reveals transcriptional divergence of B cell and T cell lineages. *Nat. Immunol.* **16**, 1282–1291 (2015).

287. Wood, B. *et al.* Patients with Early T-Cell Precursor (ETP) Acute Lymphoblastic Leukemia (ALL) Have High Levels of Minimal Residual Disease (MRD) at the End of induction—A Children's Oncology Group (COG) Study. *Blood* **114**, 9 (2009).

288. Gutierrez, A. *et al.* Absence of biallelic TCRgamma deletion predicts early treatment failure in pediatric T-cell acute lymphoblastic leukemia. *J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol.* **28**, 3816–3823 (2010).

289. Lee, M. S., Hanspers, K., Barker, C. S., Korn, A. P. & McCune, J. M. Gene expression profiles during human CD4+ T cell differentiation. *Int. Immunol.* **16**, 1109–1124 (2004).

290. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinforma. Oxf. Engl.* **30**, 2114–2120 (2014).

291. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).

292. Liao, Y., Smyth, G. K. & Shi, W. The R package Rsubread is easier, faster, cheaper and better for alignment and quantification of RNA sequencing reads. *Nucleic Acids Res.* **47**, e47 (2019).

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

294. Kusakabe, M. *et al.* Synthetic modeling reveals HOXB genes are critical for the initiation and maintenance of human leukemia. *Nat. Commun.* **10**, 2913 (2019).

295. Bolotin, D. A. *et al.* MiXCR: software for comprehensive adaptive immunity profiling. *Nat. Methods* **12**, 380–381 (2015).

296. Hänzelmann, S., Castelo, R. & Guinney, J. GSVA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinformatics* **14**, 7 (2013).

297. Dadi, S. *et al.* 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 (2012).

298. Petrie, K., Zelent, A. & Waxman, S. Differentiation therapy of acute myeloid leukemia: past, present and future. *Curr. Opin. Hematol.* **16**, 84–91 (2009).

299. Parekh, C. & Crooks, G. M. Critical Differences in Hematopoiesis and Lymphoid Development Between Humans and Mice. *J. Clin. Immunol.* **33**, 711–715 (2013).

300. Plum, J., De Smedt, M., Defresne, M. P., Leclercq, G. & Vandekerckhove, B. Human CD34+ fetal liver stem cells differentiate to T cells in a mouse thymic microenvironment. *Blood* **84**, 1587–1593 (1994).

301. Schmitt, T. M. & Zúñiga-Pflücker, J. C. Induction of T Cell Development from Hematopoietic Progenitor Cells by Delta-like-1 In Vitro. *Immunity* **17**, 749–756 (2002).

302. De Smedt, M., Hoebeke, I. & Plum, J. Human bone marrow CD34+ progenitor cells mature to T cells on OP9-DL1 stromal cell line without thymus microenvironment. *Blood Cells. Mol. Dis.* **33**, 227–232 (2004).

303. Lagergren, A. *et al.* The Cxcl12, periostin, and Ccl9 genes are direct targets for early B-cell factor in OP-9 stroma cells. *J. Biol. Chem.* **282**, 14454–14462 (2007).

304. Cho, S. K. *et al.* Functional characterization of B lymphocytes generated in vitro from embryonic stem cells. *Proc. Natl. Acad. Sci. U. S. A.* **96**, 9797–9802 (1999).

305. Awong, G., Herer, E., La Motte-Mohs, R. N. & Zúñiga-Pflücker, J. C. Human CD8 T cells generated in vitro from hematopoietic stem cells are functionally mature. *BMC Immunol.* **12**, 22 (2011).

306. Mohtashami, M. *et al.* Direct Comparison of Dll1- and Dll4-Mediated Notch Activation Levels Shows Differential Lymphomyeloid Lineage Commitment Outcomes. *J. Immunol.* **185**, 867–876 (2010).

307. de Pooter, R. & Zúñiga-Pflücker, J. C. T-cell potential and development in vitro: the OP9-DL1 approach. *Curr. Opin. Immunol.* **19**, 163–168 (2007).

308. Dervović, D. & Zúñiga-Pflücker, J. C. Positive selection of T cells, an in vitro view. *Semin. Immunol.* **22**, 276–286 (2010).

309. Mohtashami, M., Shukla, S., Zandstra, P. & Zúñiga-Pflücker, J. C. Artificial Thymus: Recreating Microenvironmental Cues to Direct T Cell Differentiation and Thymic Regeneration. in *Synthetic Immunology* (eds. Watanabe, T. & Takahama, Y.) 95–120 (Springer Japan, 2016). doi:10.1007/978-4-431-56027-2_4.

310. Poznansky, M. C. *et al.* Efficient generation of human T cells from a tissue-engineered thymic organoid. *Nat. Biotechnol.* **18**, 729–734 (2000).

311. Montel-Hagen, A. *et al.* Organoid-induced differentiation of conventional T cells from human pluripotent stem cells. *Cell Stem Cell* **24**, 376-389.e8 (2019).

312. Seet, C. S. *et al.* Generation of mature T cells from human hematopoietic stem and progenitor cells in artificial thymic organoids. *Nat. Methods* **14**, 521–530 (2017).

313. Chung, B. *et al.* Engineering the human thymic microenvironment to support thymopoiesis in vivo. *Stem Cells Dayt. Ohio* **32**, 2386–2396 (2014).

314. Sun, V. *et al.* The Metabolic Landscape of Thymic T Cell Development In Vivo and In Vitro. *Front. Immunol.* **12**, (2021).

