# To B or not to B-cell - The control of cellular identity in lymphocyte development

Tingvall-Gustafsson, Johanna

2025

*Document Version:*
Publisher's PDF, also known as Version of record

# To B or not to B-cell

## The control of cellular identity in lymphocyte development

JOHANNA TINGVALL-GUSTAFSSON
DEPARTMENT OF LABORATORY MEDICINE | FACULTY OF MEDICINE | LUND UNIVERSITY

# To B or not to B-cell

The control of cellular identity in lymphocyte development

# To B or not to B-cell

## The control of cellular identity in lymphocyte development

Johanna Tingvall-Gustafsson


LUND
UNIVERSITY

DOCTORAL DISSERTATION

Doctoral dissertation for the degree of Doctor of Philosophy (PhD) at the Faculty of Medicine at Lund University, to be publicly defended on November 6th at 13:00 in Segerfalksalen, BMC A10, Lund

*Faculty opponent:*
Professor Cornelis Murre, PhD
University of California, San Diego (UCSD)

Organization
**LUND UNIVERSITY**

Department of Laboratory Medicine
Sölvegatan 19
SE-223 62, Lund
Sweden

Author(s)
Johanna Tingvall-Gustafsson

Document name
**DOCTORAL DISSERTATION**

Date of disputation
2025-11-06

Sponsoring organization

Title and subtitle

To B or not to B-cell - The control of cellular identity in lymphocyte development

Abstract

Cell fate determination is a complex process orchestrated by lineage specific transcription factors (TFs), acting in concert with the chromatin landscape, to enforce lineage restriction and establish stable cellular identity. To gain a deeper understanding of this interplay in lymphoid development, a global screen of proximity interacting partners of key regulators was conducted in *Paper I*, revealing that the interactome of lineage-specific TFs exhibit a substantial overlap in interaction partners with enrichment of proteins involved in the general transcriptional machinery and epigenetic regulators. Among these, we identified ARID1a, a core component of the BAF chromatin remodeling complex, as a critical regulator of the chromatin landscape in T cell development, acting in concert with TCF7 and GATA3. TFs, the drivers of cellular identity, exert their gene-regulatory potential through binding to cis-regulatory elements (CREs), which despite their often distal location within the linear genome can be juxtaposed to target genes via chromatin looping. To facilitate the incorporation of genome topology into the functional interpretation of these distal CREs, we developed a bioinformatics tool, ICE-A, which we subsequently used for exploration of the gene regulatory networks (GRNs) in early lymphocyte development in *Paper II*. A highly dynamic chromatin landscape emerged, with context dependent interaction of enhancers to both lineage specific and broadly expressed genes. In addition, lineage-restricted transcription factors were found to intersect with core regulatory circuits of basic cellular functions, linking differentiation to survival and proliferative capacity. In *Paper III*, this mechanistic link, with potential implications in leukemic transformation, is exemplified by B cell TF mediated regulation of the proto-oncogene *Myc*. Opposing regulatory control by EBF1 and PAX5 on *Myc* expression through a distal enhancer region, establishes a regulatory loop that controls B cell expansion. The same TFs also serve as key regulators in a dose-dependent epigenetic switch that drive uncommitted progenitors into a state of stable B cell identity, as demonstrated in *Paper IV*. This "Big bang" of B cells development is marked by acquisition of the B lineage epigenetic and transcriptional program, combined with polycomb mediated repression of alternative lineage potential to ensure stable B cell commitment.

Key words

Lymphocyte development, Cell fate, Gene-regulatory networks, Chromatin dynamics

Classification system and/or index terms (if any)

Supplementary bibliographical information

Language
English, Swedish

ISSN and key title
1652-8220

ISBN
978-91-8021-775-0

Recipient's notes

Number of pages
132

Price

Security classification



Signature _____ Date 2025-09-25

# To B or not to B-cell

## The control of cellular identity in lymphocyte development

Johanna Tingvall-Gustafsson


LUND
UNIVERSITY

**Front Cover:** *"A guiding hand towards the epigenetic horizon"* by Johanna Tingvall-Gustafsson. Partial images were generated using Copilot and subsequently edited and complemented using Adobe Photoshop and Illustrator.



Faculty of Medicine
Department of Laboratory Medicine

ISBN: 978-91-8021-775-0
ISSN: 1652-8220

Printed in Sweden by Media-Tryck, Lund University, Lund 2025

*Blood, sweat and a jug of fireball*

# Contents

# List of publications

This thesis is based on the following publications:

I **ARID1a associates with lymphoid-restricted transcription factors and has an essential role in T cell development**

A. Astori*, **J. Tingvall-Gustafsson***, J. Kuruvilla, E. Coyaud, E.M.N. Laurent, M. Sunnerhagen, J. Åhsberg, J. Ungerbäck, T. Strid, M. Sigvardsson, B. Raught, R. Somasundaram

*Journal of immunology*, Vol. 205, Issue 5, p. 1419-1432, 2020

*these authors contributed equally

II **Chromatin interaction-based annotation of cis-regulatory elements reveals highly dynamic promoter-enhancer interactions in lymphocyte development**

**J. Tingvall-Gustafsson**, C. Jensen, J. Ungerbäck, M. Sigvardsson

*iScience*, Vol. 28, Issue 7, 112855, 2025

III **EBF1 and PAX5 control pro-B cell expansion via opposing regulation of the Myc gene**

R. Somasundaram, C. Jensen, **J. Tingvall-Gustafsson**, J. Åhsberg, K. Okuyama, M. Prasad, J.R. Hagman, X. Wang, S. Soneji, T. Strid, J. Ungerbäck, M. Sigvardsson

*Blood*, Vol. 137, Issue 22, p. 3037-3049, 2021

IV **B-lineage commitment is dependent on a reversible epigenetic switch**

**J. Tingvall-Gustafsson**, K. Hellerstedt, J. Ungerbäck, M. Sigvardsson

Accepted for publication in *Genes & Development*

Publications not included in the thesis:

I **B lymphocyte specification is preceded by extensive epigenetic priming in multipotent progenitors**

T. Strid, K. Okuyama, **J. Tingvall-Gustafsson**, J. Kuruvilla, C. Jensen, S. Lang, M. Prasad, R. Somasundaram, J. Åhsberg, S. Cristobal, S. Soneji, J. Ungerbäck, M. Sigvardsson

*Journal of Immunology*, Vol. 206, Issue 11, p. 2700-2713, 2021

II **B-lineage acute lymphoblastic leukemia causes cell-autonomous defects in long-term hematopoietic stem cell function**

C. Jensen, J. Åhsberg, **J. Tingvall-Gustafsson**, R. Somasundaram, S. Lang, J. Ungerbäck, A. Porwit, S. Soneji, M. Sigvardsson

*Haematologica*, Vol. 108, Issue 11, p. 3175-3180, 2023

III **Ontogeny shapes the ability of ETV6::RUNX1 to enhance hematopoietic stem cell self-renewal and disrupt early lymphopoiesis**

M. Eldeeb, A. Konturek-Ciesla, Q. Zhang, S. Kharazi, **J. Tingvall-Gustafsson**, J. Ungerbäck, M. Sigvardsson, D. Bryder

*Leukemia*, Vol. 38, Issue 2, p. 455-459, 2024

# Abstract

Cell fate determination is a complex process orchestrated by lineage specific transcription factors (TFs), acting in concert with the chromatin landscape, to enforce lineage restriction and establish stable cellular identity. To gain a deeper understanding of this interplay in lymphoid development, a global screen of proximity interacting partners of key regulators was conducted in *Paper I*, revealing that the interactome of lineage-specific TFs exhibit a substantial overlap in interaction partners with enrichment of proteins involved in the general transcriptional machinery and epigenetic regulators. Among these, we identified ARID1a, a core component of the BAF chromatin remodeling complex, as a critical regulator of the chromatin landscape in T cell development, acting in concert with TCF7 and GATA3. TFs, the drivers of cellular identity, exert their gene regulatory potential through binding to cis-regulatory elements (CREs), which despite their often distal location within the linear genome can be juxtaposed to target genes via chromatin looping. To facilitate the incorporation of genome topology into the functional interpretation of these distal CREs, we developed a bioinformatics tool, ICE-A, which we subsequently used for exploration of the gene regulatory networks (GRNs) in early lymphocyte development in *Paper II*. A highly dynamic chromatin landscape emerged, with context dependent interaction of enhancers to both lineage specific and broadly expressed genes. In addition, lineage-restricted transcription factors were found to intersect with core regulatory circuits of basic cellular functions, linking differentiation to survival and proliferative capacity. In *Paper III*, this mechanistic link, with potential implications in leukemic transformation, is exemplified by B cell TF mediated regulation of the proto-oncogene *Myc*. Opposing regulatory control by EBF1 and PAX5 on *Myc* expression through a distal enhancer region, establishes a regulatory loop that controls B cell expansion. The same TFs also serve as key regulators in a dose-dependent epigenetic switch that drive uncommitted progenitors into a state of stable B cell identity, as demonstrated in *Paper IV*. This "Big bang" of B cells development is marked by acquisition of the B lineage epigenetic and transcriptional program, combined with polycomb mediated repression of alternative lineage potential to ensure stable B cell commitment.

# Popular summary

The hematopoietic system consists of an ensemble of cellular actors that make up our blood, each playing a crucial part, from oxygen transport and blood clotting to providing organisms with an immune system. This thesis is focused on lymphocytes, a subset of hematopoietic cells including B- and T cells, the lead actors in adaptive immunity. Despite the diverse roles, all cells within the hematopoietic system originate from a singe type of unspecialized cells, known as the hematopoietic stem cell (HSC), residing in the bone marrow. With a shared genetic material, the cellular roles are mediated by a group of proteins known as transcription factors (TFs), acing as directors of cellular identity by controlling which genes are active. In addition, the DNA, which provide the script for cellular function, is wrapped around histone proteins, forming in a fibrous structure known as chromatin. How tightly or loosely this chromatin is packed affects whether TFs can access certain genes and turn them on or off. On top of this, chromatin folds into a complex three-dimensional (3D) structure, creating loops that bring distant parts of the genome together, allowing for a fine-tuned regulation of gene expression. In an interplay with this chromatin landscape, that set the stage for cellular identity, the TFs direct unspecialized cells into gradually more restricted roles, until a point of no return is reached and the fate of the cell is decided.

To understand how TFs shape cell identity, we carried out a screen to uncover their interaction partners (*Paper I*). We found that many TFs partner up with proteins that influence how accessible the chromatin landscape is. This highlight the importance of the two-way relationship between TFs and chromatin, where TFs guide identity by reshaping chromatin, and chromatin in turn regulates TF function. One protein, called ARID1a, stood out as a frequent partner to TFs that guide lymphocyte identity. By deleting the gene that code for ARID1a, we uncovered that the protein has a particularly important role in the development of T cells, a type of immune cells that can destroy infected cells. The BAF complex, with ARID1a in its core, was found to selectively open up regions of DNA required for T cells to form by guidance of the T cell TFs TCF7 and GATA3. TFs often impact genes by binding to DNA sites that, in the linear genome, lie

far away from the gene they control. This is enabled by folding of the genome to bring these distal sites, known as enhancers, close to the genes in the 3D space. To explore this process, we developed a tool called ICE-A in *Paper II*, that pair enhancers with genes based on chromatin loops. By applying this tool to study lymphocyte development, we uncovered a dynamic web of interactions between enhancers and genes. A web which ensures that the main plot of lineage-specific gene expression unfolds correctly, while simultaneously overseeing the supporting scenes that maintain essential cellular processes.

These interconnected narratives of cellular identity and basic cellular processes, like survival and expansion, is further explored in *Paper III*. The focus here is on B cells, the co-star of adaptive immunity which produce antibodies that can recognize and neutralize invaders. In this act, the B cell TF EBF1 and PAX5 orchestrate a regulatory loop controlling cellular expansion via a protein called MYC. This type of cell specific control of both identity and survival can help ensure normal development, but if the actors go off script, the same connection can create a weakness that leads to leukemia.

In *Paper IV*, we investigate what happens when the unspecialized lymphoid cells are faced with the existential dilemma of what fate to adopt, and just like Hamlet, ask themselves the question: "To B or not to B(-cell)". Again the TFs EBF1 and PAX5 come to rescue. By directing a switch in the chromatin landscape, the cells are guided into stable B cell identity. In this "Big Bang" of B cell development, the TFs play a dual role, by both activating B cell signature genes, while simultaneously directing a complex known as PRC2 to silence alternative cellular fates.

# Populärvetenskaplig sammanfattning

Det hematopoetiska systemet består av en samling av celler, där varje cell, likt en skådespelare, spelar en avgörande roll i allt från syretransport och koagulering, till att förse oss med ett immunförsvar. Den här avhandlingen är fokuserad på lymfocyter, en typ av hematopoetiska celler som inkluderar B- och T celler, huvudrollsinnehavarna i det adaptiva immunförsvaret. Trots sina skilda roller, härstammar alla celler i det hematopoetiska systemet från samma typ av ospecialiserade celler i benmärgen, de så kallade hematopoetiska stamcellerna (HSC). Då dessa celler har samma uppsättning av genetiskt material, är det en grupp av proteiner, så kallade transkriptionsfaktorer (TFs), som agerar regissörer av cellidentitet genom att reglera vilka gener som är aktiva. DNA, som utgör manuskriptet för cellens funktion, är lindat runt histonproteiner och bildar en fibrös struktur som kallas kromatin. Hur kompakt detta kromatin är packat påverkar om transkriptionsfaktorerna kan komma åt vissa gener och slå på eller stänga av dem. Utöver detta, så viker sig kromatinet till en komplex tredimensionell (3D) struktur, vilket skapar loopar som för samman avlägsna delar av genomet och möjliggör en finjusterad reglering av genuttryck. I samspelet med detta kromatinlandskap, som därigenom agerar som en scen där cellens identitet gestaltas, styr TFs de ospecialiserade cellerna mot gradvis mer och mer begränsade roller, tills de når en punkt utan återvändo och cellens öde avgörs.

För att förstå hur TFs formar cellidentitet undersökte vi deras samarbetspartners (*Artikel I*). Vi fann att många TFs interagerar med proteiner som påverkar hur tillgängligt kromatinlandskapet är. Detta belyser vikten av det tvåvägsförhållande som finns mellan TFs och kromatin, där TFs styr cellens identitet genom att omforma kromatin, och kromatin i sin tur reglerar TFs funktion. Ett protein, kallat ARID1a, framträdde som en frekvent partner till de TFs som styr lymfocytidentitet. Genom att ta bort genen som kodar för detta protein upptäckte vi att ARID1a har en särskilt viktig roll i utvecklingen av T celler, en typ av immunceller som till exempel verkar genom att förstöra infekterade

celler. BAF-komplexet, som ARID1a är en del av, visade sig selektivt öppna upp DNA-regioner som är nödvändiga för cellmognad under vägledning av två T cell TFs, TCF7 och GATA3.

TFs påverkar ofta geners aktivitet genom att binda till DNA regioner som i det linjära genomet ligger långt ifrån den gen de kontrollerar. Detta möjliggörs genom att genomet viks så att dessa avlägsna regioner, som kallas enhancers, förs tätt intill generna i 3D rummet. För att undersöka denna process utvecklade vi ett verktyg kallat ICE-A i *Artikel II*, som parar ihop enhancers med gener baserat på kromatinloopar. Genom att använda detta verktyg för att studera utvecklingen av lymfocyter, upptäckte vi ett dynamiskt nätverk av interaktioner mellan enhancers och gener. Ett nätverk som säkerställer att huvudakten av linjespecifikt genuttryck fortskrider korrekt samtidigt som de säkerställer att cellernas basala funktioner upprätthålls.

Dessa sammanflätade akter om cellulär identitet och grundläggande cellulära processer, som överlevnad och expansion, utforskas vidare i *Artikel III*. Här ligger fokus på B celler, medspelaren i det adaptiva immunsystemet, som producerar antikroppar som i sin tur kan känna igen och neutralisera angripare. I denna akt koordinerar B cellernas transkriptionsfaktorer EBF1 och PAX5 en regulatorisk loop som kontrollerar cellernas expansion via ett protein kallat MYC. Denna typ av cellspecifik kontroll av både identitet och överlevnad, kan bidra till normal utveckling, men bidrar även med en sårbarhet som kan ge upphov till leukemi om spelarna avviker från manus.

I *Artikel IV* undersöker vi vad som händer när de ospecialiserade lymfoida cellerna ställs inför det existentiella dilemmat att välja sitt öde, och precis som Hamlet, ställer de sig frågan: ”Att vara eller icke vara (B-cell)”. Återigen kommer transkriptionsfaktorerna EBF1 och PAX5 till undsättning. Genom att styra en omställning i kromatinlandskapet leds cellerna in i en stabil B cellsidentitet. I detta ”Big Bang” av B cellernas utveckling spelar TFs en dubbel roll, genom att både aktivera B cellernas signaturgener och samtidigt guida ett komplex, känt som PRC2, för att tysta ner alternativa cellulära öden.

# Abbreviations

| | |
|---|---|
| 4-OHT | 4-Hydroxytamoxifen |
| Ab | Antibody |
| ARID1a | AT-rich interactive domain-containing protein 1A |
| ATAC-seq | Assay for transposase-accessible chromatin using sequencing |
| B-ALL | B cell acute lymphoblastic leukemia |
| BAF | BRG1/BRM associated factors |
| BCR | B cell receptor |
| BM | Bone marrow |
| ChIP-seq | Chromatin immunoprecipitation followed by sequencing |
| CLP | Common lymphoid progenitor |
| CRE | Cis-regulatory element |
| CRISPR | Clustered regularly interspaced short palindromic repeats |
| CT | Chromosome territory |
| CTCF | CCCTC-binding factor |
| Cut&Run | Cleavage under targets and release using nuclease |
| DN | Double negative (DN4-CD8-) |
| DNA | Deoxyribonucleic acid |
| DOX | Doxycycline |
| DP | Double positive (DN4+CD8+) |
| DRE | Distal regulatory element |
| EBF1 | Early B cell factor 1 |
| ER | Estrogen receptor |
| ETP | Early thymic progenitor |
| FACS | Fluorescence-activated cell sorting |
| FL | Fetal liver |
| FLT3 | FMS-like tyrosine kinase 3 |
| FOXO1 | Forkhead box O1 |

| | |
|---|---|
| GATA3 | GATA-binding factor 3 |
| GRN | Gene regulatory network |
| gRNA | Guide RNA |
| GSEA | Gene set enrichment analysis |
| HSC | Hematopoietic stem cell |
| iCas9 | Inducible Cas9 |
| IgH | Immunoglobulin heavy chain |
| IgL | Immunoglobulin light chain |
| IKZF1 | IKAROS family zinc finger 1 |
| IL-7(R) | Interleukin 7 (receptor) |
| KO | Knock-out |
| Lin | Lineage |
| LMPP | Lymphoid-primed multipotent progenitor |
| LSK | Lineage-Sca1+Kit1+ |
| NGS | Next-generation sequencing |
| NK | Natural killer |
| Op9-DL1 | Op9 delta-like 1 |
| PAX5 | Paired box 5 |
| PCA | Principal component analysis |
| PLAC-seq | Proximity ligation-assisted ChIP sequencing |
| PRC2 | Polycomb repressive complex 2 |
| PSEA | Peak set enrichment analysis |
| PXI | Proximity interacting partner |
| RNA-seq | RNA sequencing |
| RUNX1 | Runt-related transcription factor 1 |
| SCA1 | Stem cells antigen-1 |
| T-ALL | T cell acute lymphoblastic leukemia |
| TAD | Topologically associating domain |
| TCF7 | T-cell factor 7 (also known as TCF1) |
| TCR | T cell receptor |
| TF | Transcription factor |
| TH | $Ebf1^{+/-}Pax5^{+/-}$ transheterozygous |
| TSS | Transcription start site |
| UMAP | Uniform manifold approximation and projection |
| WT | Wild type |

# Part I

# RESEARCH CONTEXT

# Chapter 1

# Theoretical background

As the title, *"To B or not to B-cell"* suggest, this thesis is focused on the existential questions of identity and fate. Sure, I know Shakespeare did not have a relatively defined developmental stage of immature lymphoid progenitors in mind when he wrote his famous Hamlet quote. However the question of identity and fate is equally relevant in a cellular context, and considering the pivotal role B- and T cells have played in evolution by providing vertebrates with an adaptive immune system it is not so far fetched to call it existential.