315. Reimann, C. *et al.* Human T-lymphoid progenitors generated in a feeder-cell-free Delta-like-4 culture system promote T-cell reconstitution in NOD/SCID/γc(-/-) mice. *Stem Cells Dayt. Ohio* **30**, 1771–1780 (2012).

316. Iriguchi, S. *et al.* A clinically applicable and scalable method to regenerate T-cells from iPSCs for off-the-shelf T-cell immunotherapy. *Nat. Commun.* **12**, 430 (2021).

317. Shukla, S. *et al.* Progenitor T-cell differentiation from hematopoietic stem cells using Delta-like-4 and VCAM-1. *Nat. Methods* **14**, 531–538 (2017).

318. Trapnell, C. *et al.* The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat. Biotechnol.* **32**, 381–386 (2014).

319. Hahne, F. *et al.* flowCore: a Bioconductor package for high throughput flow cytometry. *BMC Bioinformatics* **10**, 106 (2009).

320. Si, Y., Liu, P., Li, P. & Brutnell, T. P. Model-based clustering for RNA-seq data. *Bioinforma. Oxf. Engl.* **30**, 197–205 (2014).

321. Wong, R. *et al.* Improved resolution of phenotypic subsets in human T-ALL by incorporation of RNA-seq based developmental profiling. *Leuk. Res.* **110**, 106712 (2021).

322. Merkel, A. *et al.* gemBS: high throughput processing for DNA methylation data from bisulfite sequencing. *Bioinforma. Oxf. Engl.* **35**, 737–742 (2019).

323. Boyd, N. *et al.* 'Off-the-Shelf' Immunotherapy: Manufacture of CD8+ T Cells Derived from Hematopoietic Stem Cells. *Cells* **10**, 2631 (2021).

324. Rothenberg, E. V. Single-cell insights into the hematopoietic generation of T-lymphocyte precursors in mouse and human. *Exp. Hematol.* **95**, 1–12 (2021).

325. Wada, H. *et al.* Adult T-cell progenitors retain myeloid potential. *Nature* **452**, 768–772 (2008).

326. Li, L., Leid, M. & Rothenberg, E. V. An Early T Cell Lineage Commitment Checkpoint Dependent on the Transcription Factor Bcl11b. *Science* **329**, 89–93 (2010).

327. Rothenberg, E. V. Logic and lineage impacts on functional transcription factor deployment for T-cell fate commitment. *Biophys. J.* **120**, 4162–4181 (2021).

328. Sarbassov, D. D., Guertin, D. A., Ali, S. M. & Sabatini, D. M. Phosphorylation and regulation of Akt/PKB by the rictor-mTOR complex. *Science* **307**, 1098–1101 (2005).

329. Zufferey, R., Nagy, D., Mandel, R. J., Naldini, L. & Trono, D. Multiply attenuated lentiviral vector achieves efficient gene delivery in vivo. *Nat. Biotechnol.* **15**, 871–875 (1997).

330. Cong, L. *et al.* Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819–823 (2013).

331. Concordet, J.-P. & Haeussler, M. CRISPOR: intuitive guide selection for CRISPR/Cas9 genome editing experiments and screens. *Nucleic Acids Res.* **46**, W242–W245 (2018).

332. Conant, D. *et al.* Inference of CRISPR Edits from Sanger Trace Data. *CRISPR J.* **5**, 123–130 (2022).

333. Taussig, D. C. *et al.* Hematopoietic stem cells express multiple myeloid markers: implications for the origin and targeted therapy of acute myeloid leukemia. *Blood* **106**, 4086–4092 (2005).

334. Burgess, A. W. & Metcalf, D. The nature and action of granulocyte-macrophage colony stimulating factors. *Blood* **56**, 947–958 (1980).

335. Dong, F. *et al.* Stimulation of Stat5 by granulocyte colony-stimulating factor (G-CSF) is modulated by two distinct cytoplasmic regions of the G-CSF receptor. *J. Immunol. Baltim. Md 1950* **161**, 6503–6509 (1998).

336. Tian, S. S. *et al.* Multiple signaling pathways induced by granulocyte colony-stimulating factor involving activation of JAKs, STAT5, and/or STAT3 are required for regulation of three distinct classes of immediate early genes. *Blood* **88**, 4435–4444 (1996).

337. Mui, A. L., Wakao, H., O'Farrell, A. M., Harada, N. & Miyajima, A. Interleukin-3, granulocyte-macrophage colony stimulating factor and interleukin-5 transduce signals through two STAT5 homologs. *EMBO J.* **14**, 1166–1175 (1995).

338. Suzuki, K., Hino, M., Hato, F., Tatsumi, N. & Kitagawa, S. Cytokine-Specific Activation of Distinct Mitogen-Activated Protein Kinase Subtype Cascades in Human Neutrophils Stimulated by Granulocyte Colony-Stimulating Factor, Granulocyte-Macrophage Colony-Stimulating Factor, and Tumor Necrosis Factor-□. *Blood* **93**, 341–349 (1999).

339. Hohaus, S. *et al.* PU.1 (Spi-1) and C/EBP alpha regulate expression of the granulocyte-macrophage colony-stimulating factor receptor alpha gene. *Mol. Cell. Biol.* **15**, 5830–5845 (1995).

340. Smith, L. T., Hohaus, S., Gonzalez, D. A., Dziennis, S. E. & Tenen, D. G. PU.1 (Spi-1) and C/EBP alpha regulate the granulocyte colony-stimulating factor receptor promoter in myeloid cells. *Blood* **88**, 1234–1247 (1996).

341. Zhang, D. E., Hetherington, C. J., Chen, H. M. & Tenen, D. G. The macrophage transcription factor PU.1 directs tissue-specific expression of the macrophage colony-stimulating factor receptor. *Mol. Cell. Biol.* **14**, 373–381 (1994).

342. Shibata, Y. *et al.* GM-CSF Regulates Alveolar Macrophage Differentiation and Innate Immunity in the Lung through PU.1. *Immunity* **15**, 557–567 (2001).

343. Mossadegh-Keller, N. *et al.* M-CSF instructs myeloid lineage fate in single haematopoietic stem cells. *Nature* **497**, 239–243 (2013).

344. Choe, K. S., Ujhelly, O., Wontakal, S. N. & Skoultchi, A. I. PU.1 directly regulates cdk6 gene expression, linking the cell proliferation and differentiation programs in erythroid cells. *J. Biol. Chem.* **285**, 3044–3052 (2010).

345. Celada, A. *et al.* The transcription factor PU.1 is involved in macrophage proliferation. *J. Exp. Med.* **184**, 61–69 (1996).

346. Zhu, H. *et al.* T-ALL leukemia stem cell 'stemness' is epigenetically controlled by the master regulator SPI1. *eLife* **7**, e38314 (2018).

347. Oda, M. *et al.* DNA Methylation Restricts Lineage-specific Functions of Transcription Factor Gata4 during Embryonic Stem Cell Differentiation. *PLOS Genet.* **9**, e1003574 (2013).

348. Li, J.-Y. *et al.* Synergistic function of DNA methyltransferases Dnmt3a and Dnmt3b in the methylation of Oct4 and Nanog. *Mol. Cell. Biol.* **27**, 8748–8759 (2007).

349. Farthing, C. R. *et al.* Global Mapping of DNA Methylation in Mouse Promoters Reveals Epigenetic Reprogramming of Pluripotency Genes. *PLOS Genet.* **4**, e1000116 (2008).

350. Venugopal, K., Feng, Y., Shabashvili, D. & Guryanova, O. A. Alterations to DNMT3A in Hematologic Malignancies. *Cancer Res.* **81**, 254–263 (2021).

351. Jaiswal, S. *et al.* Age-related clonal hematopoiesis associated with adverse outcomes. *N. Engl. J. Med.* **371**, 2488–2498 (2014).

352. Xie, M. *et al.* Age-related mutations associated with clonal hematopoietic expansion and malignancies. *Nat. Med.* **20**, 1472–1478 (2014).

353. Genovese, G. *et al.* Clonal hematopoiesis and blood-cancer risk inferred from blood DNA sequence. *N. Engl. J. Med.* **371**, 2477–2487 (2014).

354. Young, A. L., Challen, G. A., Birmann, B. M. & Druley, T. E. Clonal haematopoiesis harbouring AML-associated mutations is ubiquitous in healthy adults. *Nat. Commun.* **7**, 12484 (2016).

355. Shlush, L. I. *et al.* Identification of pre-leukaemic haematopoietic stem cells in acute leukaemia. *Nature* **506**, 328–333 (2014).

356. Jeong, M. *et al.* Loss of Dnmt3a Immortalizes Hematopoietic Stem Cells In Vivo. *Cell Rep.* **23**, 1–10 (2018).

357. Waks, A. G. & Winer, E. P. Breast Cancer Treatment: A Review. *JAMA* **321**, 288–300 (2019).

358. Park, J.-E. *et al.* A cell atlas of human thymic development defines T cell repertoire formation. *Science* **367**, eaay3224 (2020).

359. Anderson, M. K., Weiss, A. H., Hernandez-Hoyos, G., Dionne, C. J. & Rothenberg, E. V. Constitutive expression of PU.1 in fetal hematopoietic progenitors blocks T cell development at the pro-T cell stage. *Immunity* **16**, 285–296 (2002).

360. Laiosa, C. V., Stadtfeld, M., Xie, H., Andres-Aguayo, L. de & Graf, T. Reprogramming of Committed T Cell Progenitors to Macrophages and Dendritic Cells by C/EBPα and PU.1 Transcription Factors. *Immunity* **25**, 731–744 (2006).

361. Franzke, A. *et al.* G-CSF as immune regulator in T cells expressing the G-CSF receptor: implications for transplantation and autoimmune diseases. *Blood* **102**, 734–739 (2003).

362. De Smedt, M. *et al.* T-lymphoid differentiation potential measured in vitro is higher in CD34+CD38−/lo hematopoietic stem cells from umbilical cord blood than from bone marrow and is an intrinsic property of the cells. *Haematologica* **96**, 646–654 (2011).

363. Drexler, H. G. *The Leukemia-Lymphoma Cell Line FactsBook*. (Academic Press, 2001).

Appendices

Appendix A Cord blood pools accessed

Cord pool	Experiments performed
C9801p	Flow immunophenotyping
C9663p	Flow immunophenotyping
C9839p	Flow immunophenotyping, media transfer experiments, RNA-seq, WGBS
C9891p	Flow immunophenotyping, media transfer experiments, RNA-seq, WGBS
C9941p	Flow immunophenotyping, media transfer experiments, RNA-seq, WGBS