The focus of early lymphocyte development, places this thesis in the intersection between, hematology, developmental biology and immunology. A biological context is given in *Section 1.1*, were all the relevant players of the hematopoietic system is presented along with their (hierarchical) organization. The general principles underlying the establishment of cellular identity, which allow this diverse repertoire of mature cells to arise from a common progenitor population is the focus on *Section 1.3*. However, unraveling of the mechanisms behind this process requires detailed investigation of the coordinated action of lineage specific transcription factors, acting in concert with the chromatin landscape, in order to drive transcriptional programs (covered in *Section 1.2*). Although the emphasize of this thesis is the development of B- and T cells, I regard it equally as an investigation into the gene regulatory networks that direct cell fate, illustrated through lymphocyte development. The same processes which drive the establishments cellular identity are closely linked to disease states, a connection further detailed in *Section 1.4*.

## 1.1 Hematopoiesis: Casting the roles of blood and immune cells

The hematopoietic system fulfills critical physiological functions, from ensuring oxygen delivery and hemostasis to orchestration of innate and adaptive immune responses. Despite their functional diversity, all blood and immune cells are derived from hematopoietic stem cells (HSCs), rare multipotent progenitors that sustain lifelong blood cell production through the tightly regulated process known as hematopoiesis [1]. Traditionally hematopoiesis has been seen as a hierarchical process, with HSCs at the top, giving rise to successive progenitor populations that undergo stepwise fate commitments to generate mature cells [1, 2]. Recent advances in single-cell omics technologies, however, have challenged this model, replacing the concept of discrete, homogeneous progenitors with a view of hematopoiesis as a dynamic and continuous landscape of cell fate decisions, where progenitor states display plasticity and overlapping potential along developmental trajectories [3–5]. At the core of this process lies a finely tuned balance of HSC self-renewal and differentiation, ensuring both long-term maintenance of blood homeostasis and the capacity to regenerate blood cells in response to specific physiological demands [6].

### 1.1.1 Introducing the hematopoietic ensemble

At the apex of hematopoiesis are HSCs that, via symmetric or asymmetric cellular division [7], give rise to multipotent progenitors (MPPs) that lack self-renewal capacity but retain multi-lineage differentiation potential [8, 9]. In the classical hierarchical depiction of hematopoiesis, it was assumed that the first lineage restriction event was the divergence of MPPs into the myeloid and lymphoid branch, each associated with a common progenitor population (common myeloid progenitors (CMP) or common lymphoid progenitors (CLP)) [10, 11]. This distinct separation between myeloid and lymphoid cells has later been replaced by a model in which the first bifurcation event is represented by the separation of megakaryocytic-erythroid progenitors (MEP) from cells with combined lympho-myeloid differentiation potential [12, 13]. The MEP precursors continue to differentiate to eventually give rise to mature oxygen transporting erythrocytes and platelets with blood clotting capacity [13]. The bipotent lympho-myeloid progenitors on the other hand, can be further separated into myeloid biased MMP3, eventually give rise to mature granulocytes and monocyte/macrophages as well as lymphoid primed MPP4 (LMPP) with the capacity to generate B-, T- and innate lymphoid cells (ILC) [14, 15]. All of these cell types have important function in the immune system, with B- and T cells re-

sponsible for adaptive immunity while the remaining cells being part of the innate defense system [16]. On top of this are dendritic cells (DC) which can not be defined to a single lineage, and provide a tight link between the innate and adaptive immune system via their antigen-presenting activity [17].

The emergence of single-cell technologies has enabled more comprehensive profiling of hematopoietic development, revealing substantial heterogeneity in the early progenitor compartment [18]. Although early single-cell studies proposed that unilineage-restricted cells emerge strictly from a continuum of low-primed, undifferentiated progenitors (known as cloud HSCPs) [3, 4], subsequent multi-omics profiling and cellular barcoding approaches have confirmed the presence of lineage priming [19–22]. Nonetheless, single cell technological have replaced

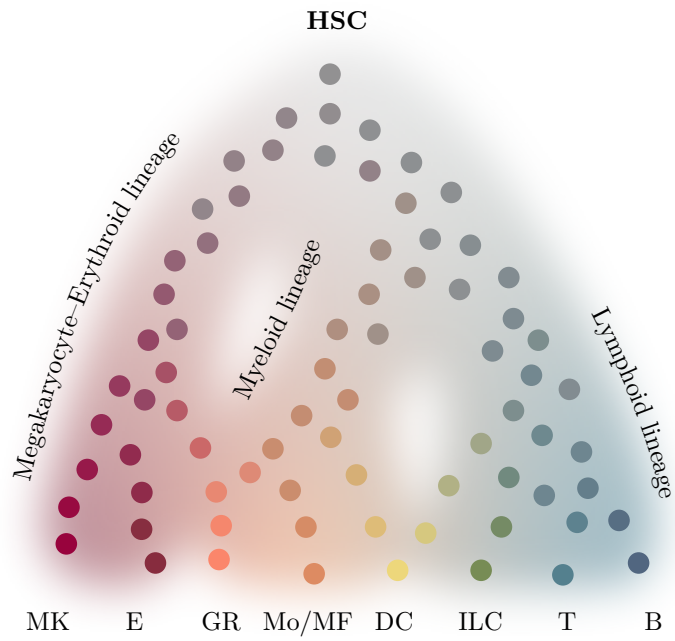



**Figure 1.1: Conceptional view of the revised model of hematopoiesis.** With the development of single-cell technologies, a novel model of hematopoiesis has emerged, where the classical stepwise progression through discrete, homogeneous progenitor populations has been replaced by a heterogeneous progenitor compartment consisting of cells with varying lineage bias. Mature cell fate is established through a gradual process in which cells progressively acquire lineage-specific characteristics while losing alternative potentials as they advance along developmental trajectories. HSC: Hematopoietic stem cell, MK: Megakaryocyte, E: Erythroid cell, GR: Granulocyte, Mo/MF: Monocyte/Macrophage, DC: Dendritic cell, ILC: Innate lymphoid cell, T: T cell, B: B cell.

the concept of discrete lineage restriction events with a more continuous landscape of progenitor cells with gradually divergent lineage-priming [23], though this continuum might be somewhat inflated by data sparsity bluing true lineage boundaries. Despite this revised perspective, the fundamental architecture of hematopoiesis appears largely conserved, even though variations in myeloid lineage branching has been reported, including a distinct early pathway to megakaryocyte pathway [24], co-segregation of additional myeloid populations (mast cells, basophil and eosinophils) with the MEPs [25–27] or the existence of two partly overlapping myeloid braces with respect to cellular output [19]. This variation likely reflects, at least in part, the heterogeneity within the early progenitor compartment, in which cells on an individual layer display variation in priming and multi-lineage potential [28]. At the population scale, however, these biases manifest as defined developmental trajectories, see Figure 1.1. Lineage specification and repression of alternative fates is tightly controlled by both cell intrinsic and extrinsic regulatory mechanism, as outlined in section *1.3*. Yet, the cellular output from hematopoietic stem and progenitor cells are likely also shaped by stochastic mechanisms.

### 1.1.2 Origin of the hematopoietic system

HSCs first emerge during early embryogenesis through fetal hematopoiesis, a process that establishes both a reservoir of undifferentiated HSCs and the mature blood cells required for embryonic development. This developmental program unfolds across distinct anatomical sites, reflecting the structural dynamics of embryogenesis and enable exposure of HSCs to diverse microenvironmental cues that regulate self-renewal and differentiation [29]. Mammalian hematopoiesis proceeds in two successive waves; primitive and definitive hematopoiesis [30]. In the yolk sac, primitive hematopoiesis is driven by transient HSCs with little or no self-renewal potential, producing mainly megakaryocyte–erythroid lineage cells to meet the embryo's oxygen demands, along with a limited repertoire of primitive myeloid and lymphoid cells [31, 32]. The emergence of definitive hematopoiesis is marked by the generation of long-term self-renewing HSCs with multi-lineage potential in the aorta-gonad-mesonephros (AGM) region [33]. These HSCs migrate to the placenta and later to the fetal liver (FL) which served as the principal hematopoietic organ during mid-gestation in both mice and humans [31, 34, 35]. In contrast to the predominantly quiescent HSCs of adult bone marrow, fetal HSCs exhibit a highly proliferative phenotype, reflecting the demands of rapid growth during developmental hematopoiesis [36]. By late gestation, HSCs colonize the BM, establishing the lifelong site of hematopoiesis and ensuring continuous production of all hematopoietic lineages. Within the BM, HSC self-renewal and differentiation

into lineage-restricted progenitors are governed by both intrinsic cellular mechanisms and extrinsic cues from specialized stem cell niches, including interactions with stromal cells and signaling from the surrounding microenvironment [37].

### 1.1.3 Lymphocytes: The lead actors of adaptive immunity

This thesis centers on the lymphoid branch of the hematopoietic system, which consist of B- and T lymphocytes, the key player of adaptive immunity. The hallmark of adaptive immunity is the ability to recognize a vast array of antigens and mount a tailored response, while simultaneously establishing immunological memory for faster and more effective responses upon re-exposure [38]. Although B- and T cells have distinct functions in adaptive immunity, their specificity is established through RAG1/RAG2-dependent V(D)J recombination, a process in which variable (V), diversity (D), and joining (J) gene segments are rearranged to form a diverse repertoire of antigen receptors [39–43]. B cells mediate humoral immunity by producing antigen-specific antibodies that neutralize pathogens and target them for destruction, whereas T cells control cell-mediated immunity by recognizing antigens on MHC molecules and coordinating cytotoxic and helper responses [44–47]. In addition, several ILC populations are lymphoid-derived, including natural killer (NK) cells, which are characterized by cytotoxic activity against virus-infected and malignant cells [48]. The full spectrum of mature lymphoid subsets are primarily derived from a heterogeneous population of bone marrow–resident lymphoid-primed multipotent progenitors (MPP4/LMPP) [49], via a process known as lymphopoiesis. Lymphocyte development is commonly described as a progression through successive progenitor stages, characterized by gradual lineage specification and restriction of alternative fates. While this framework simplifies the underlying heterogeneity and gradual lineage priming observed at the single-cell level, it remains a valuable model for both experimental investigation and conceptual understanding. Whereas this section outlines the central stages of B- and T cell development, *Section 1.3* provide a more comprehensive account of the mechanistic basis of lymphoid lineage commitment.

**B cell development**

Traditionally, B cell lineage restriction has been conceptualized as a stepwise progression from lymphoid primed multipotent progenitors (MPP4/LMPP), via a lymphoid restricted common lymphoid progenitor (CLP) population into fully committed B cell progenitor defined by CD19 surface expression [10, 12, 50]. However, the use of an *Igll1* ($\lambda$5) promoter–driven reporter mouse model [51] challenged this view by identifying CD19- B cell–committed cells within a Ly6D+ fraction of the CLP compartment [52], leading to the separation of CLP into ALP (All-lymphoid progenitor) and a BLP (B cell-biased lymphoid progenitor) associated increased expression of B cell specific genes [52]. Further dissection of the CLP compartment has revealed additional heterogeneity with regards to output of alternative lineage, consistent with the revised model of hematopoiesis. This include ILC bias in integrin a(2)b(7) (LPAM1) or CXCR6 expressing cells [53] as well as a DC primed fraction marked by BST2 (PDCA-1) [54, 55]. Although originally regarded as B cell restricted, single-cell *in vitro* assays demonstrated that a substantial fraction of the BLPs retain alternative lineage potential [56]. By stratifying the BLPs based on combined surface expression of GFRA2 and BST1, Jensen et al. [57] identified three successive BLP subsets (BLP1–3) corresponding to progressive lineage restriction. The earliest BLP1 (GFRA2-BST1-) subset retain broad lymphoid potential for all lymphoid lineages. Transition into BLP2 (GFRA2+BST1-) signifies B cell lineage specification and the loss of NK fate, while final progression into BLP3 (GFRA2+BST1+) marks the point of stable B cell commitment.

A central aspect of B cell development is the expression and assembly of functional B cell receptor (BCR), consisting of immunoglobulin heavy (IgH) and light (IgL) chains [41]. In the germline configuration, the variable (antigen-binding) part of the *Ig* genes are split into discontinuous V, D (*IgH* only) and J segments, requiring recombination based joining in order to generate functional receptor genes [58, 59]. In addition to ensuring a diverse BCR repertoire for adaptive immunity, the (pre-)BCR also has a crucial role in B cell development and proliferation [41]. IgH recombination is initiated already in uncommitted B cell progenitors by recombination-activating gene (RAG)-dependent cleavage at recombination signal sequences (RSSs) flanking the D and J gene segments [60, 61]. Random nucleotides are then added by deoxynucleotidyl transferase (TdT, encoded by *Dntt*) for enhanced diversity, before joining of the D and J segments. Subsequent recombination of a V and D-J segment in the pro-B cell stage marks complete V-DJ *IgH* rearrangement [62]. This transition from CD19- B cell progenitors coincides with activation of B cell–specific gene expression networks, driving progression along the B cell lineage [63]. In conjugation with IL-7 signaling, functionally assembled pre-BCR trigger a proliferative burst

[64] and down-regulation of RAG protein levels [65]. Upon the transition into a non-cycling small pre-B stage, RAG and TdT is reactivated, allowing for rearrangement of the V-J segments of the light chain. Once fully rearranged, the Immature B cells marked by IgM surface expression, is subjected to a self-reactivity checkpoint. Binding of IgM to endogenous antigen trigger receptor editing until ceased BCR-dependent signaling indicate lost self-reactivity [66]. Immature B cells exit the BM and migrate to the spleen as transitional B cells, where they undergo additional selection before differentiating into either Marginal zone B (MZB) or follicular B cells (FoB). Whereas MZBs mount rapid innate-like responses in a mostly T cell independent manner [67], FoBs serve as the central mediators of adaptive humoral immunity, orchestrating germinal center reactions for generation of class-switched, affinity-matured antibodies and long-term immunological memory [68]. For a schematic overview of the different stages of B cell development, including alternative classifications (Hardy fractions [62]), lineage restriction and defining cell surface markers, see Figure 1.2.

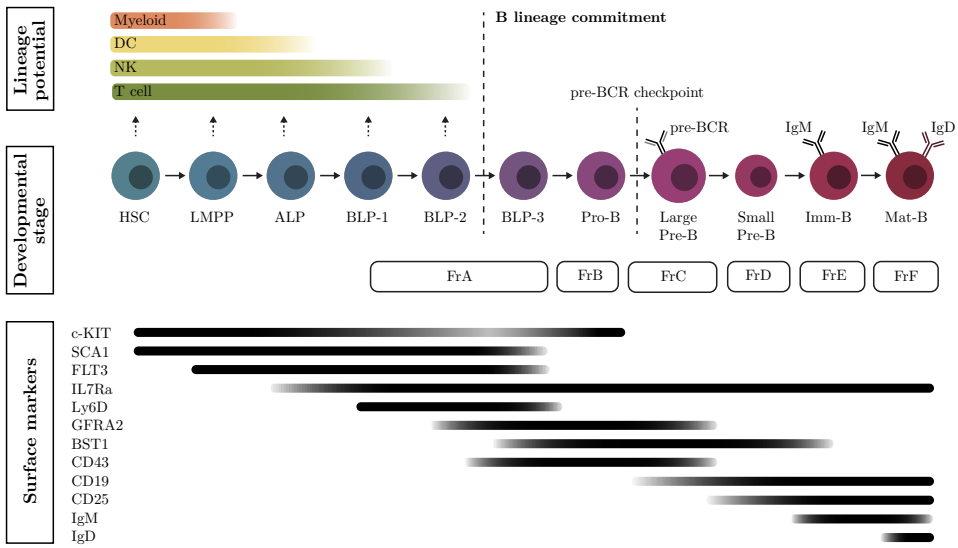



**Figure 1.2: B cell developmental progression.** Depiction of the key developmental stages in the differentiation of HSC into mature B cells, including alternative classification (Hardy fractions) and lineage restriction. The expression pattern of selected surface markers commonly used to define the developmental stages are presented.

**T cell development**

T cell development is initiated by migration of lymphoid primed progenitors (LMPP/CLP) to the thymus, which represent the primary location of T lineage restriction and commitment [69–72]. Within the thymic niche, interaction with cortical epithelial cells activate Notch signaling, which initiate the T cell differentiation program [73, 74]. The maturation process of early thymic progenitors (ETP) into mature T cells can be monitored by the surface expression of CD4 and CD8, where the double negative (DN) stage represents the earlier thymocytes. The DN population can be further subdivided into DN1-4, based on the expression of CD44, CD25 and KIT [75]. The earliest stages DN1 (CD44+CD25-) through DN2a (CD44+CD25+Kit$^{high}$), sometimes referred to as "phase 1", exhibit early T cell lineage specification but maintain the capacity to differentiate into other lineages [76–78]. Transition to the DN2b (CD44+CD25+Kit$^{int}$) marks stable T lineage commitment and entry into "phase 2" [78, 79], accompanied by activation of genes required for TCR recombination (*Rag1/2*) and signaling (e.g. *CD3d/e/g*, *Lck*) [80, 81]. Conventional $\alpha\beta$T cells first rearrange the $\beta$ locus followed by pre-TCR mediated $\beta$-section at the DN3a (CD44-CD25+CD27$^{low}$CD28-) to DN3b (CD44-CD25+CD27$^{high}$CD28+) transition. Successful TCR$\beta$ rearrangement trigger extensive proliferation and progression into the CD4/CD8 double positive state (DP) via a DN4 stage (CD44-CD25-), here TCR$\alpha$ rearrangement finalizes the assembly of the complete TCR [82]. Unlike B cell receptors which directly recognize native antigens, TCRs recognize antigen peptides only in the context of MHC presentation, creating a requirement for selection based on self-MHC recognition [83]. Thymocytes that successfully pass this checkpoint mature into either CD4 single-positive (CD4SP) T helper cells restricted to MHC class II, or MHC class I recognizing CD8 single-positive (CD8SP) cytotoxic T cells, followed by negative selection for elimination self-reactive T cells [83, 84]. At the DN2–3 stages, a subset of thymocytes undergo TCR$\gamma$ and TCR$\delta$ recombination, diverting from the $\alpha\beta$ pathway to generate $\gamma\delta$ T cells, a functionally diverse subset that bridges innate and adaptive immunity.[85, 86]. See Figure 1.3 for a an overview of T cell development in the thymus.

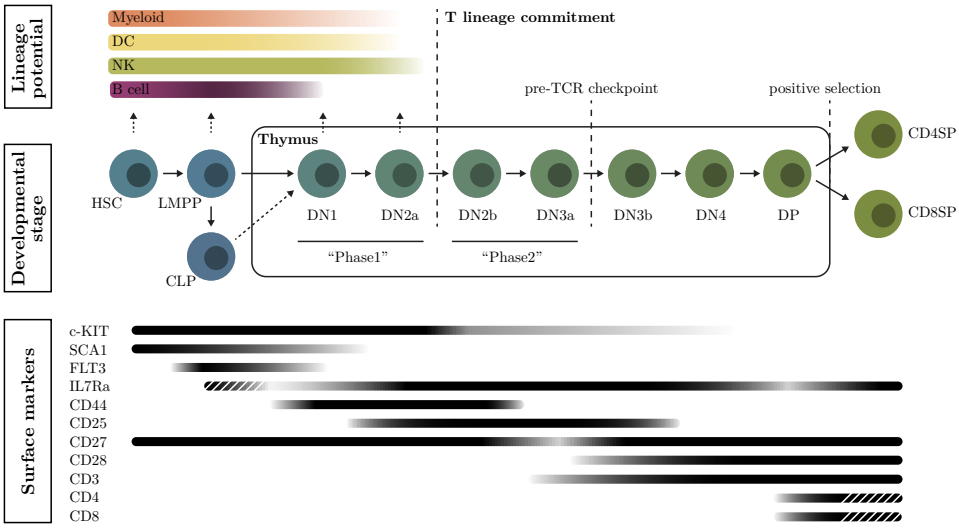



**Figure 1.3: T cell developmental progression.** Depiction of the key developmental stages in the differentiation of HSC into mature CD4/CD8 single positive T cells, including alternative lineage potential. The expression pattern of selected surface markers commonly used to define the developmental stages are presented.

### Lymphopoiesis in human versus mouse

This thesis, together with much of the background presented, is primarily based on murine studies, which offer a powerful platform for genetic manipulation and *in vivo* experimentation. Despite the evolutionary distance between mice and humans, these studies remain highly relevant to human biology due to the strong conservation of key transcriptional regulators, such as EBF1 and PAX5 in B cell development [87], and TCF7 and GATA3 in T cell development, with BCL11B marking lineage commitment in both species [88]. While the surface markers defining specific developmental stages differ between mice and humans, recent single-cell studies have shown that the core transcriptomic programs guiding B and T cell differentiation from a common progenitor are largely conserved [87, 88]. Thus, although interspecies differences warrant careful consideration during interpretation, murine models offer mechanistic insights that remain highly relevant to human lymphocyte development.