Appendix B Flow cytometry antibodies

B.1 Antibodies for discrimination of T-lymphoid and myeloid populations

Antibody	Clone	Colour/Channel	Manufacturer
CD11b	ICRF44	Per-CPCy5.5	BioLegend
CD13	WM15	PE	BioLegend
CD15	W6D3	BV421	BD Biosciences
CD1a	HI149	PE-Cy7	BioLegend
CD3	UCHT-1	BV480	BD Biosciences
CD33	WM53	PE-CF594	BD Biosciences
CD34	581	BUV563	BD Biosciences
CD38	HB7	BUV395	BD Biosciences
CD4	SK3	BUV661	BD Biosciences
CD44	IM7	BV786	BD Biosciences
CD5	UCHT2	BUV737	BD Biosciences
CD7	M-T701	APC-H7	BD Biosciences
CD8	RPA-T8	APC	BD Biosciences
HLA-DR	HIT8a	AF700	BioLegend
Live/Dead Fixable Yellow		BV570	ThermoFisher

B.2 Antibodies for analysis of G,M, and GM-CSF receptors

Antibody	Clone	Colour/Channel	Manufacturer
CD114	LMM741	PerCP-Cy5.5 and APC	BioLegend
CD114	LMM741	BV421	BD Biosciences
CD115	9-4D2-1E4	PerCP-Cy5.5	BioLegend
CD115	9-4D2-1E4	BV421	BD Biosciences
CD116	4H1	PE-Cy7	BioLegend

B.3 Antibodies for phosphoflow

Antibody	Clone	Colour/Channel	Manufacturer
pSTAT3	4/P-STAT3	AF647	BD Biosciences
pSTAT5	Clone 47/Stat5	AF647	BD Biosciences
pERK	197G2	AF647	CST
Anti-rabbit isotype		AF647	CST
Anti-mouse isotype		AF647	BD Biosciences

B.4 Antibodies for cell sorting (*in vitro* subsets)

Antibody	Clone	Colour/Channel	Manufacturer
CD7	M-T701	PerCP-Cy5.5	BD Biosciences
CD5	UCHT2	BV605	BD Biosciences
CD1a	HI149	PE-Cy7	BioLegend
CD44	IM7	APC-Cy7	BD Biosciences
DAPI		BV450	ThermoFisher

B.5 Antibodies for cell sorting (thymus populations)

Antibody	Clone	Colour/Channel	Manufacturer
CD45	HI30	AF700	BioLegend
CD123	6H6	Biotin	BioLegend
CD56	5.1H11	Biotin	BioLegend
CD14	63D3	Biotin	BioLegend
CD11c	3.9	Biotin	BioLegend
CD19	HIB19	Biotin	BioLegend
CD3	UCHT-1	BV480	BD Biosciences
CD4	OKT4	PerCP-Cy5.5	BioLegend
CD8	RPA-T8	BV650	BD Biosciences
Streptavidin		Biotin	BioLegend
DAPI		BV450	Thermofisher

B.6 Antibodies for discrimination of *in vitro* T-cell populations

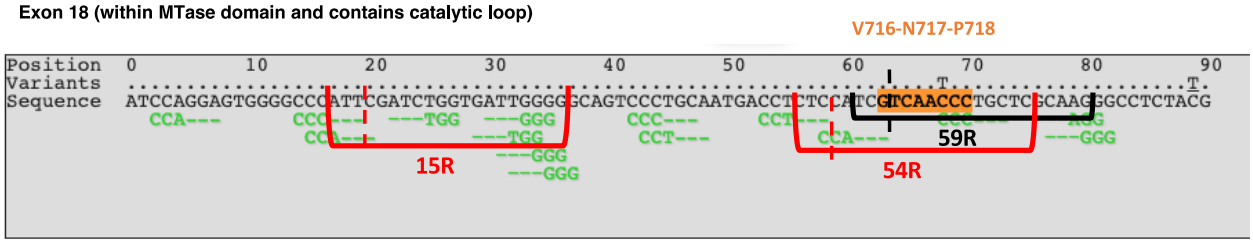
Panel	Antibody	Clone	Colour/Channel	Manufacturer
Early	CD38	HIT2	PerCP-Cy5.5	BioLegend
	CD34	581	APC	BioLegend
	CD44	IM7	APC-Cy7	BD Biosciences
	CD5	UCHT2	BV605	BD Biosciences
	CD7	M-T701	BV786	BD Biosciences
	CD1a	HI149	PE-Cy7	BioLegend
Late	TCRgd	B1	PerCP-Cy5.5	BioLegend
	CD8	RPA-T8	APC	BD Biosciences
	CD3	UCHT1	BV510	BioLegend
	CD4	RPA-T4	BV605	BD Biosciences
	TCRab	IP26	PE	BioLegend

Appendix C *In vitro*-derived T-cell subsets and thymus populations isolated for RNA-seq and WGBS

Population	# replicates		Sorted phenotype
	RNA-seq	WGBS	
Day 7	3	1	CD7+CD5-CD44+CD1a-
Day 10	3	1	CD7+CD44+CD1a-
Day 14	3	1	CD7+CD44-CD1a-
Day 36	3	1	CD7+CD44-CD1a+
Day 42	3	0	CD7+CD4+CD8+
DN PNT	1	1	CD45dimCD123-CD56-CD14-CD11c-CD19-CD3-CD4-CD8-
DP PNT	1	1	CD4+CD8+

Appendix D sgRNA targeting of the *DNMT3A* locus

Location of guide RNAs within Exon 18 of *DNMT3A*. The region highlighted in orange encodes the catalytic loop within the methyltransferase domain of *DNMT3A*.