## 1.2 Gene regulation: The script that guide cell fate

The ability of an multicellular organism to generate different specialized cell types and respond to environmental ques is dependent on the ability for spatial and temporal control of the expression of genes. Although post-transcriptional regulation exist, the control of expression is to a large extent dependent on the transcription of genes into ribonucleic acid (RNA) [89]. Transcription is mediated by RNA polymerase enzymes, which via promoter recognition, open up the deoxyribonucleic acid (DNA) duplex and initiate templated RNA synthesis. An elongation complex then extend the newly synthesized RNA molecule until it encounter a termination signal [89]. The fine-tuned regulation of gene transcription is largely mediated by *transcription factors* (TFs), DNA binding protein that through sequence recognition, can instructed the transcriptional machinery and activator/repressor complexes to target genes [90, 91]. The TFs act within a complex *chromatin landscape*, where accessibility and histone modification provide an additional layer of gene regulation [92]. Beyond local chromatin states, the *three-dimensional (3D) genome architecture* shapes the regulatory landscape by coordinating interactions between genes and distal regulatory elements [93]. For an overview of the different layers of regulation that collectively control gene expression, see Figure 1.4. The subsequent sections aims to describe this layered regulation from a top down perspective, starting with the spatial organization of the genome and proceeding through the dynamics of local chromatin state to transcription factor activity.

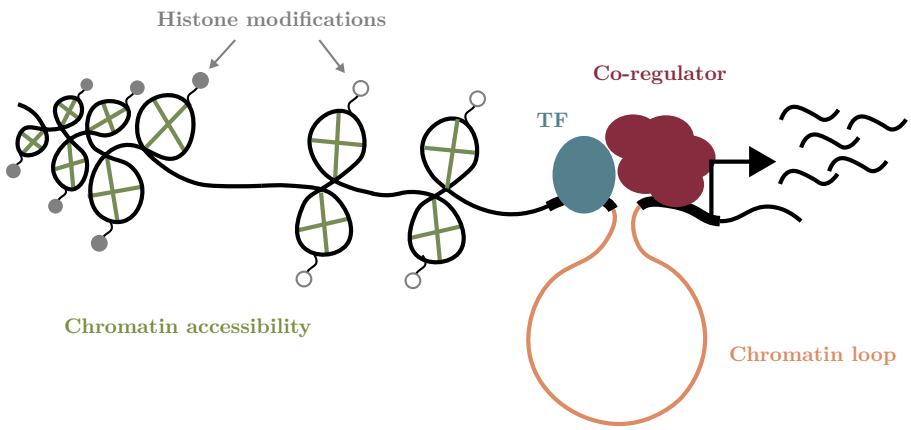



**Figure 1.4: Layers of gene regulatory mechanisms.** Schematic representation of different layers of gene regulatory mechanisms, including transcription factor (TF) recruitment of co-regulatory complex, chromatin accessibility and histone modification as well as 3D chromatin loop connecting distal regulatory elements to target genes.

### 1.2.1 The chromatin landscape: Setting the stage for transcriptional dynamics

In the nucleus, the genome is organized into a complex structure of protein, DNA (and RNA), known as chromatin. The smallest unit of chromatin is a nucleosome, consisting of ∼147 bp of DNA wrapped around a core of 8 histone protein (two of each H2A, H2B, H3, H4). The nucleosomes can assemble into chromatin fibers whose degree of compaction is tightly connected to chemical modifications of the histones proteins and the DNA itself [94]. Chromatin, in turn, folds into higher-order structures that serve both to compact the genome and provide a framework for dynamic multi-layered control of gene-regulation [95]. It is, however, important to recognize that these regulatory layers do not function in isolation, rather their states are mutually influenced as well as modulated by trans-acting factors, collectively shaping a chromatin landscape that supports stable cell identity while retaining the flexibility to respond to external cues or undergo differentiation.

Today we often think of genome topology as a relatively novel concept in gene-regulation. It is indeed true that the study of higher order structures of the genome and its dynamics has greatly benefited from recent technological advancement, however the idea of a dynamic organization of the genetic material dates back over a century. The term chromatin was first coined by Walther Flemming in 1882 to describe the substance found in the nucleus [96], which around the same time was identified as a mixture of nucleic acid and protein [97–99]. Although no uniform theory of cellular differentiation and heredity existed at the time, many of the general principles we today take for granted can conceptual be traced back to the late 1800s (more or less substantiated by experimental findings) [100]. These include the theory of chromatin as the hereditary substance [100–102] which, despite its equal distribution in all cells, was hypothesis to enable differentiation of a fertilized egg into all the specialized cell types of the adult organism by undergoing progressive changes in response to variation in external factors (chemical or physical), thereby triggering context-dependent cellular response [103, 104]. Not long after the discovery of chromatin, the concept of chromosome territories was introduced by Rabl and Boveri [105, 106]. Also the conceptual idea that the higher-order organization of the chromatin is important for cell fate determinant was introduced before the 20th century, as indicated by the following quote from Edmund B. Wilsons book *"The cell in development and inheritance"* written in 1896 [100], in which he discuss the theories of Karl W. Nägeli [107].

> *"The [chroamtin]* was conceived as an extremely complex substance consisting of elementary complexes of molecules known as micellce. These are variously grouped to form units of higher orders, which, as development proceeds, determine the development of the adult cells, tissues, and organs. The specific peculiarities of the [chroamtin] are therefore due to the arrangement of the micellae; and this, in its turn, is owing to dynamic properties of the micellae themselves. During development the [chroamtin] undergoes a progressive transformation of its substance, not through any material change, but through dynamic alterations of the conditions of tension and movement of the micellae. These changes in the [chroamtin] cause reactions on the part of surrounding structures leading to definite chemical and plastic changes, i.e. to differentiation and development."*
>
> * Nägeli used the word idioplasm to describe chromatin

Nägeli's theories centered around structural changes directly affecting cellular output, however the concept of nucleus derived formative substances as a mediator of cellular processes (like the later discovered mRNA) was proposed in parallel [103, 108, 109]. While these theories are structurally and mechanistically lacking, and we know now that the smallest subunit of the chromatin (here referred to as micellae) is in fact nucleosomes, is is still quite remarkable that these theories was proposed more than 50 year before the discovery of the DNA structure and before the existence of high resolution microscopes. In the 100 years plus since Wilson's quote, technological advances first into microscopy and later with sequence-based methods for capturing genome conformation, has provided us with a much more detailed understanding of genome topology and its role in gene-regulation. Still, the general concept of a higher-order structure of chromatin as a key factor in the processes that guide cellular identity holds true, and much is still to be discovered regarding the underlying mechanisms.

### Genome topology

Within the nucleus, the genome adopts a hierarchical three-dimensional organization across different size scales. The earliest recognized higher-order structures are *chromosome territories* (CTs), formed by the spatial segregation of chromosomes into distinct nuclear domains during interphase [110]. The distribution of the CTs are not random, instead regions of high gene-density and transcriptional activity tend to occupy a more central location in the nucleus whereas transcriptionally repressed and gene-poor regions more frequency are position closer to the nuclear lamina [111]. Within CTs, chromatin is further partitioned into *compartments* based on chromosome condensation and transcriptional activity, originally retrieved by principal component analysis of a Hi-C derived contact map. The A compartment is comprised of active, accessible euchromatin, whereas the B compartment is characterized by condensed and transcriptionally inactive regions [112]. The segregation of the genome into compartments is not static,

rather, genomic regions can switch between compartments in a context- or stage-specific manner [113, 114]. In B cell development, for instance, lineage-specific genes (e.g. EBF1) transition from B to A compartments to enable transcription, while genes associated with alternative lineages shift in the opposite direction [115, 116]. On a smaller scale, non-random folding of the chromatin into loops result in *topologically associated domains* (TADs), characterized by a high degree of internal interactions [117]. The TAD boundaries are enriched with binding sites for CTCF, an architectural protein that, together with cohesin, facilitates loop extrusion and insulates adjacent domains [118]. By concentrating interactions within their domains, TADs facilitate synchronized gene regulation and limit the impact of external regulatory elements. While originally viewed as stable structural units, TADs has been shown to display dynamic, cell type–specific organization supporting lineage-restricted enhancer activity [119]. The emergence of single-cell conformation capture methods, alongside advances in microscopy-based approaches, has further challenged the notion of TADs as conserved structures. While compartments remain relatively consistent across individual cells, TADs show substantial cell-to-cell variability, leading to a revised understanding of TADs as snapshots of transient chromatin interactions rather than rigid representations of folding patterns in single cells. Although TADs may better reflect a probabilistic interaction domains rather than rigid structures, they still provide valuable information about the relative contribution of chromatin interactions a cell undergoes during interphase. There is still much to be learned about how genome topology shape transitional program, however, *enhancer-promoter (E-P) loops* is known to be of great importance in this process, by enabling regular activities of distal regulatory elements via

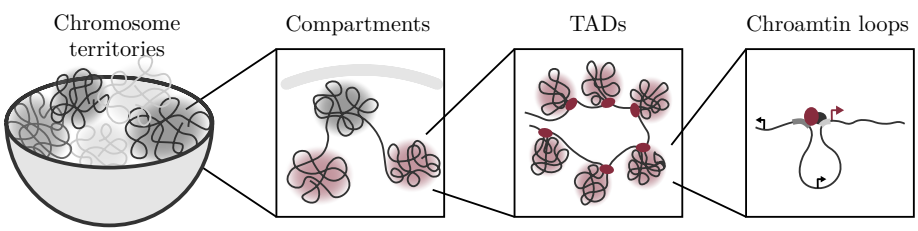



**Figure 1.5: Hierarchical representation of 3D genome topology.** The genome is organized into a complex non-random organization at different size scales. The largest structure are chromosome territories (CTs) occupying distinct spaces in the nucleus. The CTs are partitioned into compartments based on chromatin condensation and transcriptional activity, which in turn can be subdivided into clusters of chromatin interactions known as topologically associated domains (TADs). The smallest structure are represented by individual chromatin loop that can link distal regulatory elements to their target genes.

physical proximity in the 3D space. The mechanism of regulatory elements and how they can impact gene-expression of (often distally located) target genes is the focus of next section. For an overview of the hierarchical 3D genome organization, see Figure 1.5.

### The cis-regulatory landscape

While protein-coding sequences comprise only ∼1.2% of the genome, regulatory elements embedded in the non-coding portion of the genome represent a much larger fraction and is highly relevant for gene regulation [120]. These regions, referred to as *cis-acting regulatory elements* (CREs), can mediate spatiotemporal specificity of gene-regulation in response to context specific activity of trans-acting factors (e.g. transcription factors) [93]. This specificity is encoded in the genome sequence, both within the element itself (e.g. combinatorial TF binding sites) and the relative position of the CRE compared to genes and other regulatory element in the linear genome. CREs can be classified based on function and location relative to genes. *Promoters*, located at the 5' end of genes, severe as the binding site for the transcriptional machinery, whereas *enhancers* and silencer are distal elements that exert positive or negative effects on gene expression, respectively [93]. However as the direction of transcriptional change is context dependent, the term enhancers is often used to describe distal CREs irrespective of activating or repressing activity. In addition, *insulators* are a distinct class of CREs that function primarily as boundary elements of regulatory domains, and are characterized by the presence of CTCF binding sites [121, 122].

With the median enhancer–target gene distance estimated to be on the order of 100 kb [123–126], a central question emerges: by what mechanisms do enhancers exert their regulatory activity over such long genomic distances? Initially, one-dimensional models were proposed, suggesting that transcriptional machinery or chromatin remodellers scan along the linear genome to locate target promoters [127, 128]. However, these models could not account for phenomena such as promoter skipping [129–131] and were therefore largely abandoned in favor of 3D models, in which enhancer activity is explained by physical proximity within the 3D space. Broadly the 3D models of E-P interactions can be subdivided into direct models and action-at-a-distance model [132]. In the contact models the E-P loops is mediated by a direct connection between the CRE and target promoter, either via cohesion mediated loop extrusion and/or other architecture proteins inducing the mediator complex [133] and YY1 [134], all of which have been implicated in *IgH* locus contraction during recombination [135, 136]. Also transcription factors, with or without co-factors has been proposed to mediate E-P loops [93]. The action-at-a-distance model also relies on spatial proximity within the 3D genome, but at a larger scale (on the order of hundreds

rather than tens of nanometers), with the formation of *condensates* proposed as the leading mechanistic explanation [132]. Condensates are macromolecule rich structure, generated via phase-separation, a phenomenon in which a solution separates into two distinct phases when a certain threshold concentration is reached due to the intermolecular interactions being energetically favored over solute–macromolecule interactions [137]. It has been suggested that condensates can assemble into hubs, allowing for the coordinated control of multiple genes and the integration of regulatory inputs from several CREs onto a single target gene [138, 139]. Extensive evidence supports the existence of chromatin loops connecting validated CREs with their target genes, in some cases spanning extreme distances as for the regulation of *Myc* by the BENC enhancer located ∼1.7 Mbp from transcription start site (TSS) [140]. Still limited correlation between between E-P distance and nascent transcription has been observed [141, 142]. A proposed mechanistic explanations for this has been a hit-and-run model in which transient contacts trigger a delayed transcriptional response [132] or gradient-based diffusion model [143]. It is also possible that the E-P served as priming mechanism enabling rapid transcriptional response upon additional input, or that weak correlation simply reflects the failure to account for combinatorial CRE activity [144].

Contact models aside, a number of mechanisms have been proposed to explain the actual transcription modulating activity of enhancers, including transfer of enhancer recruited proteins to promoters, recruitment upon contact or direct impact on gene-expression via modulation of chromatin state. For condensate-based model, the establishment of a local increase in density of regulatory proteins has also been proposed as a possible mechanism [132]. Enhancers and promoters are typically treated as distinct entities, yet chromatin capture data demonstrate widespread promoter–promoter contacts [145, 146], and experimental evidence confirm the ability of promoters to act in a enhancer-like manner [147], highlighting the flexible nature of CRE activity. The observation that ∼50% of enhancer bypass their closes gene [129, 130], highlight the importance of enchanter selectivity. This selectivity has been proposed to arise from a number of mechanisms, including chromatin-state barriers, sequence compatibility, protein interactions and spatial constraints within condensates [132]. To add further complexity, these modes of selectivity are not mutually exclusive and can act combinatorial and quantitatively to achieve fine-tuned regulation of target gene expression. The expression of genes is not necessarily a continuous process, rather transcription has been observed to occur in distinct bursts [148, 149]. This *transcriptional bursting* contributes to cellular heterogeneity and has been implicated to be a mediator of rapid cellular responses and a driver for developmental transitions [150, 151]. The transient hit-and-run mechanism

for E-P contact have been proposed as a potential explanation for this non-continuous behavior [152]. Although the primary focus on CRE is transcriptional activation, repressive mechanism are just as important, especially for the establishment of lineage restriction. Repression can be achieved via numerous mechanisms, including blocking, interference, non-coding RNAs, recruitment of repressor complexes or modulating of chromatin state [153]. In a addition, target gene repression can be mediated by imposing structural constrains for gene activation, including regulation of insulator distribution or E-P looping (e.g. via interference or diversion) [153, 154]. The extent to which the various models outlined here contribute, and how they integrate to establish the cis-regulatory landscape and gene regulatory programs, remains to be determined. Deciphering this cis-regulatory code, i.e. predicting transcriptional output from the CRE landscape, remains a key frontier in molecular biology, with implication in both understanding of normal developmental and disease.

### Local chromatin states

The potential of CREs to exert gene regulatory functions is determined by composition, post-translational modifications and the topological organization of the nucleosomes within the genomic regions, as well as modifications of the DNA itself [155]. These factors, collectively known as the chromatin state, determine how permissible the element is to regulatory proteins, which subsequence determine the transcriptional output [156]. *Chromatin accessibility*, the degree of physically access to chromatininized DNA, is determined by the occupancy, density and organization of the nucleosomes [155]. Although only ∼2–3% of the genome is considered "open", these regions represent 90% of the TF bound regions [157], highlighting the importance of chromatin accessibility in CRE selectivity. However, this relationship is not unidirectional, trans-acting factors can actively modulate chromatin accessibility, either by promoting nucleosome eviction [158] or through the recruitment of chromatin-remodeling complexes [159]. One chromatin remodeller that has a central role in Paper I, is the an ATP-dependent BRG1/BRM Associated Factors (BAF) complex [160–165], which employs ATP hydrolysis to reposition the nucleosomes [165]. The canonical BAF complex is composed of a core set of subunits, including an ATPase (SMARCA2/4) and an DNA-binding AT-rich interactive domain containing protein (ARID1a/b) [166].

The plethora of post-translational modifications (PTMs) deposited on histone proteins, often referred to as *histone marks*, constitute another layer of chromatin regulation. These modifications are intimately linked to chromatin accessibility and transcriptional activity, with specific patterns associated with distinct classes of regulatory elements and transcriptional states (e.g., active versus repressed chromatin). Histone modifications have been proposed to act as a dynamic "epigenetic code" [167], which can be written, read, and erased by a diverse set of histone-modifying enzymes [168] (see Table 1.1). Histone modifications can play instructive roles in gene regulation, either by directly inducing structural alterations of chromatin, indirectly through the recruitment of chromatin remodellers or effector proteins, or alternatively by being deposited as a downstream consequence of other regulatory processes [169]. Tri-methylation of histone 3 lysine 36 (H3K36me3) is associated with gene bodies whereas the same modification of lysine 4 (H3K4me3) is associated with active promoters. Although H3K4me3 has been shown to be able to recruit the transcriptional machinery [170], its instructive effect seems to be modest [171] and alternatives functions such as reinforcing transcription or reduce noise has been proposed [172–174]. Another histone mark associated with active chromatin is acetylation of histone H3 at lysine 27 (H3K27ac). In contrast to H3K4me3, which is largely restricted to active promoters, H3K27ac is also enriched at active enhancer elements, thereby serving as a key signature of enhancer activity [175]. The same lysine residue can also be modified with tri-methylation (H2K27me3) which, together with the closely linked monoubiquitylation of histone 2A lysine 119 (H2AK119ub), instead mark repressed elements and plays a crucial role in maintaining cellular identity during development [176]. H3K27me3 is deposited by the Polycomb Repressive Complex 2 (PRC2), composed of core subunits EZH1/2, EED, SUZ12, and RBBP4/7, with EZH2 providing the primary catalytic activity [177]. EED in turn is responsible for the propagation of the repressive mark, trough allosteric activation of EZH2 when bound to tri-

**Table 1.1:** Selected histone modifications, along with their writers and erasers. In addition, the type of elements each histone marks is associated with is presented.

| **Modification** | **Mark** | **Element** | **Writer** | **Eraser** |
|---|---|---|---|---|
| Acetylation | H3K27ac | Active elements | KAT3A/B | HDAC1/2 |
| Methylation | H3K4me3 | Active promoters | SET1, KMT2A/B | NO66, JARID1 |
| | H3K36me3 | Gene bodies | SETD2 | JHDM3 |
| | H3K27me3 | Repressed elements | EZH1/2 | UTX/Y, JMJD3 |
| | H3K9me3 | Heterochroamtin | SUV39H1/2 | JHDM3 |
| Ubiquitination | H2AK119ub | Repressed elements | Ring1A/B | BAP1 |

methylated H3K27 [178]. During development, many genes display a *bivalent chromatin* state marked by both H3K4me3 and H3K27me3, keeping key regulatory loci poised for activation or repression during lineage commitment [179]. In contrast to the developmentally regulated repression marked by H3K27me3, tri-methylation of histone H3 at lysine 9 (H3K9me3) is associated with stably silenced heterochromatic regions [180]. This modification plays an instructive role in chromatin compaction through the recruitment of HP1 [181, 182], which has been proposed to self-associate and drive phase separation [183–185].

### 1.2.2 Transcription factors: The directors of cellular identity

#### Mechanisms of action

The concept of transcription factors (TFs) can be traced back to the discovery of shared DNA sequences in the promoter regions of genes with a similar expression pattern [90, 91, 186]. These short sequences (i.e. motifs) were shown to be selectively targeted by DNA binding proteins with gene regulatory potential, the TFs [91]. TF sequence-specificity is mediated by a DNA binding domain, often used to classify TFs into families [187–190]. Although TF mediated gene-regulation can provide cell-type specific expression via regulation of TF expression or activity, this alone can not explain the full complexity of gene regulatory networks (GRN). Indeed, since then it has been shown that the TF recognition is guided by a complex *motif grammar*, defined not only by the DNA sequence by also the nucleosome topology [191] and the pattern of combinatorial TF binding sites with varying affinities [192, 193]. As most TFs are restricted to DNA binding in open chromatin environments, the epigenetic state of the cell provide an additional layer of regulation on top of motif sequence and organization, restraining TF activity to epigenetically primed regions in the genome [92]. However, a group of TFs, known as *pioneer factors*, can bypass this barrier by their intrinsic ability to bind and induce nucleosome eviction at sites of closed chromatin regions [158], which can be extended to entire elements via recruitment of chromatin remodeling complexes (e.g. the BAF complex) [194, 195]. Although a continuum of pioneering activity dependent on chromatin state, binding site affinity and TF concentration is probably a more accurate description than a strict binary classification [195], the fact that many of the key drivers of lineage specific programs poses pioneering activity substantiate the importance of pioneer factors is development [194, 196–198].

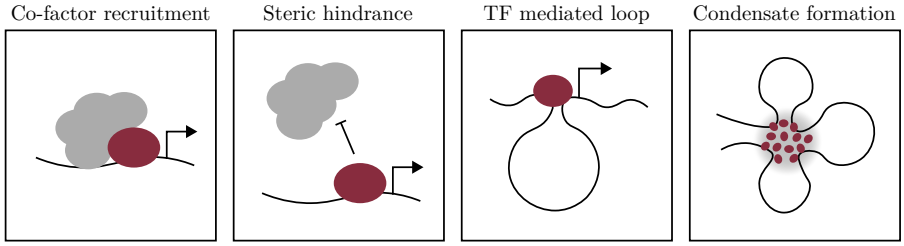



**Figure 1.6: Mechanisms of action for transcription factors.** Transcription factors can exert gene regulatory function via numerous mechanisms, including steric hindrance, direct recruitment of the transcriptional machinery or through recruitment of co-factors or chromatin remodellers. In addition, TFs can impact gene expression by modulation of the genome 3D structure, by direct contact or condensate formation.

The mechanisms by which transcription factors regulate their target genes are diverse, reflecting both their inherent properties and the cellular context. While some TFs function by direct recruitment of the transcriptional machinery [199] or steric hindrance [200], most eukaryotic TFs exert their gene regulatory function by *co-factor recruitment* [201]. The ability of TFs to interact with a wide array of co-activator and repressor complexes enable context specific TF activity. Although TFs were initially discovered to bind within the promoter region of their target genes, it is now known that the majority of TF responsive element are located in distal regulatory elements often located tens to hundreds of kilobases away from the target promoter [123–126, 131]. The ability of TFs to act over long genomic distance can be achieved via the formation of 3D chromatin loops bringing distal elements into physical proximity of their target promoter [202], either by direct interaction or via condensates formation [93]. The switch-like dynamics of condensates has positioned it as a attractive model to explain the rapid state transitions observed in many biological processes [203, 204]. An overview of the diverse mechanisms of TF action is provided in Figure 1.6.

### Transcription factors in lymphocyte development

The developmental programs that guide the differentiation from multipotent progenitors into lymphoid primed progenitors and later fully commit B- and T cell involve the coordinated action of a number of TFs. Among these are PU.1, a broadly expressed TF with pioneering activity that can orchestrate different lineage fates in a dose dependent manner [205]. Also RUNX1 exhibit a multi lineage expression pattern, and can mediate different cell-type specific activity based on interaction partner [206, 207], and together with IKZF1 [208] and TCF3 (also known as E2A) [209], is important for lymphocyte priming in

multi-protein progenitors. The key TFs involved in the B- and T cell specific programs are introduced below:

***B lineage specific TFs***

**FOXO1:** Forkhead box O1 (FOXO1), activated by E-proteins in early B cell development [210], is a critical regulator driver of the early B cell program [211, 212]. FOXO1 control the expression of the *Rag* genes [211] as well as engaging in a feed-forward loop with TCF3 to drive the expression of the key B cell factor EBF1 [213].

**EBF1:** Although *Ebf1* (Early B cell factor 1) has a rather broad tissue expression profile, in hematopoietic system it is restricted to the B cell lineage [214–217], and first detected in the BLP stage [63, 218]. EBF1 is a key factor in the activation of the B linage genes, including *Pax5*. Mechanistically, EBF1 can act as a pioneer factor [197, 219, 220] and has also been promote condensate via recruitment of chromatin remodelling complexes [221]

**PAX5:** Paired box gene 5 (PAX5) is an important factor in the establishment of B cell identity, as indicated by the alternative lineage potential observed in *Pax5* deficient pro-Bs [222–227]. Together with EBF1, PAX5 is a major contributor to the establishment of the B cell transcriptional program as well as the repression of alternative lineage genes [228, 229].

***T lineage specific TFs***

**TCF7:** TCF7 (also known as TCF1) is activated by Notch signaling and is a key regulator of T cell development [230–232]. Mechanistically, TCF7 drive the early T cell program, exemplified by its ability to activate a *Gata3*, and *Bcl11b* [231]. The importance of TCF7 in T cell development is further substantiated by the finding that *Tcf7* deficiency cause differentiation defects in early T cell development upon migration to the thymus [230, 231].

**GATA3:** GATA3 belongs the GATA family of zinc finger transcription factors. Despite its relatively broad expression pattern throughout T cell development [233, 234] as well as in other tissues [235–237], GATA3 provide stage specific regulation of T cell development [80, 238]. In addition to it role in initiation of the T cell specification of thymic progenitors [234, 239], GATA3 is also involved ILC differentiation and in the promoting a CD4+ helper cell fate over CD8+ fate [234, 238].

**BCL11B:** The activation of BCL11B at the DN2a-to-DN2b transition correlate with T lineage commitment. This is consistent with the important role of BCL11B in promoting T lineage over alternative fates, as well as to drive the Phase 2 program [240–243]. BCL11B has been shown to exert its regulatory potential both directly via requirement of chromatin remodellers and histone modifiers, as well as via TF circuitry [243–245].

### Transcription factor networks

Similar to TF–cofactor interactions, TF–TF crosstalk provides an additional layer of gene regulatory control, mediated either through direct physical contact or indirectly at the transcriptional or chromatin level [246]. It has been shown that *DNA-mediated cooperativity* (via composite motifs) can stabilize DNA binding [247] and impose lineage-specificity on broadly expressed TFs, as illustrated by the lineage specific occupancy pattern of ETS1 [248, 249]. TF–TF interactions are not always synergistic, and alternative modes of cross-talk include co-factor [246] or *binding site competition* [250], as well as *depletion via redirection*, exemplified by PU.1 during T cell development [207].

The gene regulatory potential of TF enable indirect interactions via transcription factor networks. These networks can be stratified into classes of small regulatory circuits. While such sub-networks are often referred to as motifs in network theory, the term circuits will be used here to avoid confusion with DNA sequence motifs. Feed-back loops (FBL) and feed-forward loops (FFL) represent two groups of well characterized circuits with relevance in TF networks [251]. *Feed-forward loops* are three node circuit in which a factor (X) activate a target (Z) both direly and via an intermediate (Y). Depending on whether the direct and indirect regulatory interactions reinforce or counteract one another, the FFLs can be further divided into coherent (C-FFL) or incoherent (I-FFL) subtypes [251]. Coherent FFLs act as persistence detectors, and noise filter, a property that can be be beneficial in the earlier stages of cell fate decisions as it ensures that only persistent signals trigger lineage-specific programs. Example of this type of circuit can be found in both early B cell development with an TCF3-FOXO-EBF1 loop [210, 213, 252] and in T cell development with a Notch-TCF7-GATA3 loop [230, 231, 253]. Incoherent FFL on the other hand can generate transient pulse like expression or delayed timing of transcriptional programs as observed for the initiation of T cell effector and innate programs in T cell development, provided by the opposing role of Notch signaling on the regulation of TCF7/GATA3 and several of their target genes [253]. AND/OR logic within in these feed-forward circuits can provide additional complexity to the transcriptional output [251]. The other class of circuits, the *feed-back loops* are most prominently exemplified by the classical negative FBL providing stability

to biological systems by tuning of expression levels of transcriptional programs and buffering fluctuations. In lymphoid development exemplified by the activation of ID3 and IKZF by PAX5 [254, 255], which in turn suppress the early B cell transcriptional program (via E2A and EBF1 respectively), leading to reduced pre-BCR signaling and subsequent control of proliferation [256–259]. In contrast, positive feedback loops reinforce signaling and generate bistability, as seen in the reciprocal activation between EBF1 and PAX5 which ensure stable B lineage commitment [260–262]. Figure 1.7 summarizes the different classes of regulatory circuits and illustrates them with examples from early lymphoid development. Although isolation of these small sub-networks facilitate the conceptual understanding of the underlying processes, in reality the circuits are highly intertwined in complex TF network [263], enabling precise and temporally controlled expression of lineage-specifying programs.

Transcription factor activity can also be indirectly modulated through *TF mediated chromatin remodeling*, which alters accessibility for subsequent TF binding. This can be mediated via pioneering activity or recruitment of chromatin remdodellers or histone modifying complexes [246]. The key B cell transcription factor EBF1 provide an en example of pioneering TF that also has been shown to impact the chromatin level via BRG1 recruitment [197, 219–221] and phase-separation [221]. In addition, TFs can influence the activity of other TFs through *topological constraints* imposed by chromatin looping. An indirect example is provided by PAX5, which facilitates *Igh* loop extrusion and V gene recombination by PRC2-mediated repression of *Wapl*.

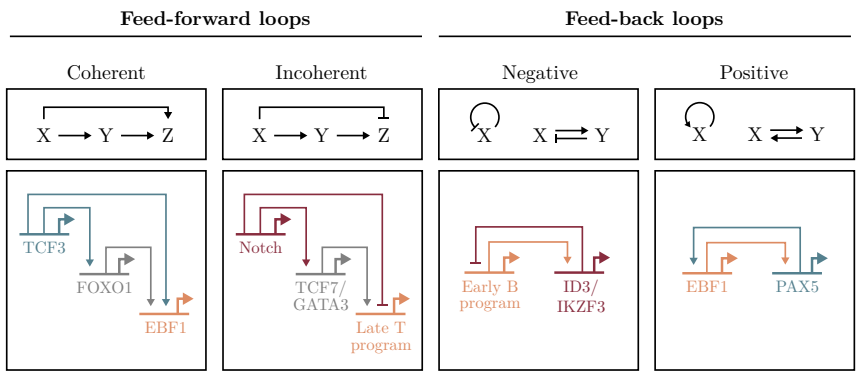



**Figure 1.7: Regulatory circuits in TF networks.** Conceptual representation of selected regulatory circuits, with example in lymphoid development. The depicted circuits are coherent/incoherent feed-forward as well as positive/negative feed-back loops.

### 1.2.3 Signaling pathways: A cellular dialogue

The mechanism covered so far in this section are cell intrinsic, however, the cells does not exist in isolation but instead occupy a distinct niche which provide the cell with cell-cell interaction and external factors (e.g. cytokines). These factors trigger signaling cascades which integrate with the cell intrinsic regulatory mechanisms in order to drive cellular processes, such as migration, proliferation and differentiation. Two of these signaling pathways with relevance to this these are described below:

#### IL-7 signaling

IL-7 is a soluble globular protein produced by stromal cells in the BM and thymus [264], that can trigger IL-7 signaling by binding to the heterodimeric IL-7 receptor consisting of an $\alpha$-chain (IL-7R$\alpha$, CD127) paired with common cytokine receptor $\gamma$-chain ($\gamma$c, encoded by the *Il2rg*) [265]. Binding of IL-7 cause a conformational change in the receptor, trigger Janus kinase 1 (JAK1) and 3 (JAK3) mediated phosphorylation of IL-7R$\alpha$, and subsequent activation of signal transducer and activator of transcription (STAT) proteins. Phosphorylation of STAT cause homodimerizes and translocation to the nucleus where it can activate the expression of target genes [266]. IL-7 signaling is critical for lymphopoiesis and deficiency cause severe impairment in both B- and T cell lineages, including developmental block in the transition to the pro-B cells stage of B cell development [267, 268], although the dependence of IL-7 for B cell development seems to be less pronounced in the human setting [269]. It has been suggested that IL-7 signaling play an instructive role in B cell development by activating *Ebf1* in the common lymphoid state [270–272], which is supported by the finding that ectopic expression of EBF1 is sufficient to partially rescue B cell development in IL-7 deficient mice [270]. In addition, IL-7 signaling act in conjunction with the pre-BCR to stimulate a proliferative burst [64] and reduce RAG levels [65]. IL-7 signaling has also been shown to be important for early T cell development in the thymus, supporting the survival and expansion of thymocyte progenitors [267, 273, 274].

#### Notch signaling

Notch signaling is a highly conserved signaling pathway with a central role in the regulation diverse cell fate decisions [275]. In the hematopoietic system, the most prominent effect of Notch signaling is observed in the T cell lineage, where it is indispensable for initiating the T cell transcriptional program [276]. The signaling cascade is initiated by binding of the trans-membrane Notch receptor to a delta-like ligand, which trigger proteolytic cleavages of the receptor, and release of the Notch intracellular domain (NICD). NICD translocate to the nu-

cleus where it, via interaction with DNA-binding protein RBP-J$\kappa$ (also known as CSL) and co-activators, drive transcription of target genes [275, 277]. In T cell development, multipotent progenitors are exposed to high levels of Notch signaling via delta-ligand expressing stromal cells in the thymic niche [73, 74], causing the activation of several genes with critical functions in the T cell developmental progression (e.g. *Tcf7* and *Gata3*) [276]. The essential role for Notch signaling i T cell development is supported by the arrest early T cell development and intrathymic accumulation of B cells in the absence of Notch-mediated signals [278, 279], as well as the ability of stromal expressing delta-ligand 1 (Op9DL1) to recapitulate T cell development *in vitro* [261].

## 1.3 Cell fate determination: A three part act about identity and commitment

Cell fate determination, the ability of cells to acquire distinct cellular identities through the process of differentiation, is central to all multicellular life. The most apparent example being the ability of a single cell (the zygote) to give rise to all cell types in the fully formed organism. The concept of cell fate decisions is not only applicable to embryonic development, but is also relevant in the context of differentiation of adult stem and progenitor cells into more specialized cells. A conceptual representation of this process was proposed by Waddington in 1939 [280, 281], referred to as the Waddington landscape, in which differentiating cells were visualized as balls rolling down a landscape of bifurcating valleys, representing alternative developmental trajectories (see Figure 1.8). Extrinsic signals from the microenvironment, together with intrinsic gene regulatory programs, guide cells along these paths, establishing and maintaining cellular identity. Although the discovery of induced pluripotent stem cells (IPS) [282] and direct reprogramming [283] has challenged certain aspects of the Waddington landscape, most notably its implied uni-directionality and the inaccessibility of certain states, the framework continues to serve as a valuable conceptual model of cell fate determination. At the molecular level, differentiation is orchestrated by a network of more or less lineage-specific transcription factors acting in concert with a dynamic chromatin landscape and extrinsic signals, in order to establish lineage-specific transcriptional programs and repress alternative fates. This process is continuous and dynamic, encompassing intermediate states with varying degrees of plasticity, ensuring a balance between adaptability and robustness. The mechanistic basis of cell fate determination within the lymphoid lineage of hematopoiesis is detailed below.

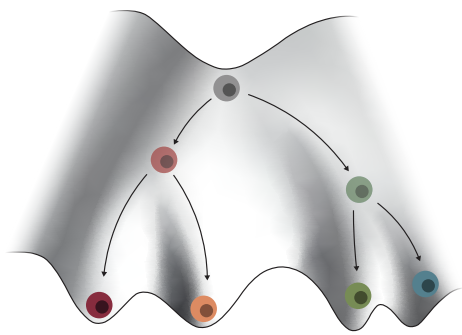


**Figure 1.8: The Waddington landscape.** Conceptual representation of the epigenetic landscape that guide cell fate determination.

### 1.3.1 Opening act: Lymphoid priming

In the initial stages of hematopoiesis, the B- and T cells share a common trajectory, diverging *Gata1* expressing megakaryocytic-erythroid progenitors away from a shared myeloid/lymphoid population characterized by FLT3 surface expression and high PU.1 (encoded by *Spi1*) levels [12, 284]. PU.1 is also a critical regulator for the subsequent bifurcation into myeloid/lymphoid lineage, where high PU.1 levels achieved via elongation of cell cycle duration [205, 285] favor myeloid differentiating, while a low expression of *Spi1* drive cells into a lineage restricted stage. This common lymphoid progenitor population is marked by the expression of IL-7R [286], although there is not clear evidence pointing toward an instructive role of IL-7 signaling at this divergence [271]. Instead, IKZF1 serve a key role in early lymphoid restriction, by regulating PU.1 levels via activation of the *Spi1* repressor GFI1 [287], as well as to create a permissive chromatin landscape for lymphoid development by recruitment of the nucleosome-remodeling MI-2BETA and histone deacetylase (NuRD) to target genes [288]. The priming for lymphoid transcriptional program is also achieved on the level of genome topology, with IKZF1 mediated chromatin interaction driving the compartment switch of lymphoid-specific genes into a permissive euchromatin state to facilitate the progression of lymphoid development [289].

### 1.3.2 Middle act: B- and T lineage specification

IKZF1 is also involved in B cell lineage specification, and appears to collaborate with TCF3 [209] to activate lymphoid-restricted genes (e.g. *Rag1*, *Dntt*) [209, 290–292]. TCF3 (also known as E2A) in turn establishes a FFL with FOXO1 [210], culminating in the activation of EBF1 at the transition to the BLP2 stage [57, 213]. EBF1, a key regulator of the B development, orchestrate B cell specification through activation of a set of B cell genes [63, 218]. Given that initiation of the B-lymphoid program is associated with structural alteration in genome organization and changes in chromatin states of B-lineage associated elements [115], the significance of EBF1 in B cell specification may, at least in part, be attributed to its ability to modify the epigenetic landscape [197, 219–221]. Mechanistically, this can be achieved by EBF1's pioneering activity and ability to promote phase-separated structures at B-lineage target genes via interaction with chromatin-remodeling complexes (e.g. BRG1) [197, 219–221]. Among EBF1 target are genes encoding components of the pre-BCR, whose assembly represent an import step in early B cell development and marks the pro-B to pre-B transition [254, 293–298]. The importance of EBF1 in B cell development is substantiated by the ability of endogenous EBF1 expression to activate most of the early B-lineage program [227, 299], and the fact that loss of EBF1 result

in cellular plasticity and de-differentiation [300]. However, despite this crucial role of EBF1 in B lineage specification, commitment to the B lineage require the action of another key B lineage restricted TF, PAX5 [222, 224, 227].

T cell specification is instead dependent on migrating of lymphoid primed progenitor with varying degree of lymphoid commitment to the thymus [69, 72, 301], where the cells converge on the T cell developmental trajectory with stage dependent kinetics [70, 71]. Notch signaling, mediated by delta-ligand expressing thymic stromal cells, initiate the T lineage specification program which initially is predominantly influenced by the action of bone marrow inherited TFs [302]. Although these TFs are permitted to promote proliferation and expression of progenitor genes, Notch signaling selectively modulate their activity for compatibility with a T cell fate. A critical example being PU.1, which are allowed to use its pioneering activity for T lineage priming [303], while its ability to promote myeloid differentiation is suppressed by Notch mediated inhibition of its obligatory myeloid partners of the C/EBP and IRF families [302, 304–306]. Notch signaling also induce the expression of the lineage restricted TFs TCF7 and GATA3 [307], essential for initiation of the T cell identity program [230, 306, 308–312]. Among the T-lineage signature genes, many rely on non-redundant input from multiple regulators for activation, highlighting the complex regulatory network that drive T cell specification [302]. Observation of changes in 3D architecture and local chromatin state in regions associated with some of these T cell restricted genes [232, 313–316] similar to what has been observed in B cell development, suggesting that modulation of the chromatin landscape is an important regulatory mechanism for lineage specifying program in general [302].