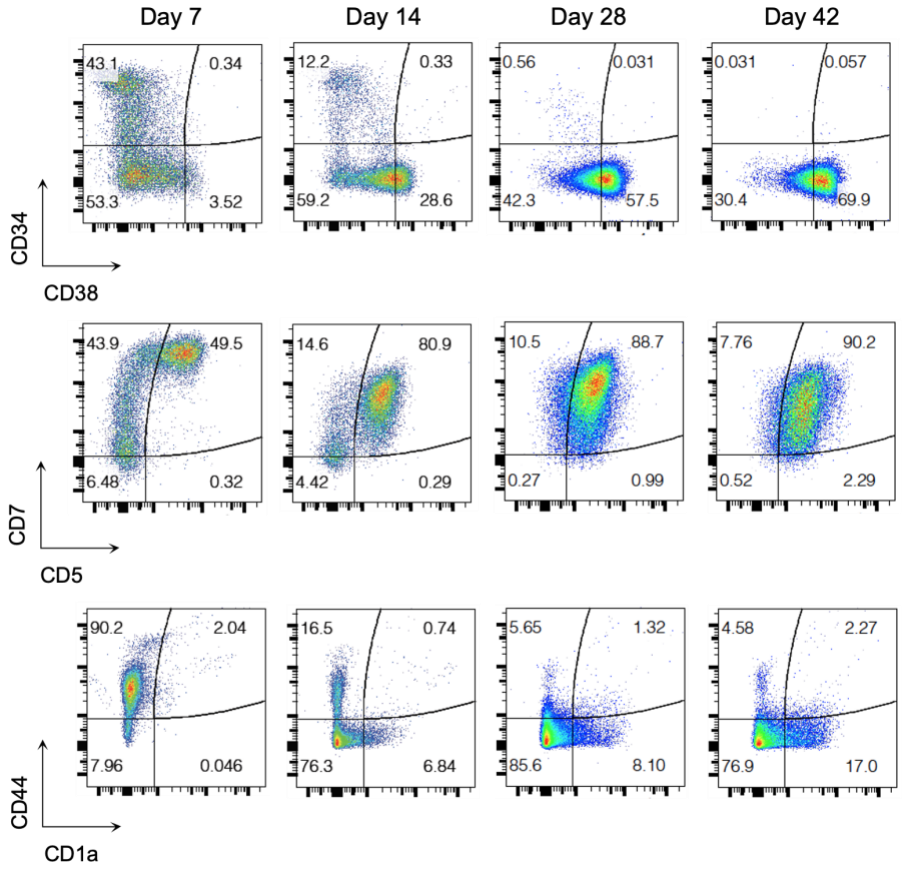


Appendix E shRNA, sgRNA, and primer sequences

Class	Name	Sequence
shRNA	shDNMT3A TRCN0000035754	CCCAAGGTCAAGGAGATTATT
	shDNMT3A TRCN0000035755	CCGGCTCTTCTTTGAGTTCTA
	shDNMT3A TRCN0000035756	GCCTCAGAGCTATTACCCAAT
	shDNMT3A TRCN0000035758	CCACCAGAAGAAGAGAAGAAT
	shScrambled	CTAAGGTTAAGTCGCCCTCG
sgRNA	DNMT3A 15R	CCCCAATCACCAGATCGAATGTTTTAGAGCTATGCT
	DNMT3A 54R	GAGCAGGGTTGACGATGGAGGTTTTAGAGCTATGCT
	DNMT3A 59R	CTTGCGAGCAGGGTTGACGAGTTTTAGAGCTATGCT
	EMX1	GAGTCCGAGCAGAAGAAGAA
Primer	DNMT3A FW	TGTCCATGGTTGCAGCTAGG
	DNMT3A RV	GCGATCATCTCCCTCCTTGG
	DNMT3A FW Nested	TGGGTCTCCTCTCTTTCGTG
	DNMT3A RV Nested	CAGGAGGCGGTAGAACTCAA
	EMX1 FW	AGCAGCTCTGTGACCCTTTG
	EMX1 RV	TTGTCCCTCTGTCAATGGCG

Appendix F Immunophenotypes of *in vitro*-derived T-cell subsets in additional cord pools

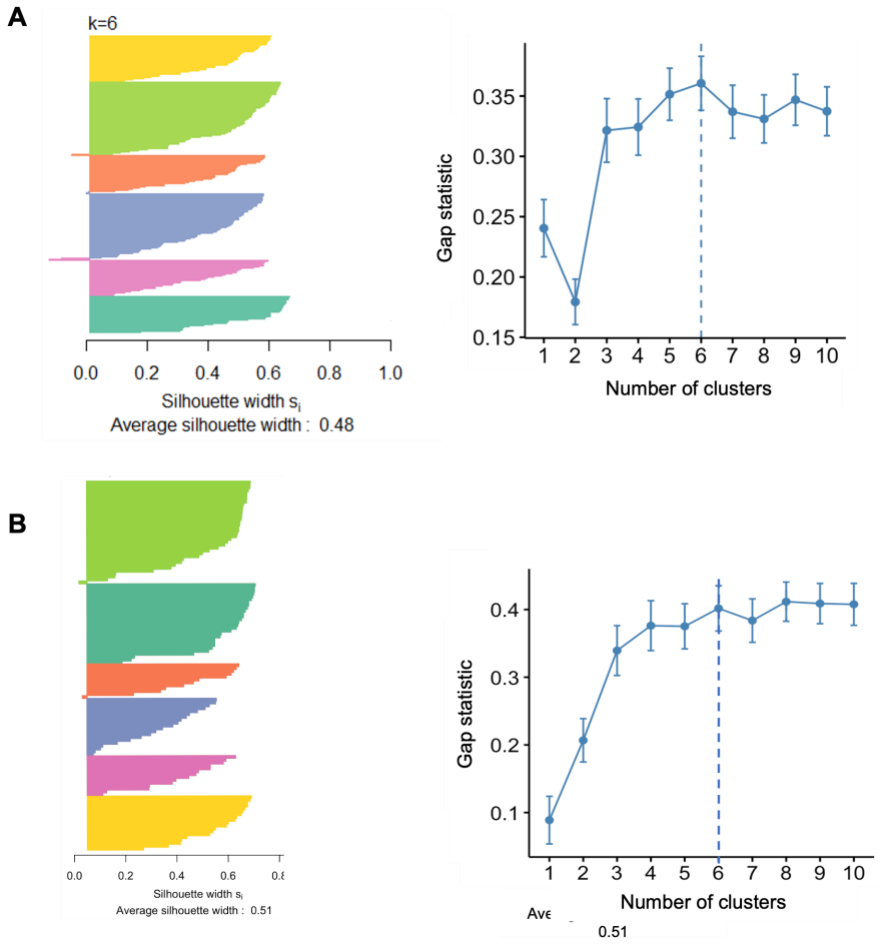
Flow cytometric analysis of T-cell subsets derived in vitro from CD34+ cord blood. Cells were collected at the indicated time points throughout differentiation. Results of a second cord blood donor are shown.