### 1.3.3 Final act: Commitment & beyond

Quite paradoxically, many of the factors driving the early Phase 1 T cell or even progenitors programs are also involved in driving the functionally important DN2a-to-DN2b transition, which marks stable T lineage commitment. This include the E2A:HEB E protein dimers which activate a set of signature T cell genes including *Rag1* [317, 318] and CD3 TCR co-receptor components [319]. Also RUNX1 and IKZF1 show multifaceted and stage specific activities [320, 321]. During the Phase 1 to Phase 2 transition, which correlate with commitment, RUNX factors undergo genome-wise relocation from sites enriched for co-binding with PU.1 to new sites enriched for co-binding with TCF7, E protein as well as the novel TF BCL11B [207, 243, 253, 321], whose expression concise with functional commitment at a single-cell level [79]. The delayed activation of *Bcl11b* is to a large extent caused by epigenetic constraint [313, 314, 322, 323]. Although the regulators of *Bcl11b* (RUNX1, TCF7, GATA3), are

all expressed already at DN1, their regulatory potential for *Bcl11b* is initially limited by its regulatory elements being marked by repressive H3K27me3 and sequestered in inactive compartment B [253, 314, 321, 324]. As enforced expression of *Runx1* is sufficient to overcome this chromatin barrier, enable RUNX1 to bind high-affinity binding sites and subsequently turn on *Bcl11b* expression at an earlier stage, TF dose titration is most likely a contributing mechanisms [253]. BCL11B itself has a dual role in the promoting T cell identity, it blocks NK and ILC differentiation by repressing *Id2* and other innate regulators to ensure stable T cell commitment while simultaneously drive the progression of T lineage program via activation of signature genes [243, 325]. In contrast to Phase 1 regulators which often act in an incoherent manner to ensure robustness and timing, the Phase 2 program is characterized by more coherent TF circuits enforcing T cell identity [302]

B cell commitment, on the other hand, is dependent on the PAX5, as demonstrated by the ability of *Pax5* deficient B cell progenitors to differentiate into both myeloid and T cells [222–227, 326]. Similar to the activity of BCL11B in T cell commitment, PAX5 impose B cell lineage restriction by activation of B signature genes while simultaneously imposing lineage restriction by repression of alternative lineage genes (e.g. colony-stimulating factor receptor 1 (Csf1r or cfms)[228] and Notch1 [327]). This dichotomous role, which is also observed for IKZF1 and EBF1, can at least partly be explained by the TFs ability to interact with both activator and repressor complexes, including PRC2 [220, 221, 328–332]. Among these repressed PAX5 target genes are the gene encoding then Cohesin release factor WAPL, enabling the recombination of distal IgH V-genes by facilitating increased DNA loop extension [331]. In addition, the ability of PAX5 to activate *Ebf1*, establishes a positive feed-back loop between the two TFs [333], which enforces B cell identity and drive the progression of B cell development into the pro-B cell stage and beyond. These two factors also play an important function regulation of assembly and signaling trough the pre-BCR. While EBF1, in collaboration with TCF3, activate critical component for pre-BCR assembly at the large pre-B stage, including signal transducers CD79$\alpha/\beta$ and surrogate light chains IGLL1 ($\lambda$5) and V-set pre-B cell surrogate light chain 1 (VPREB1) [294, 295], PAX5 ensure that the subsequently triggered proliferative burst is suppressed to enable light chain recombination [334, 335]. Mechanistically, PAX5 drive cell cycle exit by activation of the genes encoding IKZF3 and ID3, resulting in reduced expression of surrogate light chain genes by EBF1 binding site competition [257] and impaired TCF3 activity [256, 258, 259], respectively. The functional placticity of B cell progenitors with combined heterozygous deletion of *Ebf1* and *Pax5* (*Ebf1*$^{+/-}$*Pax5*$^{+/-}$), not observed for single heterozygous deletion [333] provide further support for the important

non-redundant functions of EBF1 and PAX5 in B development, and in addition highlight the importance of TF dose in linage commitment as observed also for T cell development. For an overview of transcription factor network controlling priming, specification and commitment of B- and T lymphocyte, see Figure 1.9.

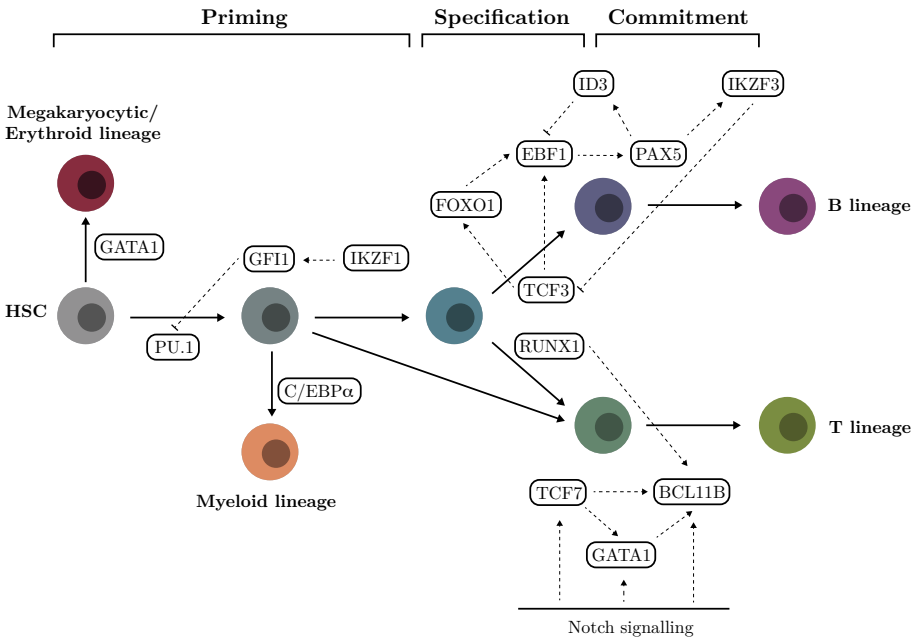



**Figure 1.9: Transcription factors network in lymphoid cell fate determination.** Schematic representation of the transcription factor network that control lymphoid priming, B- and T lineage specification and finally commitment.

## 1.4 Going off script: From normal development to leukemic transformation

### 1.4.1 Leukemic transformation

While gene regulatory networks ensure fine-tuned regulation of cellular identity during normal development, their interconnected structure also makes them sensitive to disruptions, as confirmed by the dose-dependent effects seen in B- and T cell development. Collectively, this sensitivity, combined with cellular plasticity, and rapid proliferative capacity renders developmental TF circuits prone to oncogenic hijacking and malignant transformation [336]. Indeed, partial loss-of-function mutations, including heterozygous loss, of key regulatory TFs are frequency detected in B-lineage leukemia [337, 338], often in a combinatorial pattern and in combination with fusion proteins (e.g. BCR-ABL or ETV6-RUNX1) [330, 337–339]. Although the developmental block often observed for B cell acute lymphoblastic leukemia (B-ALL) has been suggested to be the cause of the leukemic transformation [340–342], in appears that disrupted stage specific control of the cells basic machinery (e.g. metabolism and proliferation) is the most prominent driver [343]. This principle is illustrated by IKZF1 and PAX5, which not only guide B cell identity but also regulate glucose metabolism, thereby linking developmental programs with control of cellular energy demands. [344, 345]. In T cell lymphoblastic leukemia (T-ALL), especially the high risk ETP-ALL subtype, many of the key regulators can act as proto-oncogene if the regulatory mechanisms controlling stage specific activity is disrupted [346–353]. The dysregualtion of TFs can also be mediated by mutations targeting non-coding regulatory elements, as seen for the accelerated development of T-ALL via loss of distal *Pten* enhancer [354], or focal amplification of non-coding region driving lineage-ambiguous leukemia by inducing aberrant BCL11B expression [355].

The signaling pathways that normally drive cell expansion under tight control of the TF networks can become oncogenic, if targeted by mutations that disrupt its regulation. This is most clearly exemplified by the frequency detection of Notch mutations in T-ALL, causing ligand-independent hyperactivation and prolonged Notch driven proliferation [302]. Similarity, in B-ALL, the IL-7/stats signaling pathway can turn oncogenic by gain-of-function mutations in genes of the JAK/STAT proteins or in the downstream RAS signaling pathway [356, 357]. These mutations has been shown to synergize with heterozygous mutations in *Ebf1* and *Pax5*, thereby reducing the normal control these TFs have on JAK–STAT signaling [358–360]. Development is accompanied by profound changes in genome architecture and dynamic remodeling of chromatin states, represent-

ing another regulatory axis vulnerable to leukemic transformation. Alterations affecting epigenetic regulators are frequently detected in acute lymphoblastic leukemia (ALL), most notably exemplified by KMT2A (MLL) rearrangements, where fusion of TFs to the histone methyltransferase leads to a decoupling of normal transcriptional programs [361].

### 1.4.2 Lineage placticity in leukemia

A potential consequence of disruption of the gene regulatory networks are lineage placticity. Indeed, the EPT-ALL subtype, often marked by broad retention of a Phase 1 transcriptional state alongside expression of post-commitment associated TCR complex components, retain some myeloid potential. This suggest that leukemic transformation is associated with failed repression or retrograde differentiation [302]. In B-ALL, lineage plasticity is especially pronounced in MLL-rearranged cases [362–365], where relapse is frequently associated with conversion to a myeloid–monocytic phenotype [365–367]. This lineage switch is marked by silencing of the signature B TFs *Ebf1* and *Pax5*, suggested to be mediated by epigenetic dysregulation [366, 368]. Lineage plasticity has also been observed in both normal and malignant B cell progenitors with trans heterozygous deletion of *Ebf1* and *Pax5* in mice [333, 369], although there is little support for loss-of-function mutation in these genes to cause lineage placticity in the human B-ALL [336]. While lineage conversion after conventional therapy is rather rare, a shifts in tumor cell surface antigens is more common in patients treated with targeted therapies (e.g. chimeric antigen receptor (CAR)-T cells). Even though this epitope masking seen in relapse after CD19 CAR-T treatment can be attributed to mutations in the *Cd19* gene itself rather then a complete lineage switch [370–372], the ability of these cells to escape treatment or avoid detection still poses a serious challenge.

# Chapter 2

# Methodology

## 2.1 Experimental models

This thesis is based on data generated from a combination of primary cells from mice models as well as cell lines. The majority of the work have been conducted in murine models, which may limit the direct translatability of findings to the human setting. However, recent phenotypic characterizations of human hematopoiesis indicate that despite difference in the surface marker profiles, the principles of early lymphocyte development are similar between human and mice [22, 373].

### 2.1.1 Mice models

#### CD45 congenic mice

CD45, encoded by the *Ptprc* gene, is a trans-membrane protein tyrosine phosphatase involved in cell activation and signaling [374–376]. The broad surface expression of CD45 on all hematopoietic cells except erythrocytes and platelets, has made the protein a widely used pan-hematopoietic marker [377, 378]. Different isoforms of CD45 have been identified. While C57BL/6 mice carry the common CD45.2 isoform encoded by the $Ptprc^b$ allele, the $Ptprc^a$ allelic variant translating into CD45.1 has been identified in the SJL mouse strain [379, 380]. The different isoforms enable the use of CD45 as a congenic marker, facilitating tracking and quantification of reconstituted hematopoietic donor cells in a transplantation setting [381, 382].

### iCas9 mice

The CRIPSR associated protein 9 (Cas9) is an endonuclease that can introduce guide RNA (gRNA) directed double stranded breaks in genomic DNA [383, 384]. As cells respond to double stranded breaks primarily through the error prone non-homologous end joining (NHEJ) pathway which often introduces mutations [385], the CRISPR/Cas9 system can be used for targeted inactivation of genes or TF binding in regulatory elements. For generation of a mouse line with inducible Cas9 activity (iCas9), TetO-Cas9 mice (JAX:#029476, B6.Cg-*Col1a1*$^{tm1(tetO\text{-}cas9)Sho}$/J) expressing tetracycline responsive element (TetO)-Cas9 from the *Col1a1* promoter were crossed with R26m2rtTA [386] mice expressing a tetracycline-controlled transactivator. The iCas9 mouse strain provide temporal control of Cas9 expression *in vivo* and *in vitro* through doxycycline (DOX) administration.

### *Arid1a*$^{flox/flox}$*-hCD2iCre* mice

For investigation of the role of BAF subunit ARID1a in lymphocyte development, a conditional *Arid1a* knock-out mouse strain was generated. *ARID1a*$^{flox/flox}$ mice carrying loxP sites flanking exon 8 from Gao et al. [387] were crossed with human CD2 promoter–driven iCre mice [388]. The resulting mice strain (*Arid1a*$^{flox/flox}$*-hCD2iCre*) is characterized by selective inactivation of *Arid1a* in early lymphocyte development [389] due to nonsense-mediated decay of the mutant *Arid1a* transcript [387].

### *Ebf1*$^{+/-}$*Pax5*$^{+/-}$ mice

The *Ebf1*$^{+/-}$*Pax5*$^{+/-}$ transheterozygous (TH) mice model, first described by Ungerbäck et al. is characterized by targeted mutations in the DNA-binding domains of *Ebf1* and *Pax5* causing functional inactivation of the two B cell TFs [333]. The mice model is on a C57BL/6 background and was generated by crossing *Ebf1*$^{+/-}$ [218] and *Pax5*$^{+/-}$ [390] mice. Reduced levels of *Ebf1* and *Pax5* cause lineage placticity in the CD19+ pro-B cell compartment. In contrast to their WT counterpart which are committed to a B lineage fate, the TH proB cells are able to undergo T lineage conversion if exposed to Notch signaling [333]. In addition, the *Ebf1*$^{+/-}$*Pax5*$^{+/-}$ mice are preconditioned malignant transformation. At ∼25-40 week of age, a large fraction of TH mice develops lymphoid leukemia with phenotypic features resembling human B-ALL [339].

### 2.1.2 Cell lines

#### Mice cell lines

***230-238***

230-238 is murine B cell progenitor cell line generated by Abelson transformation of BM cell from BALB/c mice [391, 392]. The 230-238 cells represent a pro-B cell like stage [393–395] and can be maintained in RPMI1640 supplemented with 10% heat-inactivated fetal calf serum, 25mM HEPES, 50 g/ml gentamicin and 50 $\mu$M $\beta$-mercaptoethanol.

***Scid.adh.2C2***

The Scid.adh.2C2 cell line is a DN3-like murine T progenitor cell line. The cell line originate from a subclone of the Scid.adh cell line derived from a spontaneous thymic lymphoma in SCID mutant mouse [396]. Scid.adh.2C2 cells can be cultured in RPMI1640 with 10% fetal bovine serum, sodium pyruvate, non-essential amino acids, gentamicin and 50 $\mu$M $\beta$-mercaptoethanol.

#### Human cell lines

***Flp-In T-REx-293***

Flp-In T-REx-293 is derived from the 293 human embryonic kidney cell line [397], designed for generation of stable cell lines expressing a gene of interest [398]. Flp-In T-REx-293 can be maintained in DMEM media with 10% FBS.

***NALM6***

NALM6 is a B precursor leukemia cell line derived from peripheral blood of a 19-year old male with B-ALL [399]. NALM6 has been used widely in B-ALL research and has been extensively characterized as part of the Cancer Cell Line Encyclopedia project [400–402].

### 2.1.3 Isolation of primary cells

#### Cell extraction

Stem- and progenitor cells as well as B cell progenitors are isolated from BM in hind bones of adult mice or FL, whereas T cell progenitor are extracted from thymus. Cells are brought into single-cell suspensions by tissue homogenization and filtering. Stage or lineage-specific surface proteins (i.e. markers) enable isolation of defined cell populations using antibody-based separation techniques, including magnetic cell separation and fluorescence-activated cell sorting (FACS).

### Magnetic cell separation

Magnetic cell separation is a method for selection of specific populations from a heterogeneous mixture of cells based on antibody mediated selective attachment of cells to magnetic microspheres (or beads) [403]. There are two forms of magnetic cell separation, positive selection (i.e. enrichment) and negative selection (i.e. depletion). For depletion, the cells unbound to the magnetic beads are retained. Enrichment instead is based on respsuspension of cells from the bead-bound fraction (Figure 2.1). While magnetic cell separation provide an efficient and high-throughput strategy for coarse cell separation, the level of multiplexing and accuracy is not sufficient for isolation of pure cell populations defined by a combination of different surface markers. Thus, extraction of rare progenitor populations are often achieved by depletion of lineage positive cells followed by more precise and higher resolution cell separation through fluorescence-activated cell sorting (FACS).

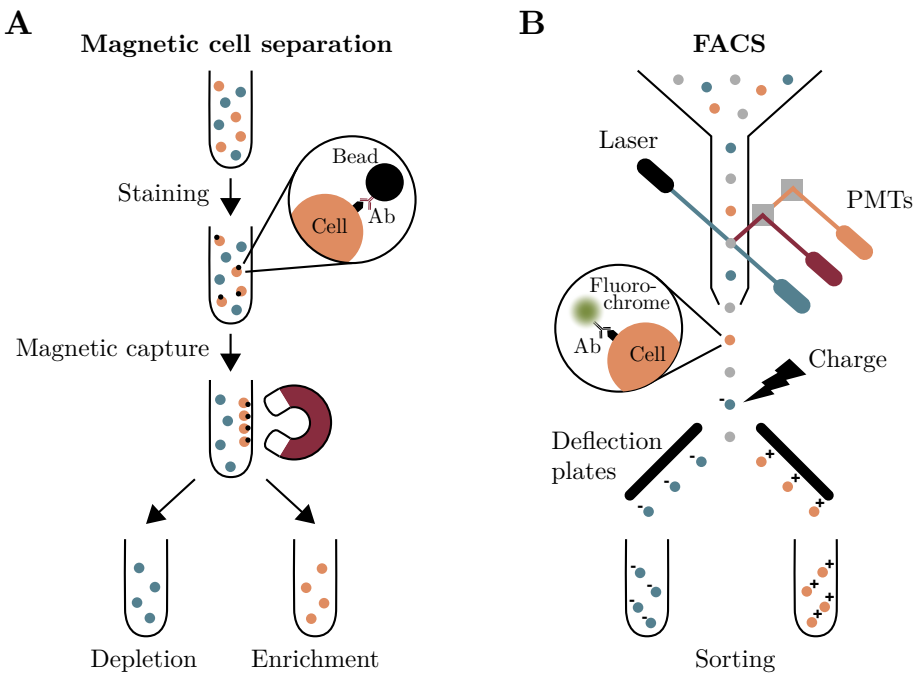



**Figure 2.1: Schematic illustration of magnetic cell separation and FACS.** **A**. Depletion or enrichment of cells in a heterogeneous mixture can be performed by selective antibody mediated attachment of cells to magnetic beads. **B.** FACS provide a fluidics based method for characterization and isolation of cells by electrostatic sorting of fluorescently labeled cells.

### Fluorescence-activated cell sorting

Flow cytometry refers to the measurement of physical and chemical characteristics of single cells (or other particles) in a fluid stream [404]. Since the first implementation of flow cytometry for measurement of cellular volume 75 year ago [405], great technological progress has been made, including the development of fluorescence-based flow cytometry in 1969 [406], and today flow cytometry is a routinely used method for complex cell characterization in both research and in the clinic with the potential to analyze 30+ markers simultaneously [404]. In addition to cell phenotyping using fluorochrome coupled antibodies, flow cytometry based assays for cell viability, proliferation and apoptosis among others are available. Fluorescence-activated cell sorting (FACS) is an extension to flow cytometry, that enable sorting of cell based on a combination of surface markers at single-cell resolution by electrostatic charging of cell enclosed droplets [407–409]. The principle of FACS is based on hydrodynamic focusing of fluorescently labeled cell info a single file stream. As the cells pass through laser beams, fluorochromes used for the detection of target proteins is exited and emit light which is converted into electrical signal by photo-multiplying tube (PMT) detectors. If the sorting requirement is met, an electrostatic charge is applied to the stream, which enable droplet based sorting of cells in a deflection field [404].

In hematology, FACS has been used extensively for analysis and isolation of defined cell populations. In the work presented in this thesis, FACS has primarily been used for quantification and isolation of progenitor populations belonging to the lymphoid lineage, see Table 2.1 for details.

**Table 2.1:** Cell surface markers for defined cell populations in lymphoid development.

| **Lineage** | **Population** | **Sorting strategy** |
|---|---|---|
| Progenitors | LSK | Lin$^{*}$-Sca1+Kit+ |
| | LMPP | Lin$^{*}$-Sca1+Kit+Flt3+ |
| | CLP | Lin$^{*}$-Flt3+IL-7R+Kit$^{\text{int}}$Sca1$^{\text{int}}$ |
| B lineage | ProB | Lin$^{**}$-CD19+IgM-IgD-CD43+Kit+ |
| | PreB | Lin$^{**}$-CD19+IgM-IgD-CD43-Kit- |
| | ImmB | Lin$^{**}$-CD19+IgM+IgD- |
| | MatB | Lin$^{**}$-CD19+IgM+/-IgD+ |
| T lineage | DN | Lin$^{***}$-CD4-CD8- |
| | DN1 | *DN*: CD44+CD25-Kit$^{\text{high}}$ |
| | DN2a | *DN*: CD44+CD25+Kit$^{\text{high}}$ |
| | DN2b | *DN*: CD44+CD25+Kit$^{\text{int}}$ |
| | DN3 | *DN*: CD44-CD25+ |
| | DN4 | *DN*: CD44-CD25- |
| | DP | Lin$^{***}$-CD4+CD8+ |
| | CD4SP | Lin$^{***}$-CD4+CD8- |
| | CD8SP | Lin$^{***}$-CD4-CD8+ |

* Ter119, Mac1, Gr1, CD11c, NK1.1, CD4, CD8, CD19

** Ter119, Mac1, Gr1, CD11c, NK1.1, CD4, CD8

*** Ter119, Mac1, Gr1, CD11c, NK1.1, CD19

## 2.2 Cell manipulation and characterization

### 2.2.1 *In vivo* assays

#### Transplantation

Since the groundbreaking work by Till and McCulloch in the early 1960s taht demonstrated the existence of multipotent hemtopoetic progenitor cells with self-renewal capacity via injection of bone marrow cells into irradiated recipient mice [410–412], transplantation has remained a gold standard experimental method for studying hematopoiesis and immune cell development. The use of congenic mice strains (e.g. CD45.1/CD45.2) for discrimination of donor cells from host cells, enable competitive assays for evaluation of stem- and progenitor cell fitness and the ability to specifically investigate cellular function and lineage reconstitution of donor cells in a biologically relevant context [381, 382]. In addition, transplantation can be used for functional evaluation of genetic modification and the ability to separate cell-intrinsic and environmental effects. During transplantation the donor cells are transferred to the host by intravenous injection via the tail vein. To achieve efficient engraftment, pre-conditioning by irradiation or chemotherapy is often required to eliminate host competition, and free up bone marrow niches [413–415].

### 2.2.2 *In vitro* assays

#### Op9 co-culture systems

Op9 is a murine bone marrow derived stromal cell line, that lack functional macrophage colony-stimulating factor (M-CSF) and can support the generation of hematopoietic cells of several lineages from ES cells [416], including functional B cells [417]. B cell progenitors can be differentiated from multipotent hematopoietic progenitors and expanded in culture in Op9 co-cultures using a cytokine cocktail of c-Kit ligand (KL), Flt3 ligand (Flt3L) and IL-7 [418]. In the *in vivo* setting, T cell development is dependent on Notch signaling in the thymus [278, 419]. Ectopic expression the Notch ligand Delta-like-1 in Op9 stromal cells (Op9DL1) is sufficient to induce T cell lineage restriction and generation of functional $\alpha\beta$- and $\gamma\delta$-TCR+ T cells in culture [261].

#### Flow cytometry based assays

As previously described (Section 2.1.3), flow cytometry is a frequency used single-cell method for analysis of cells in a fluid stream. For evaluation of cell population distributions and tracking of differentiation assays, conventional flow

cytometry based on cell surface staining is suitable. In work presented in this thesis, flow cytometry is also used to evaluate lineage placticity, including T lineage conversion of B cell progenitors (*Paper IV*), which uses the combined negative expression of the B cell marker CD19 and positive surface expression of Thy1 to define a population of cells in the process of lineage conversion [333], see Figure 2.2A. For analysis of intracellular targets (e.g. TFs, histone modifications), cell fixation and permeabilization of the cell membrane is required in order for the fluorescently-labeled antibodies to gain access to target antigen [420]. In addition, numerous different flow cytometry based assays have been developed to investigate different aspects of cellular states and functions, including cell cycle and apoptosis assays.

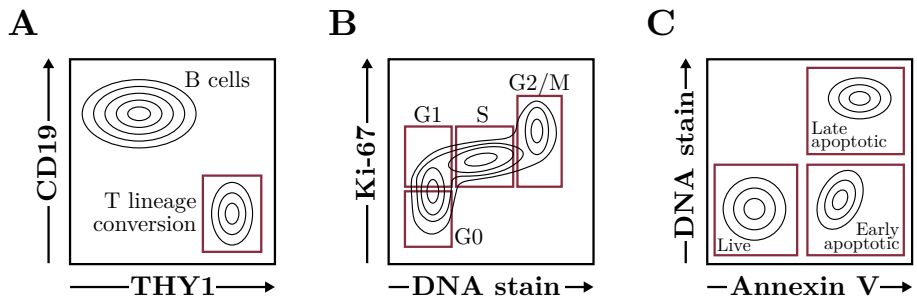



**Figure 2.2: Schematic illustration of gating strategy for flow cytometry based assays.** **A**. T linage conversion assay using B cell using CD19 and THY1 surface expression. **B**. Cell cycle analysis based on the proliferative marker Ki-67 and DNA content. **C**. Apoptosis assay with phospholipid-binding protein Annexin V.

### *Cell cycle assay*

Proliferating cells undergo sequential transition between G1, S, G2 and M phases [421]. As cell cycle progression involve replication of DNA, measurement of DNA content fluorescent DNA staining (e.g. 4,6-diamidino-2-phenylindole (DAPI), Propidium iodide (PI)) can be used to separate three major phases of the cell cycle (GO/G1, S, G2M) [422]. If combined with targeting of the proliferative marker Ki-67, the quiescent G0 cell can be discriminated from G1 [423, 424], see Figure 2.2B. To further isolate mitotic cells, anti-phosphorylated (ser10) H3 antibody can be included [425, 426].

### *Apoptosis assay*

During the early stages of apoptosis, phosphatidylserine (PS) normally located on the inner surface of the cell membrane are translocated to the cell surface. Fluorescently labeled $Ca^{2+}$-dependent phospholipid-binding protein Annexin V can be used to detect these early apoptotic cells through binding to cell sur-

face localized PS [427]. Combined with a DNA dye, the early apoptotic cells can be discriminated from late apoptotic or necrotic cells with a damaged cell membrane [428], see Figure 2.2C.

### 2.2.3 Manipulation of gene expression

#### Viral vector based overexpression

The dependence of host cells for replication of viruses has through evolution tailored the viral particles for efficient gene delivery, thus making them a useful tool for gene manipulation. Overexpression of a desired target gene can be achieved by transfection of cells with recombinant viral particles, containing the gene of interest [429]. Among the most commonly used viruses for gene delivery are retroviruses, a family of enveloped RNA viruses that use reverse transcription to generate double stranded DNA that is subsequently integrated into the host genome [430, 431]. Most retroviruses require the host cells to be actively cycling, however lentivirus, a subtype of retrovirus, have the ability to infect non-dividing cell [432]. Recombinant viruses are obtained using engineered packaging cell (e.g. Plat-E and 293T) that via a trans-complementation process of viral proteins are able to generate replication-defective viruses with the capacity to deliver the transgene into target cells [429].

#### CRISPR/Cas9

As a defense mechanism against plasmids and viruses, prokaryotes have evolved an adaptive defense system known as clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR-associated 9 (Cas 9) [433–435]. The ability of CRIPSR/Cas9 system to detect and silence DNA in a sequence-specific manner has lead to re-purposing of the system as a tool for gene editing. Mechanistically, the sequence-specific gene editing capabilities of CRISPR/Cas9 is achieved through guide RNA (gRNA) directed targeting of a Cas9 endonuclease to the target loci, followed by induction of a double stranded break (DSB) [383, 384]. The default pathway for DNA repair of Cas9 induced DSB are the non-homologous end joining (HNEJ), an error-prone repair pathway that often result in frameshift mutation and subsequent gene knockout. For more precise editing (e.g. corrections, insertions), donor template can be provided to facilitate DNA repiar via H´homology-directed repair (HDR) [436].

### Temporal control of gene expression

For precise control of cellular processes and dissection of individual protein activities, the ability for temporal control of gene expression can provide a great advantage. Among the most commonly used systems for temporal control of gene expression are Tamoxifen and Doxycycline inducible systems.

#### *Tamoxifen inducible systems*

The tamoxifen inducible system originate from the discovery that the human estrogen receptor (ER) can be used for gene induction if fused to a protein with gene regulatory function [437, 438]. To overcome the effect of endogenous estrogen, a modified version of ER with reduced affinity for normal estrogen but high affinity for tamoxifen and 4-hydroxy tamoxifen (4-OHT) (ERT) was later developed [439, 440]. The system is based on the principle that binding of tamoxifen cause the cytoplasm sequestering ERT-fused protein to translocate to the nucleus where is can exerts its gene regulatory function (see Figure 2.3A). The tamoxifen inducible system is often used in combination with the Cre-lox system to induce site-specific DNA recombination [441], but is also useful for temporal control over TF activity.

#### *Tetracycline-Inducible Systems (Tet-On/Tet-Off)*

In contrast to the indirect control of gene expression via protein translocation provided by the tamoxifen inducible system, tetracycline-based induction allow for precise and reversible control of individual target genes. The regulatory mechanism is mediated by a transactivator that control target gene expression via a Tet response element (TRE), a synthetic promoter containing a Tet operator (TetO) sequences [442]. For negative regulation of gene expression (Tet-Off), the transactivator consist of a Tet repressor protein (TetR) fused to the activating domain of the herpes simplex virus VP16 protein, resulting in an tetracycline inducible transcriptional activator (tTA). In the absence to tetracycline (or its more stable derivative doxycycline (Dox) [443]), tTA bind TRE and activate transcription. Binding of Dox triggers a conformational switch in tTA, causing its release from TRE and target gene to be turned off [442]. For induced transcriptional activation (Tet-On) a reverse transcriptional activator (rtTA) can instead be used. The rtTA exhibit a reverse response to Dox administration, causing it to selectively bind TetO and activate gene expression in the presence of Dox [444], see Figure 2.3B.

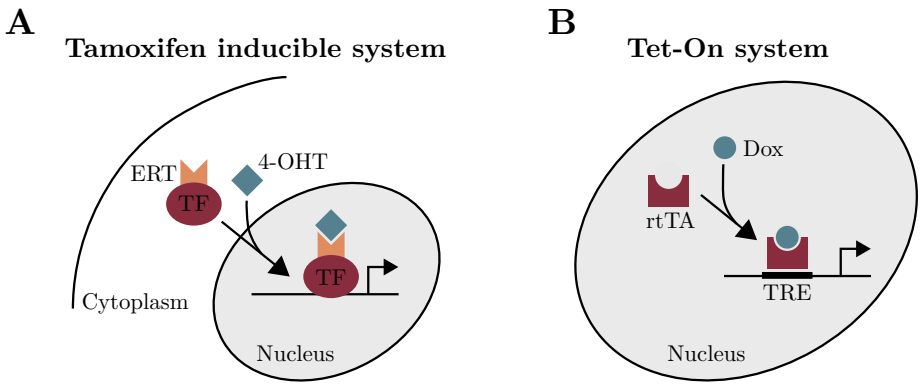



**Figure 2.3: Schematic illustration of systems for temporal control of gene expression. A**. The tamoxifen inducible system can be used for inducible TF activity by 4-OHT driven nuclear translocation. **A**. The Tet-On system provide temporal control of a target gene via doxycycline triggered rtTA mediated gene activation.

## 2.3 Methods to study gene regulatory mechanisms

Deciphering developmental processes demands integrative analysis of regulatory mechanisms across multiple layers, collectively controlling the level, timing and cell type specificity of gene expression. These regulatory layers are presented below, with selected methods in italic. The most direct measurement of gene regulation is evidently the quantification of gene transcription (*RT-qPCR, RNA-seq*), but binding and activity of trans-acting regulatory proteins (*ChIP-seq, Cut&Run, EMSA*) to cis-regulatory elements (*Luciferase reporter assay, CRISPR/Cas9*) as well as aspects of the chromatin itself including accessibility (*ATAC-seq*) and distribution of histone modification (*ChIP-seq, Cut&Run*) plays a pivotal role. On top of this the three-dimensional topology of the genome (*Hi-C, PLAC-seq*) provide an additional layer of regulation by orchestrating dynamic chromatin compartmentalization and long-range enhancer–promoter interactions. The proteins exhibiting gene- or chromatin-regulatory functions are not acting as isolated entities, but instead engage in dynamic interactions with other proteins either directly via physical contact (*BioID*) or indirectly via transcriptional control in intricate TF networks.

This thesis is mostly focused on the mechanisms involved in transcriptional control, however it should be noted that regulation of RNA processing and stability, translational regulation as well as protein modification and degradation provide additional layers that can fine-tune the intricate regulatory machinery. Further limitation of this thesis lies in the restricted investigation of cell–cell interactions and the influence of the microenvironment, factors well established to have a significant impact on cellular function.

### 2.3.1 Principles of next-generation sequencing

Although the methods presented in this thesis target distinct molecular features, many are based on next-generation sequencing (NGS) and adhere to its core principles (see Figure 2.4):

1. **Target capture:** NGS methods begin with the capture of the molecular target, a step that is method-specific and may involve crosslinking, target extraction, or antibody-based isolation. The detailed workflows and key considerations for each technique are described in the following sections.

2. **Library preparation:** During the library preparation step, the captured nuclei acids are converted into sequencing-ready molecules through fragmentation, adapter ligation, amplification, and purification [445]. In contrast to the conventional workflow in which fragmentation and adapter

ligation are performed separably, tagmentation-based approaches use a transposase to simultaneously fragment and tag DNA [446, 447].

3. **Massively parallel sequencing:** Library fragments are clonally amplified and sequenced by cyclic incorporation and imaging of fluorescently labeled nucleotides [448]. Sequencing can be performed in two modes; single-end, where only one end of the DNA fragment is sequenced, and paired-end which provide improved accuracy, coverage and detection of DNA rearrangements by sequencing of both ends.

4. **Data analysis:** To prepare NGS data for downstream analysis according to the biological question, the raw data is preprocessed by quality-filtering, trimming and mapping of reads to the reference genome [449].

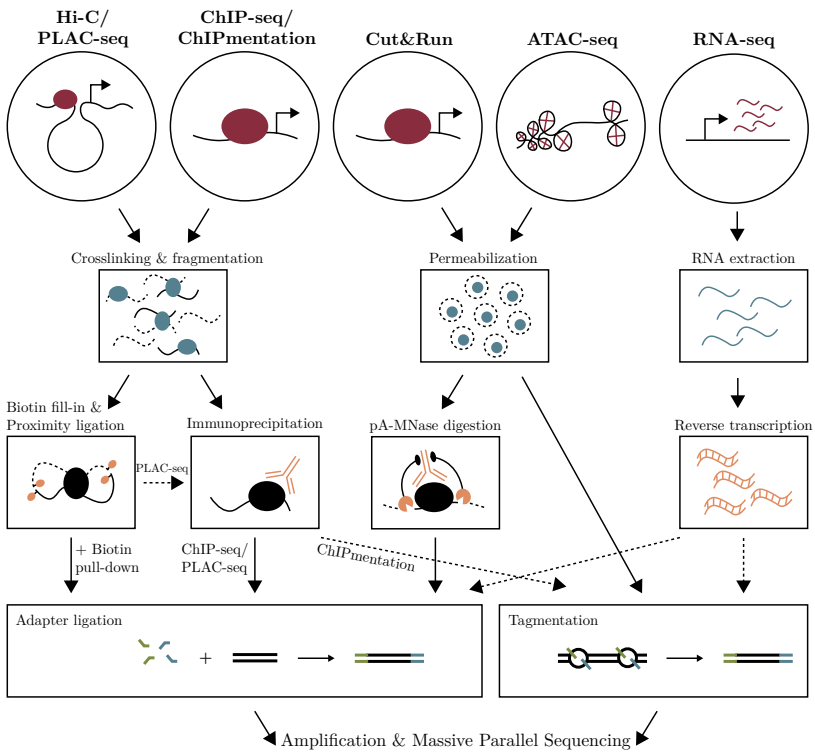



**Figure 2.4: Flowchart of next-generation sequencing (NGS) methods.** The main steps of target capture and library preparation for a number of NGS methods are presented in aggregated form.

### 2.3.2 Gene expression

***Real-time quantitative PCR***

Real-time quantitative PCR (RT-qPCR) is a widely used technique for quantifying specific RNA transcripts. To generate the input material for RT-qPCR, the RNA is extracted and converted into complementary DNA (cDNA) by reverse transcription. The cDNA serves as the template for PCR amplification of target sequence, monitored in real time using fluorescent dyes or probes [450]. Sequence specific TaqMan probes enable precise detection of target sequence by using a reporter-quencher system that ensure that fluorescent signal is only emitted during amplification when the reporter is cleaved by the Taq polymerase [450, 451]. As the fluorescent signal is proportional to the amount of amplified target, real-time monitoring allow precise quantification of initial RNA levels. qPCR have application beyond RNA quantification, including detection of DNA mutations and V(D)J recombination analysis [452].

***RNA-sequencing***

RNA-sequencing (RNA-seq) is a next-generation sequencing based method for transcriptomic profiling [453]. Although many RNA-seq protocols have been developed, the general workflow of RNA-seq involve RNA extraction followed by enrichment of mRNA by poly(A) selection. The mRNA is then fragmented and converted into double-stranded cDNA, subsequently prepared for sequencing either through conventional adapter ligation [453] or via transposase-mediated library construction [454]. Preservation of complete transcript information can be obtained by the Smart-seq2 protocol, which utilize a template-switching activity of reverse transcriptase for full-length cDNAs enrichment [455].

### 2.3.3 Activity of cis-regulatory elements

***CRISPR/Cas9-based evaluation of CRE activity***

CRISPR/Cas9 is not restricted to gene editing, the function of regulatory elements (e.g. TF binding sites) can also be evaluated by targeting of Cas9 to these non-coding regions in the genome. In addition, modified versions of the CRISPR/Cas9 system has been developed for cleavage free control of gene expression. In CRISPR interference (CRISPRi), a catalytically dead Cas9 (dCas9) lacking endonuclease activity is used for repression of gene transcription via physical obstruction of RNA polymerase or TFs [456]. To further enhance the repressional capacity, dCas9 can be fused to a repressor domain (e.g. KRAB). Similarly, CRISPR activation (CRISPRa) can be achieved via fusion of dCas9 to a transcriptional activator (e.g. VP64) [457].

#### *Luciferase reporter assay*

Luciferase reporter assay is designed to measure the activity of a promoter or other regulatory elements by using a reporter construct in which the luciferase enzyme is placed under the control of the target element [458, 459]. After transfection of the reporter construct into the cell of interest, luciferase will be produced in response to target element activation. Subsequent addition of luciferase substrate enable quantification of element activity via detection of light emitted in the enzymatic reaction [458].

### 2.3.4 Protein interactions

#### Protein-Protein interaction

#### *BioID*

BioID is a technique used to identify interaction partners of a target protein through *in vivo* proximity-dependent labeling [460]. The method utilizes a mutant version (R118G) of the *Escherichia coli* biotin protein ligase (BirA*)[461] which enable promiscuous biotinylation in a proximity-dependent manner [461]. Fusion of this mutant BirA* to the target protein (the bait) facilitate proximity-dependent biotinylation of neighboring proteins [462]. After harsh cell lysis and protein denaturation, biotinylated proximity interactors (PXI) can be isolated using streptavidin affinity purification, and identified by mass spectrometry (MS) [460], see Figure 2.5. As BioID relies on physical proximity rather than direct binding, the method is well-suited for detection of interactions with larger protein complexes (e.g. chroamtin remodelling complexes).

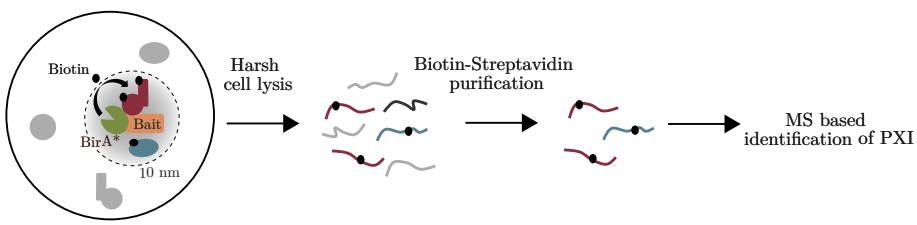



**Figure 2.5: Schematic illustration the BioID workflow.** Fusion of a promiscuous biotin ligase (BirA*) to a target protein (bait), enable proximity-dependent biotinylation of neighboring proteins. Biotinylated proximity interactors (PXIs), isolated by streptavidin affinity purification, are identified by mass spectrometry (MS).

### DNA-Protein interaction

#### *Electrophoretic mobility shift assay*

Electrophoretic mobility shift assay (EMSA) is an electrophoresis based method for detection of protein interaction with DNA (or RNA) [463–465]. The protein of interest is incubated with a labeled (e.g. radioactive) nucleic acid, followed by separation of the bound complexes by gel electrophoresis. As the bound protein will slow down DNA migration, protein-DNA interaction can be identified as a shift in the signal compared to unbound control [463, 464]. To distinguish sequence-specific interactions from non-specific binding radioactively labeled competitor DNA with known protein binding can be introduced along with unlabeled target fragments. In this case, loss of signal indicates that the protein specifically binds the target fragment, as it can out-compete binding to the positive control sequence.[466].