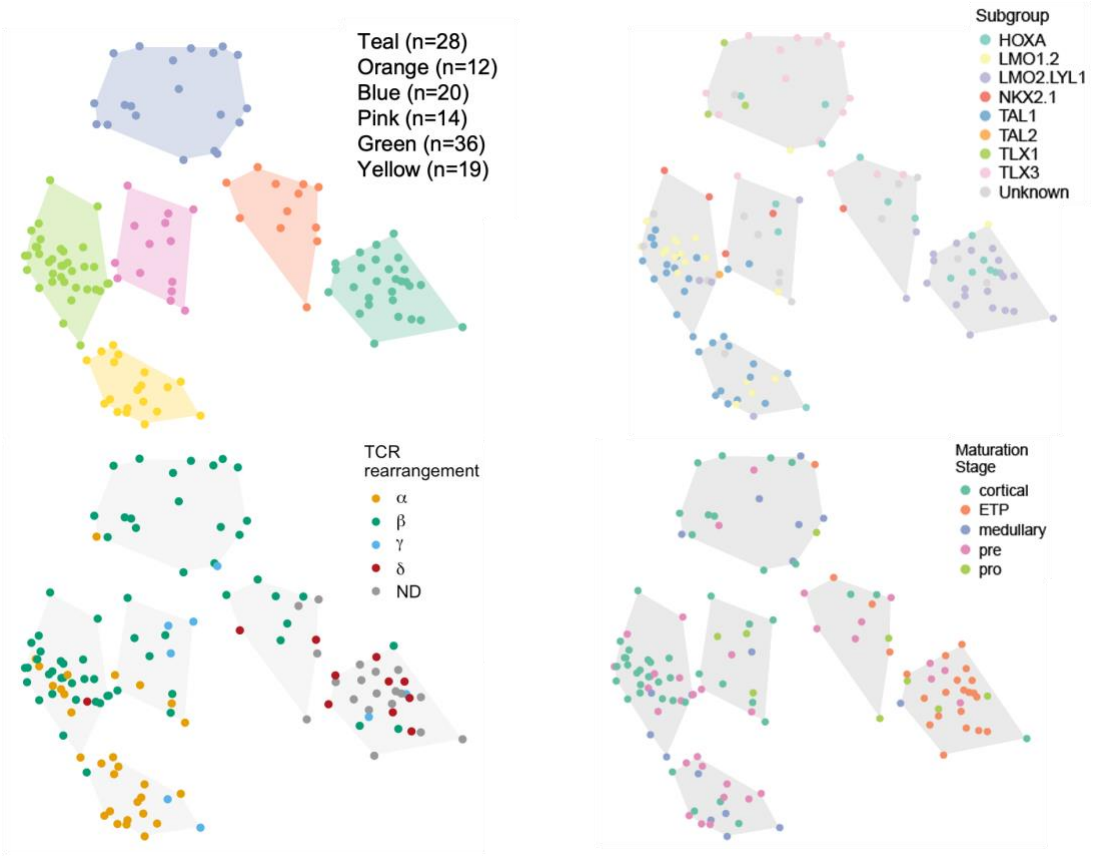
Appendix G Supplementary figures related to Chapter 2

G.1 TARGET and CGAD *k*-medoid clustering measures

Silhouette width and gap statistic measures of clustering robustness for (A) TARGET and (B) CGAD RNA-seq datasets.