#### *Chromatin Immunoprecipitation Sequencing*

Chromatin immunoprecipitation sequencing (ChIP-seq) provides a genome wide method for mapping of target protein to DNA [467]. To preserve the structural integrity of the sample, the initial step in the ChIP-seq workflow is covalent binding of DNA-protein interaction (i.e. crosslinking). While formaldehyde (FA) crosslinking is sufficient for efficient crosslinking histones to DNA [468], the stabilization of interaction between DNA and proteins with weaker or more indirect interactions (e.g. TFs, subunits of protein complexes) can be improved by combining FA crosslinking with a bridging agent (e.g. Disuccinimidyl glutarate (DSG)) [469, 470]. Next, the chromatin is fragmented by sonication and subjected to immunoprecipitation with a target specific antibody to pull down fragments bound by the protein of interest [471]. The bound DNA is released by reverse-crosslinking and recovered DNA fragments are purified to enable subsequent library preparation and sequencing [467]. ChIPmentation, a method that combine chromatin immunoprecipitation with a tagmentation-based library preparation strategy provide a more streamlined protocol by introducing sequencing-compatible adapters directly on the protein-bound chromatin [472].

#### *Cut&Run*

Cleavage under targets and release using nuclease (Cut&Run) is an alternative method for genome-wide protein-DNA profiling, offering reduced input requirement and improved signal-to-noise ratio compared to ChIP-seq [473]. In Cut&Run, antibody-based target capture is performed *in situ* in isolated nuclei [473] or directly in permeabilized cells [474] bound to concanavalin A–coated magnetic beads. Protein-bound DNA fragment are retrieved by calcium mediated activation of a protein A/G and micrococcal nuclease fusion protein (pA-

MNase) that is recruited to the target site and induce DNA cleavage [473]. Diffused DNA fragment are collected for library preparation and subsequent sequencing. Analogous to ChIPmentation, an alternative tagmentation-based method called Cut&Tag has been developed for Cut&Run [475].

### 2.3.5 Chromatin accessibility

#### Assay for Transposase-Accessible Chromatin using sequencing

Assay for transposase-accessible chromatin using sequencing (ATAC-seq) is a next-generation sequencing–based method for genome-wide profiling of chromatin accessibility [476]. ATAC-seq is based on the ability of hyperactive Tn5 transposase [477], when loaded with sequencing adapters, to simultaneously fragment and tag chromatin in a processes known as "Tagmentation" [446]. Transposons preferentially target nucleosome-free regions *in vivo* due to reduced steric hindrance [478], a property also exhibited by the prokaryotic Tn5 transposase [476]. Consequently, Tn5-mediated transposition in permeabilized cells produces open-chromatin–enriched DNA fragments compatible with high-throughput sequencing. Omni-ATAC, an optimized version of original ATAC-seq protocol provide increased compatibility with diverse sample types as well as improved signal-to-noise ratio and reduced mitochondrial DNA contamination [479]

### 2.3.6 Genome topology

With increasing evidence for the crucial role of genome topology in gene regulatory processes, including its capacity to physically connect distal cis-regulatory elements (CREs) with their target genes, a wide range of techniques have been developed to investigate the genome's three-dimensional organization. These approaches can be broadly classified into the following categories:

**Microscopy-based approaches:** Imaging-based methods enable direct visualization of genome organization within the native nuclear environment [480]. Building on fluorescence *in situ* hybridization (FISH), which uses fluorescently labeled DNA probes to detect the spatial location of specific genomic loci [481], microscopy-based approaches have advanced substantially in both resolution and multiplexing [482], establishing them as powerful tools for investigating genome organization at a single-cell level.

**Sequencing-based approaches:** Sequencing-based techniques enable genome-wide, high-throughput mapping of genome organization [480]. Many sequenced-based approaches belong to the chromosome conformation capture (3C) family, characterized by quantification of contact frequencies between genomic loci via proximity ligation.[483]. Except for methods belonging to the 3C family, non–proximity ligation–based methods for chromatin interactions including the barcode-based split-pool recognition of interactions by tag extension (SPRITE) [484] and genome architecture mapping (GAM) based on cryosectioning [485]. Furthermore, tyramide signal amplification sequencing (TSA-seq) can provide additional spatial information by enabling distance mapping in 3D [486].

### Chromosome conformation capture

Chromosome conformation capture (3C), initially developed for quantification of pairwise interaction between genomic loci [487] has evolved into an entire family of methods, some of which possesses the ability to generate global maps of the chromatin interactome [112]. Irrespective of assay, the 3C methods are based on the principle of proximity ligation, i.e. the ligation of DNA fragment located close in the 3D space, for quantification of chromatin interactions [480].

#### *Hi-C*

In Hi-C, proximity ligation is coupled with massive parallel sequencing, enabling quantification of interaction frequencies across all pairs of genomic loci [112]. To generate sequencing ready Hi-C libraries, the chromatin is first crosslinked to preserve the spacial interaction, followed by restriction enzyme (RE)-based DNA digestion. Next fragment ends are biotin labeled and proximity ligated to capture spacial interactions. Biotin pull-down of de-crosslinked and sheared chromatin selectively enrich for DNA fragments that contain contact information, which are subsequently subjected to adapter ligation and amplification [112]. While the initial megabase resolution contact maps were sufficient to identify segregation of the genome into transcriptionally active (A) and inactive (B) compartments [112], higher resolution experiments have allowed for the discovery of more fine-scaled structure including TADS and enhancer-promoter loops [117]. By replacing the RE-based fragmentation by micrococcal nuclease digestion, enhanced resolution down to the nucleosome level is offered in the modified Micro-C protocol [488]. The effective resolution can also be improved by target approaches, that enrich for predefined genomic regions (e.g. capture Hi-C [145]) or interaction associated with a specific DNA-binding protein (e.g. HiCHiP [489], PLAC-seq [490]).

***HiChIP/PLAC-seq***

HiChIP [489] and proximity ligation-assisted ChIP-seq (PLAC-seq) [490] are parallelly developed methods for genome-wide targeted mapping of chromatin interactions. Both methods build on the Hi-C protocol, incorporating an additional chromatin immunoprecipitation step after proximity ligation to selectively enrich for chromatin contacts associated with a specific DNA-bound protein [489, 490]. Despite the shared principles of HiCHiP and PLAC-seq, small deviations exist in their respective workflow, e.g. the use of tagmentation (HiChIP) vs conventional adapter ligation (PLAC-seq) based library preparation [489, 490]. Collectively these target approaches offer a cost effective alternative to Hi-C for investigation of functionally relevant chromatin interactions. By selecting for interactions linked to a specific histone modification or regulatory protein, these targeted strategies enable focused interrogation of enhancer-promoter loops and their role in gene regulatory mechanisms.

### 2.3.7 Combined methods

Regulation of gene expression is a complex process, shaped by the chromatin state and the activity of regulatory proteins, which often act via distal CREs brought into proximity with target gene via chromatin looping. While the methods described previous in this section are effective for dissecting a certain aspect of gene regulation, single-modality assays are limited in their ability to link regulatory potential to transcriptional output. An ability also restricted by the inherently population-averaged nature of bulk sequencing methods, which mask cellular heterogeneity and obscure rare and transitional cellular states. The development of high-throughput single cell sequencing technologies [491–493], particularly methods that enable simultaneous profiling across multiple modalities, has greatly enhanced our ability to resolve cellular heterogeneity and to link molecular profiles to distinct cellular states [494].

***Single cell Multi-Ome (ATAC/RNA) sequencing***

Single-cell joint profiling of chromatin accessibility and gene expression (scMulti-Ome) [495, 496] has emerged as powerful tool for gene regulatory network interference because of its ability to dissect how the chromatin state and transcriptional output are coordinated within individual cells [494]. The workflow of scMulti-Ome, commercially available as a streamlined protocol by 10x genomics [497], is initiated by tagmentation of accessible chromatin in isolated nuclei. Next, a microfluidic chip is used for partitioning of individual nuclei into droplet containing uniquely barcoded beads called gel bead-in emulsions (GEMs). Inside the GEMs, both mRNA reverse transcribed into cDNA and

tagmented ATAC fragment are tagged with a cell barcode. In addition, each cDNA molecule will also carry an unique molecular identifier (UMI) to allow discrimination between unique transcripts and PCR artifacts [497]. Barcoded molecules are recovered and amplified libraries from the two modalities are separately sequenced. The use of shared barcodes enable computational integration of the two datasets. See Figure 2.6 for a schematic overview of the scMulti-Ome workflow.

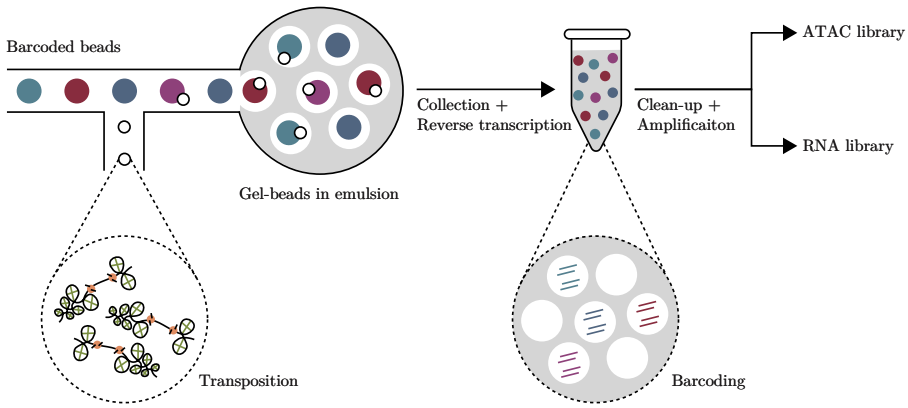



**Figure 2.6: Overview of the 10x workflow for joint profiling of chromatin accessibility and gene expression in single cells.** First isolated nuclei are subjected to transposition in bulk. Next GEMs (Gel Bead-in-emulsion) are used for capturing of individual nuclei, for reverse transcription and barcoding of cDNA and ATAC fragments. After clean-up and amplification, the scMuti-ome libraries are separably sequenced and computationally integrated based on the cell specific barcodes.

## 2.4 Computational methods

### 2.4.1 Processing of NGS data

#### General pre-processing

Although NGS methods address different aspects of gene regulation, their shared output (raw sequencing data) warrants a common set of core processing steps, outlined below with suggested tools highlighted in italic:

1. **Base calling:** In an initial step, known as base calling, the fluorescence signal recorded for each cycle are translated into nucleotide sequences (reads) in FASTQ format [498]. Base calling is often performed automatically by platform specific software.

2. **Quality control & Trimming:** Next reads are quality assessed (*FastQC* [499]) and trimmed to remove adapters and low-quality bases (*TrimGalore* [500]).

3. **Mapping:** In the mapping (or alignment) step, reads are assigned to a reference genome. Mapping of unspliced reads (e.g. ATAC-seq, ChIP-seq) are typically performed with Burrows–Wheeler aligners (BWA). The BWAs (e.g. *Bowtie2* [501]) locates short exact matches (seeds) in the reference, which are subsequently extended to full alignments with allowed mismatches or gaps [502]. Splice aware aligner (e.g. *STAR*) has been specifically developed for RNA-seq reads containing exon-exon junction. The STAR aligner uses a two step approach; first maximal mappable segments of each read are located in an iterative process based on suffix array indexing. Next the segments are clustered based on proximity and stitched together across splice junctions to generate full read alignments [503].

4. **Post-alignment processing:** This step is dependent on data format and downstream analysis but often involve sorting and index (*SAMtools* [504]). For paired-end sequencing, duplicate removal (*Picard* [505]) is often performed at this stage to prevent over-representation and PCR bias.

5. **Feature identification & quantification:** The final step in the pre-processing of NGS data is the assignment and quantification of reads associated with distinct features (e.g. gene transcripts or genomic regions). For methods based on genomic DNA lacking pre-defined units, read assignment can be made by genome binning or by identification of regions with enriched read count over background (i.e peak-calling).

### Method specific processing

#### *RNA-seq*

Due to mRNA splicing, RNA-seq data will contain reads spanning exon-exon junctions. Hence, the mapping of RNA-seq reads require specific splice-aware aligners (e.g. *STAR*), which are based on algorithms compatible with non-contiguously mapping [503]. The existence of alternative isoforms, generated via alternative splicing of mRNA, provide another challenge in RNA-seq data processing. In *RSEM*, a widely used tool for estimation of transcript abundance, the process of read assignment is determined by probabilistic modeling [506].

#### *Peak based methods*

Unlike the RNA-seq based measurement of transcript abundance, genomic DNA–based assays produce reads that lack an intrinsic unit of quantification. Although arbitrary binning of the genome can be used, identification of regions with statistically significant enrichment of accumulated signal over background provide a more biological relevant unit of measurement [467]. The process of obtaining these enriched regions, i.e. peaks, are known as peak-calling and is a crucial step in the workflow of ATAC-seq, ChIP-seq and similar methods. A popular tool for peak calling is *Macs2*, an updated version of the Macs algorithm [507], that model read coverage using a local Poisson distribution and uses a sliding window-based approach to identify enriched regions. For ChIP-seq data, the peak-calling is typically performed in comparison with a control sample (e.g. input or mock IP sample) to distinguish true binding sites from regions of elevated background due to local bias [507]. *SEACR*, a more recent peak-calling tool specifically tailored for the sparse background signal often observed in Cut&Run data, instead define the threshold for what is considered a peaks based on a global background distribution [508]. Peak sets may be refined using the irreproducible discovery rate (*IDR*) framework, which identifies reproducible peaks across replicates through significance score ranking, thereby controlling false positives [509]. To obtain a measurement of chromatin accessibility or protein binding, read coverage within peaks can be quantified by *Diffbind* among other tools [510].

#### *Chromosome conformation capture*

To extract chromatin interaction information from the chimeric reads generated by the characterizing proximity ligation step of 3C methods, a number of specific pre-processing step are required. *HiC-Pro* packages these processing steps into a streamlined workflow for generation of contact maps from raw chimeric reads [511]. To enrich for valid ligation products, only unique read pairs mapping to different location in the genome are retained. Next a contact matrix is generated by genome binning and subsequent read pair quantification within bins. It is this

bin size, in conjugation with the constraints imposed by the restriction enzyme that determine the resolution of the final chromatin interactions. To account for systematic biases, a number of normalization strategies have been developed for 3C contact maps, including the ICE method based on iterative matrix balancing with the assumption of equal visibility (i.e. equal total contacts within each bins) [512]. From the normalized contacts maps, interaction calling can be performed by tools like *Fit-Hi-C* [513]. The interaction-calling algorithm in Fit-Hi-C is based on comparison of observed contact frequencies against a background model that account for technical biases and distance decay. As data generated by HiCHIP/PLAC-seq invalidates some of the assumption of Hi-C processing (e.g. the assumption of equal visibility), the separate tools tailored specifically for HiCHIP/PLAC-seq data has been developed (e.g. *FitHiCHIP*) [514]. By default FitHiCHIP implements a coverage normalization strategy that adjust for differences in ChIP enrichment. Protein-anchored chromatin interactions are called using a reference peaks set, either inferred from the HiCHIP data or separately provided [514].

***scMulti-ome***

For multi-ome (ATAC/RNA) data processing, the initial preprocessing steps are performed separably for each modality. For ATAC-seq the first step is validation and counting of barcodes. Next reads are trimmed and aligned to the reference genome using a modified version of the BWA-MEM algorithm with preserved barcode information. Only unique read (based on start and end position as well as cellular barcode) are retained for ATAC-seq peak calling in pseudo-bulk and subsequent peak-barcode matrix construction. For the gene expression modality, reads are trimmed and aligned using *STAR* [503]. Next reads are grouped based on barcode and UMI, followed by filtered (based on most supporting reads) to only allow one gene annotation per UMI-barcode combination. Each unique filtered combinations is recorded as a count in the count matrix. The scMulti-ome processing steps is packed into an automated pipeline, *cell-ranger arc*, developed by the 10x platform [497]. This workflow also facilitate integrative analysis of the two modalities. For downstream analysis including quality control, quantification, dimensionality reduction and clustering, numerous tools have been developed for for both separate and integrative analysis. Among these are *Signac*, a comprehensive toolkit for scATAC-seq data analysis that via compatibility with the *Seurat* package [515] facilitate multi-modal analysis [516].

### 2.4.2 Downstream analysis

#### Dimensionality reduction

Sequence-based omics profiling, particularly at the single-cell level, produces inherently high-dimensional data. This high dimensionality gives rise to the *curse of dimensionality*, a phenomenon in which the sparsity of data in large feature spaces diminishes statistical power and dilutes meaningful relationships between observations [517]. Dimensionality reduction, the process of transforming the high dimensional data into a lower dimensional space, mitigates these effects by reducing the complexity of the data while maintaining the most informative structures [518]. In addition, the ability of directionality reduction to embed high-dimensional data into two or three dimensional spaces enable visual exploration of underlying patterns [519]. Different dimensionality reduction approaches serve distinct purposes, with linear methods (e.g. PCA) capturing global variation while non-linear methods (e.g. t-SNE, UMAP) can preserve more of the local and fine-grained structures of the data [519].

##### *Principal component analysis*

Principal component analysis (PCA) is a linear dimensionality reduction technique that transform data into a set of uncorrelated variables known as principal components (PCs). These PCs, which are linear combinations of the original variables, are generated by sequential maximization of variance [520]. The simple idea of PCA, to reduce dimensionality while preserving as much variability as possible, can be traced back to the early 1900s [521, 522] and has remained a widely used strategy for identification of major sources of variation, visualization and detection of batch effects [523, 524]. Despite the explosion of novel dimensionally reduction methods in the single-cell era, PCA remains frequently used in pre-processing to reduce complexity prior to clustering and non-linear embedding [525].

##### *Uniform Manifold Approximation and Projection*

Uniform Manifold Approximation and Projection (UMAP), has rapidly emerged as the preferred non-linear dimensionality reduction approach for single-cell data. The UMAP algorithm can be described as a two step process; first a weighted k-nearest neighbor graph is constructed in the high-dimensional space to capture local relationships and approximate the data as a fuzzy structure. Next cross-entropy is used to optimize a low dimensional embedding with respect to preserving both local neighborhoods and global structure [526]. The resulting low-dimensional representation enables visualization, clustering, and exploration of complex patterns in high-dimensional single-cell data sets [519].

### Clustering

Clustering is the process of grouping samples, cells or features based on similarity, allowing discovery of cell types, co-regulated feature and other patterns in biological data. A common approach is partition-based clustering, such as k-means, which partitions the data into a fixed number of clusters by iteratively assigning points to the nearest cluster centroid [527]. Hierarchical clustering, not restricted by a predefined number of cluster, instead uses pair-wise distances between data points to construct a dendrogram representing the nested relationship of the data [528]. For large datasets, graph-based clustering is often preferred, where a similarity graph is constructed and algorithms such as Louvain detect communities by maximizing modularity [529].

### Differential feature analysis

Differential feature analysis refer to the identification features (e.g. genes, peaks) with a statistically significant difference in abundance or other quantitative measure between groups of samples (e.g. cell types, experimental conditions). A widely used tool, originally designed for RNA-seq data, is *DESeq2*. DESeq2 models read counts with a negative binomial generalized linear model, estimating gene-specific dispersion and normalizing for differences in sequencing depth [530]. Next statistically significant features are detected using a Wald test [531]. To identify differential chromatin accessibility (ATAC-seq) or protein-DNA binding (e.g. ChIP-seq), the peak-centered *Diffbind* framework can be used. In addition to the actual statistical analysis (often performed by DESeq2), Diffbind facilitate the generation of a consensus peaks set, assignment of read count to peaks and method specific normalization strategies [510]. To control the false discover rate in differential feature analysis, p-values obtain from the statistical test is typically adjusted for multiple testing using the Benjamini–Hochberg procedure [532].

### Enrichment-based analysis

Enrichment-based analysis refers to the computational approach used to determine whether a predefined set of features (e.g. gene associated with a biological pathway or set of genomic regions) is enriched in an input list (often a set of significant features from differential analysis). In an over-representation test, a statistical test (e.g. Fisher's exact or hypergeometric tests) is used to evaluate if the feature set is statistically enriched over a background set [533]. As an alternative approach, gene (or peak) set analysis (GSEA/PSEA) enable the discovery of coordinated differences in levels between two biological conditions [534]. Instead of using an arbitrary threshold for the selection criteria of the input list, GSEA/PSEA uses the entire feature list (ranked based on differential

expression statistic) to calculate an enrichment score that reflect concentration of the feature set in either end of the ranked list.