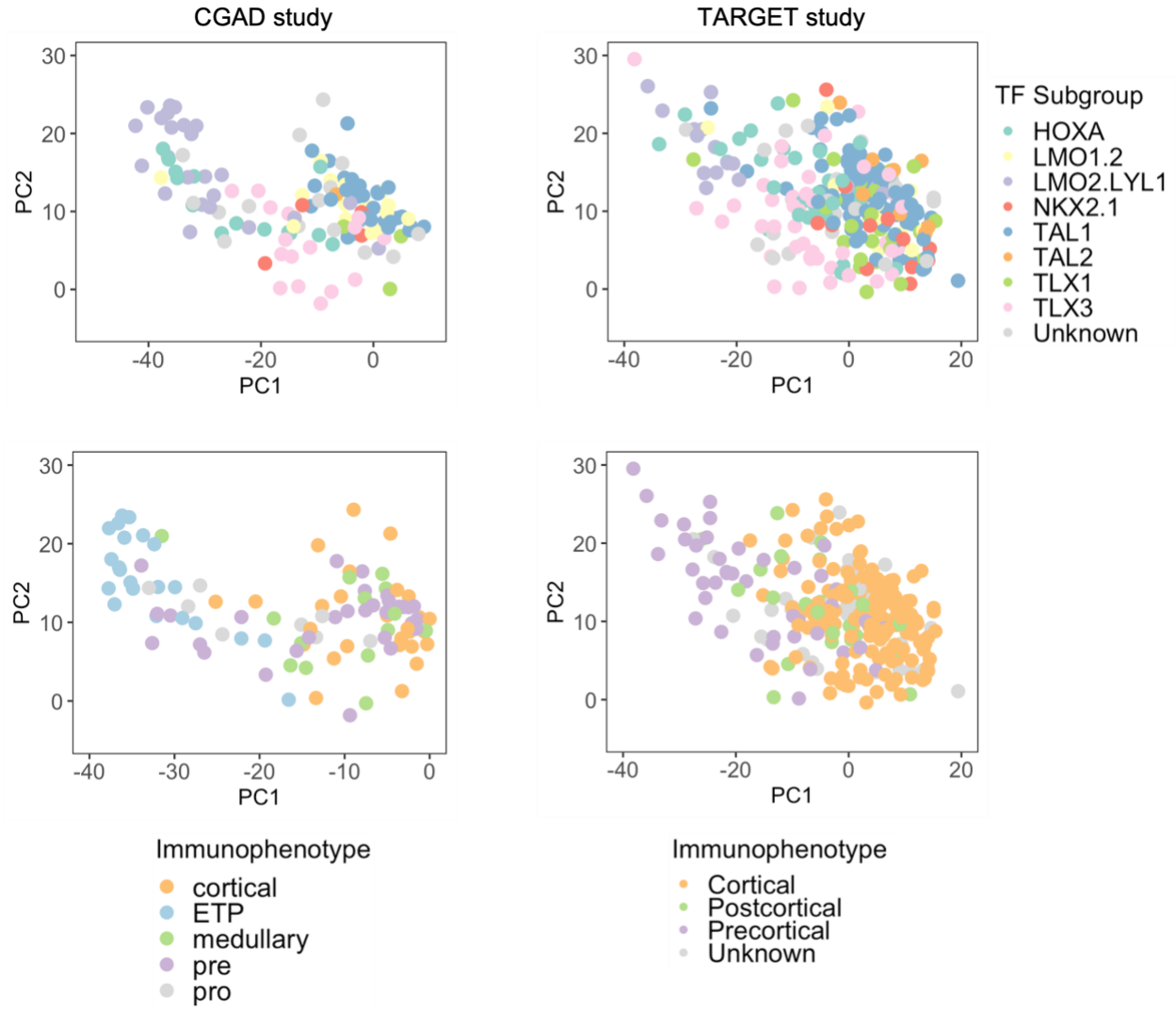


G.2 *k*-medoid clustering of CGAD T-ALL samples.



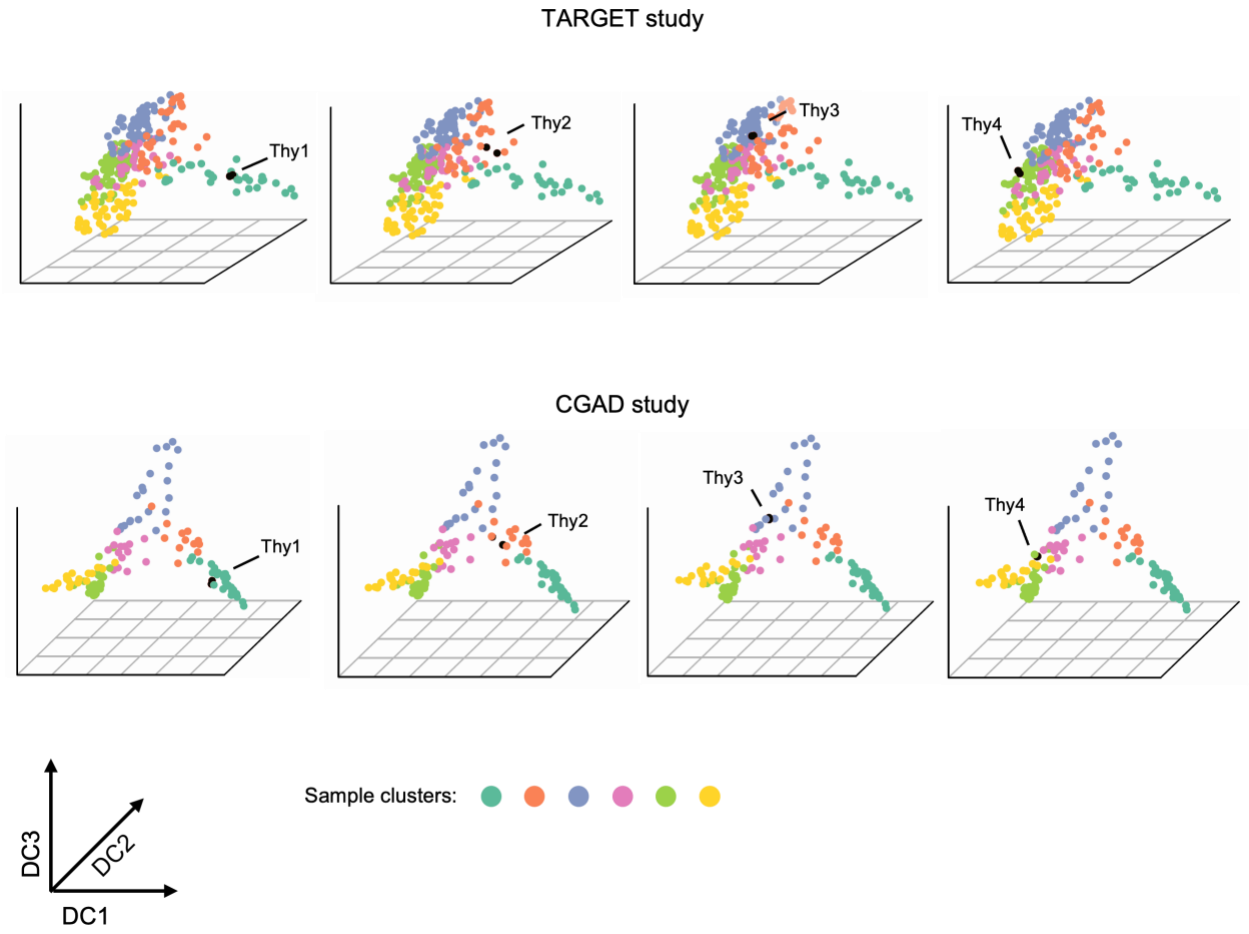
G.3 Projection of CGAD and TARGET T-ALL cohorts into PCA space defined by normal T-cell subsets.

T-ALL samples are coloured based on transcription factor (TF) membership (top) or immunophenotype (bottom).



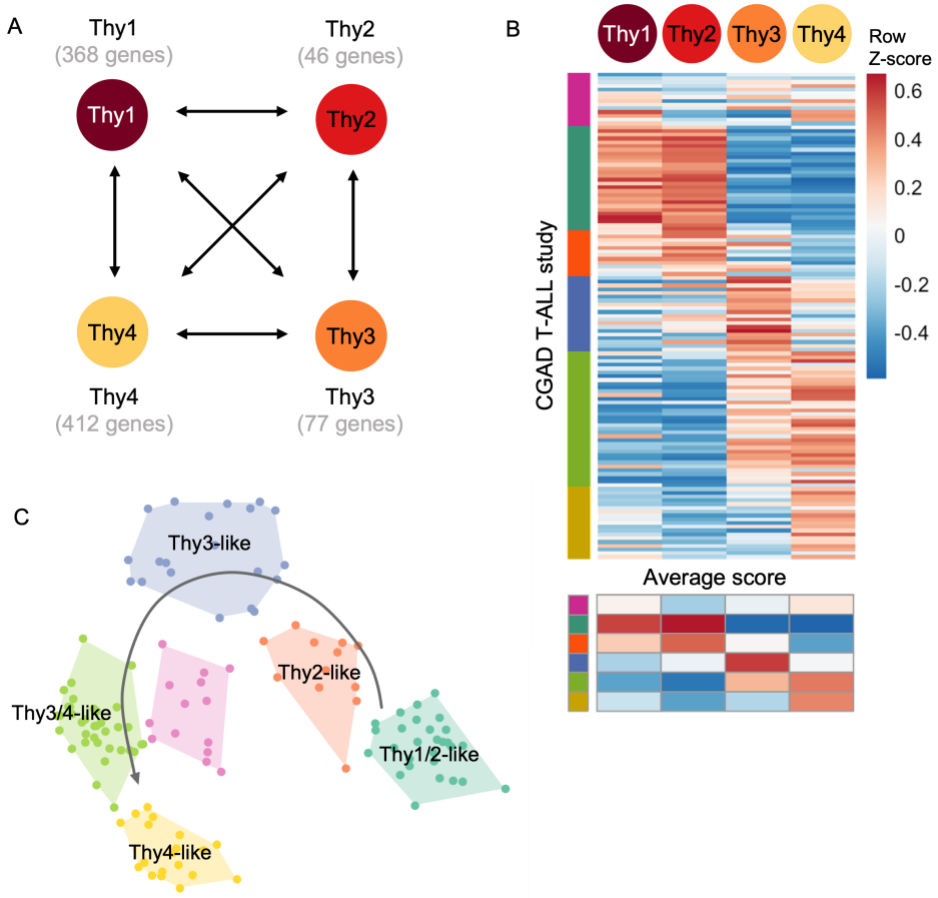
G.4 Trajectory analysis of TARGET and CGAD T-ALL cohorts.

TARGET and CGAD T-ALL cohorts were analyzed along with normal thymic progenitors by rank normalization of 2,824 genes implicated in T-cell development. Dimensionality reduction was performed by diffusion map.



G.5 Derivation of Thy1-4 subset signatures.

A) Differential gene expression comparisons performed to generate gene set signatures.
B) Gene set variation analysis (GSVA) was performed on the CGAD cohort.
C) PCA plot of CGAD samples with optimal k=6 clustering and Thy1-4 signature annotations. A similar result was obtained for TARGET samples.



G.6 Clinical outcome correlations.

A) Event-free survival probability for each sample cluster in the TARGET cohort *(Log-rank test)*.
B) Day 29 MRD divided by ETP type and TF subgroup.
C) PCA plot of the TARGET samples with k=6 optimal clustering and annotated by ETP type.
D) Day 29 MRD divided by ETP type for Cluster Teal samples only.
For panels B and D, all significant differences are shown. **, p<.05; **, p<.01; ***, p<.001; ****, p<.0001. (Dunn's test with Benjamini-Hochberg multiple test correction).*