### Motif analysis

Motif enrichment analysis refers a computational process of identifying short sequences (i.e. motifs) that are statistically over-represented in a set of genomic regions [535, 536]. As DNA-binding protein often exhibit a sequence specific binding pattern, motif analysis can be used to infer (the potential for) transcription factor activity [537]. Typically the input sequences are scanned against databases of known motifs, and significant motif enrichment is determined by a statistical test (e.g. Fisher's exact or hypergeometric tests) [535]. Motifs can also be discovered *de novo* without prior knowledge of motif sequences using tool like *HOMER*, that implement a hypergeometric test for over-representation of short k-mers against background. In an iterative process, the most significancy enriched motifs are identified by refinement of the k-mers [536].

### Annotation

As regulatory elements have an important role in gene regulation, the ability to assignment these, often distal, elements to target genes (i.e. annotation) emerges as an important analysis step for mechanistic understanding of gene regulation. *Proximity-based* annotation, i.e. assignment based on minimal distance in the 1D genomic space [536], is a wisely used approach. While conceptually and computationally simple, proximity annotation is restricted by an upper distance limit determined by the local gene density which obstruct identification of distal enhancers especially in gene-dense regions. *Correlation-based* approaches overcome the constrains set by the linear representation of the genome by instead using co-variation between element activity (e.g. chromatin accessibility) and gene expression across samples or cells to assign CRE-gene links [538]. While being useful for large multi omics data sets (e.g. scMulti-ome) [496, 539], the strategy is limited by the requirement of matching dataset of sufficient sample size to achieve statistical power for correlation analysis. Furthermore, correlation-based annotation is based on the assumption that chromatin accessibility (or alterative measure of element activity) is positively correlated with gene expression, an assumption that is not always valid [540]. In *Interaction-based* annotation, chromatin conformation capture data (e.g. Hi-C, HiCHIP/PLAC-seq) is used to link distal elements to putative target genes based on experimental evidence on physical proximity [121, 489, 541], enabling dynamic and cell-type specific annotation. However, interaction-based annotation is limited by the availability of biological relevant interaction-data of sufficient resolution, and it should also be noted that structural evidence of interaction does not guarantee functional activity. Finally *integrative approaches*, e.g. the Activity-by-Contact (ABC)

model, has been shown to be successful at predicting functional enhancer-gene links [542]. However model assumptions can introduce bias and the use of activating histone mark as a proxy for element activity limit the model to prediction of activating enhancer-gene relationship. For a conceptual representation of different annotation strategies see Figure 2.7.

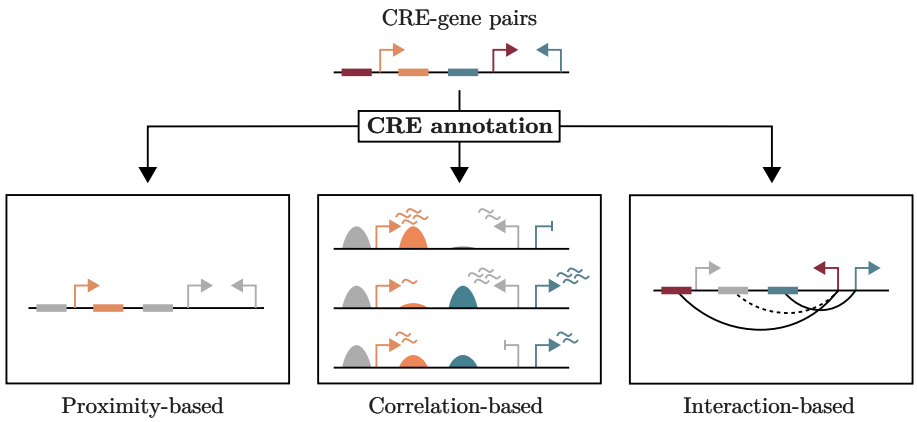



**Figure 2.7: Schematic illustration of different strategies for CRE annotation.** Proximity-based annotation assign CREs to the most proximal gene in the linear genome. In correlation-based strategies co-variance between CRE activity (e.g. chromatin accessibility) and gene expression is used for annotation. Interaction-based annotation use chromatin conformation capture data (e.g. Hi-C, HiCHIP/PLAC-seq) to identity putative target genes for distal CRE based on physical proximity in the 3D space.

## 2.5 Ethical considerations

While the use of laboratory animals raises ethical concerns and remains a subject of ongoing debate, it is undeniable that animal research has significantly contributed to advances in both fundamental biology and clinical medicine. All animal research in this thesis were conduced in accordance with ethical permits approved by the local ethics committee at Linköping and Lund University. The work adhered to the 3R principle minimizing animal suffering: replacing animal models where possible with alternative strategies, reducing the number of animals required through appropriate experimental design and statistical planning, and refining procedures to enhance animal welfare [543]. In this thesis the 3R principles has been implemented by substituting animal research with cell culturing or computational alternatives where feasible. Were mice models have been deems required (for harvest of primary cells or *in vivo* experiment), animals have been euthanized by approved methods (cervical dislocation or $CO_2$ asphyxiation) and where applicable, carefully monitored using predefined humane endpoints.

All data collection, processing, and analysis were performed with a commitment to research integrity and reproducibility. Experimental protocols and data analysis workflows have been documented and reported in detail. Developed bioinformatics pipelines and code related to data processing and figure generation (Paper II & IV) have been made freely available on github (`https://github.com/Tingvall`) for full reproducibility.

To support transparency and the broader scientific community, all datasets generated for this thesis, along with relevant metadata has been made publicly available on the Gene Expression Omnibus (GEO) repository upon publication.

# Chapter 3

# Aims of the thesis

The overarching aim of by this thesis was to unravel the gene regulatory mechanisms that control early lymphocyte development.

The specific aims for the individual projects were:

**Aim 1** Investigation of how lineage-specific TFs, in interplay with the chromatin landscape, drive lineage restriction and cell identity in normal lymphocyte development (*Paper I & IV*).

**Aim 2** Explore the link between lineage-restricted TFs networks and malignant transformation in B cell development (*Paper III*).

**Aim 3** Development of a bioinformatics tool to facilitate the use of chromosome conformation capture data in CRE annotation and GRN analysis (*Paper II*).

# Chapter 4

# Summary of included papers

The papers included in this thesis are all focused on gene regulatory mechanisms that control lineage restriction and establishment of cellular identity in early lymphoid development. This section provide a summary of each paper, with a schematic overview presented in Figure 4.1.

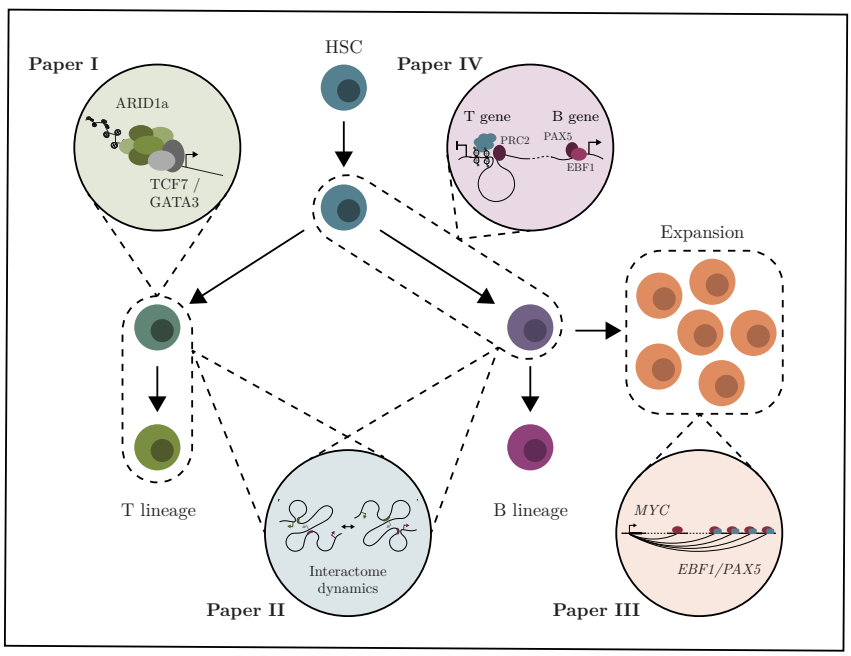



**Figure 4.1: Schematic overview of the papers.**

## 4.1 Paper I

### ARID1a associates with lymphoid-restricted transcription factors and has an essential role in T cell development

Development of lymphoid cells is controlled by stage and lineage-restricted TFs acting in concert with the chromatin landscape to regulate gene expression. To gain insights into this regulatory network, the interactome of a set of key regulators of lymphoid development were analyzed using BioID, an *in vivo* method based on proximity-dependent biotinylation [460, 462]. A substantial overlap in terms of proximity interacting partners (PXIs) were found for the lineage-restricted TFs. Among these shared PXIs, many proteins involved in the general transcription and chromatin remodeling machinery were found, as supported by enrichment of gene ontology categories related to transcriptional regulation, chromatin remodeling and histone modifiers. Several subunits of the BAF chromatin remodeling complex was found to interact with both T cell TFs (GATA3 and TCF7) and the B cell TF PAX5, with the subunit ARID1a [544, 545] being among the most prominently biotinylated proteins of this complex. To investigate the role of ARID1a in early lymphoid development, *Arid1a*$^{flox/flox}$ mice [387] were crossed with a human *CD2* promoter–driven Cre [388] strain, in order to cause deletion of *Arid1a* in early lymphoid progenitors. Investigation of the effect of *Arid1a* deletion in BM and spleen, revealed little to no effect on the early multipotent progenitors, as expected by the use of a lymphoid-restricted Cre strain, and only a modest effect in the B cell compartment. However, a dramatic reduction of the number of CD3+ cells in the spleen was found, indicating an important role of ARID1a in early T cell development. Cut&Run-based profiling of the DNA binding sites ARID1a in T cell progenitors, reveled a substantial overlap with TCF7 and GATA3 occupancy, providing an independent line of support for ARID1a as a co-factor for key T cell TFs.

Given the putative role for ARID1a as a co-factor for T cell restricted TFs and the dramatic reduction of peripheral T lymphocytes, investigation of the role of ARID1a in T cell development in the thymus was warranted. The essential role for ARID1a in thymocyte development was confirmed by a 95% reduction in thymic cellularity in *Arid1a*-deficient mice, reflected in the ability to generate more mature DP, CD4SP and CD8SP populations. Investigation of the earlier stages of T cell development reveled a relative accumulation of cells at the DN3 state in *Arid1*$^{-/-}$ mice, followed by a considerable reduction of the more mature DN4 cells. The developmental arrest in the *Arid1*$^{-/-}$ mice at the DN3 stage coincides with $\beta$-selection [546], in line with the finding that rearrangement of the TCR$\beta$-gene is dependent on a functional BAF complex [547].

However, investigation of recombination status and expression of VDJ encoding TCR$\beta$ failed to detect any considerable differences between ARID1a deficient and control DN3 cells, arguing against failed TCR$\beta$ recombination or expression as a consequence for the observed developmental block. Because ARID1a is part of a chromatin remodeling complex, we next decided to investigate potential ARID1a associated epigenetic alterations using ATAC-seq. Comparison of the chromatin accessibility between *Arid1a*$^{-/-}$ and *Wt* DN3 cells, revealed that among the ∼2300 dynamically regulated sites in the DN3-to-DN4 transition, approximately a third displayed lower accessibility in the absence of ARID1a. Proximity based annotation identified several ARID1a dependent putative regulatory elements at genes with importance in T cell development (e.g. *CD8s*, *CD44*, *Rag2* and *Lmo2*), providing support for the role of ARID1a as an important factor in the establishment of a normal epigenetic landscape in DN3 cells.

To summarize, out work in *Paper I* has identified ARID1a, a subunit of the BAF chromatin remodeling complex, as en essential factor for T cell development. ARID1a associate with the T lineage restricted TFs TCF7 and GATA3 and is required for the establishment the epigenetic landscape associated with transition from the DN3 to DN4.

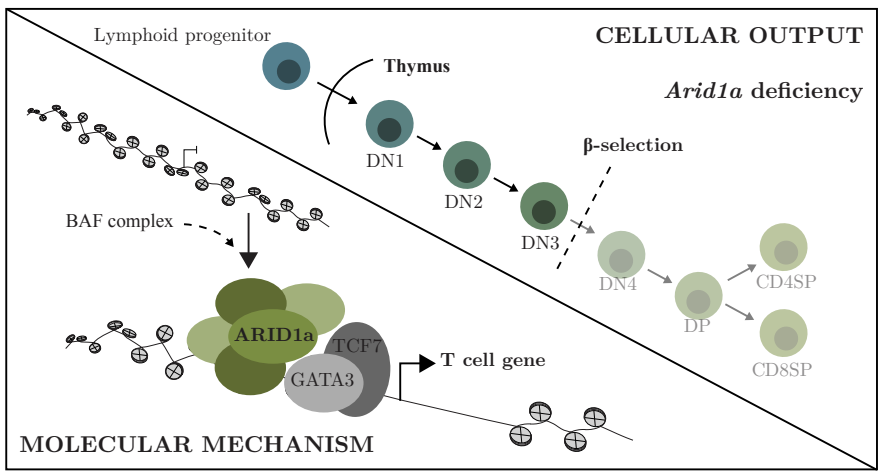



**Figure 4.2: Schematic representation of the role of ARID1a in T cell development.** ARID1a, a subunit of the BAF chromatin remodeling complex, interacts with the T lineage restricted TFs GATA3 and TCF7 and enable chromatin accessibility at elements associated with T lineage genes. ARID1a deficiency lead to impaired T cell development, and a developmental block at the DN3 stage, coinciding with $\beta$-selection.

## 4.2 Paper II

### Chromatin interaction-based annotation of cis-regulatory elements reveals highly dynamic promoter-enhancer interactions in lymphocyte development

Stage- and lineage-specific gene expression patterns are controlled by a complex interplay between TFs, the epigenetic landscape, and the 3D organization of the genome. The 3D structure allows for the formation of DNA loops that juxtaposition distal cis-regulatory elements (CREs) to target promoters, allowing for context dependent control of gene expression [112, 117, 121, 548–552]. Target genes assignment (i.e. annotation) is an important process in the functional interpretations of non-coding regions and their role in gene regulation. Current standards for CRE annotation are often based on the concept of proximity, defined as the distance between the CRE and the target gene in the linear DNA. Thus, these methods are unable to incorporate the dynamics of CRE usage and are restricted by an upper distance limit decided by the local gene density. With advancements in chrosomsome conformation capture methods (i.e. Hi-C, HiChIP), interaction-based annotation emerges as an alternative approach that by shifting from a 1D- to 3D-based view of distance can bypass the limitations of proximity-based methods.

To facilitate the exploration of complex gene regulatory networks based on chromosome configuration data, the easy-to-use tool ICE-A was developed. ICE-A implement a combined proximity- and interaction based approach to improve target gene assignment of distal elements, while still counteracting the negative aspects associated with use of predefined interactions for peak annotation (i.e. bin size and minimum distance for interaction calling). In addition, ICE-A provide options for data integration and visualization. Benchmarking ICE-A against proximity based methods in two independent datasets from different cellular context, a CRISPR-FlowFISH validated enhancer-gene pairs in a leukemia cell line [542] and a set of enhancers essential for cancer fitness in a colon cell line [553], reveled that ICE-A outperformed the proximity-based methods in target gene assignment of more distally located enhancers with maintained precision.

Using ICE-A for exploration of the enhancer-promoter dynamics in early lymphocyte development reveled that lineage-restricted TFs target regulatory elements annotated to both lineage-specific and broadly expressed genes. Among the CREs assigned to lineage-specific genes, we also observed context dependent interactions with alternative promoters revealing a highly dynamic landscape of enhancer-promoter interplay in lymphocyte development. Furthermore, inves-

tigation of the interactome in B cell progenitors deficient in the key B-lineage restricted TF EBF1 reveled an interactome intermediate to those observed in the pro-B and pro-T cells, characterized by the presence of lineage-restricted enhancer-promoter interactions for key genes in both B- and T cell development (e.g. *Gata3*). These data underline the complexity of gene regulatory networks in lymphoid development and pinpoint the importance of EBF1 in establishment of the chromatin landscape during B cell development.

In summary, in *Paper II* we developed a bioinformatics tool for chromosome configuration data based annotation of cis-regulatory elements. Using this tool for exploration of gene regulatory networks in early lymphocyte development, revealed a highly dynamic interactome and an important role for the EBF1 in B cell associated enhancer-promoter interactions. Furthermore, our work demonstrate how efficient annotation procedures for linking distal regulatory elements to target genes can provide valuable insights into gene regulatory networks.

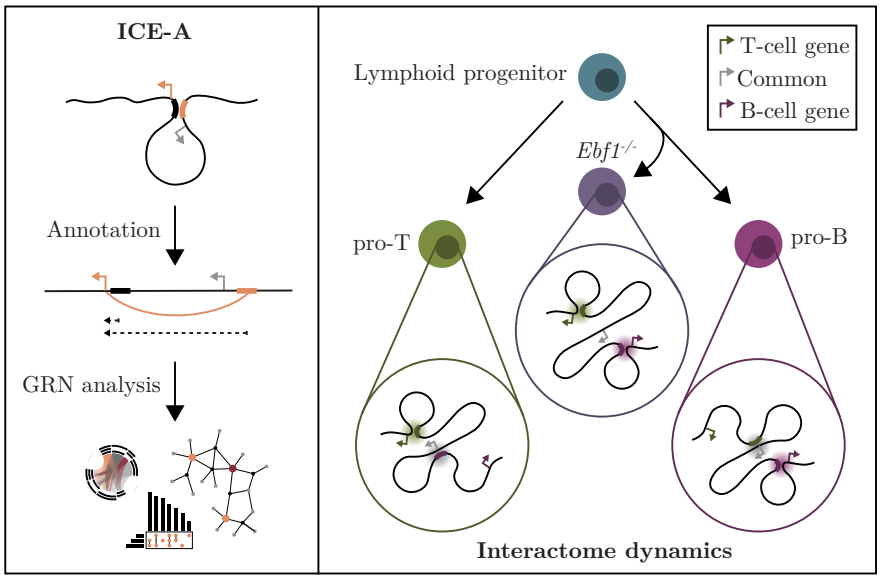



**Figure 4.3: Schematic representation of ICE-A and the interactome dynamics in lymphocyte development.** ICE-A can facilitate GRN exploration by incorporation of chromosome conformation data in CRE-annotation. ICE-A based analysis reveled a highly dynamic enhancer-promoter landscape in early lymphocyte development, with a key role of EBF1 in the establishment of the B cell restricted chromatin landscape.

## 4.3 Paper III

### EBF1 and PAX5 control pro-B cell expansion via opposing regulation of the Myc gene

The generation of lineage restricted and finally commited progenitors during hematopoiesis is tightly regulated by TF networks. However, several of these stage- and lineage restricted TFs are frequently mutated in leukemias, suggesting a direct role of TF network in malignant transformation [554]. In B cell development, two key regulators that are frequently mutated in B cell acute lymphoblastic leukemia (B-ALL) are EBF1 and PAX5 [337, 338, 555, 556]. EBF1 is critical for lineage specification and survival of B cell progenitors [300, 557, 558], and by using a system that allowed us to dynamically control EBF1 activity, we found that loss of functional EBF1 cause an impaired pro-B cell expansion as a result of cell cycle arrest. In addition, EBF1 loss was associated with reduced expression of the proto-oncogene *Myc*, providing a possible link between EBF1 and MYC in the control of B cell proliferation. The ability of ectopic expression of *Myc* to partially rescued B cell expansion in the absence of EBF1 both *in vitro* and *in vivo*, provided further support for direct EBF1-mediated regulation of *Myc*. It should however be noted that the role of EBF1 in promoting B lineage progression could not be substituted by *Myc* expression.

The regulation of *Myc* is highly complex and regulatory elements located as far as ∼1.7 Mbp from TSS have been reported [140]. Integration of EBF1 ChIP-seq data with chromatin interaction data from H3K4me3 PLAC-seq allowed us to identify several distal EBF1 bound element interacting with the *Myc* promoter in a B cell progenitor context, including an EBF1 occupied region 0.7 MBp upstream of TSS as well as several binding sites in the blood enhancer cluster (BENC) superenhancer region located ∼1.6-1.7 MBp from the *Myc* promoter. EBF1-responsiveness of the bound elements could be confirmed by luciferase reporter assay. Furthermore, functional validation by CRISPR-Cas9–mediated targeting of the putative regulatory elements identified one EBF1 binding site in the BENC regions of key importance for *Myc* expression and pro-B cell expansion.

Closer investigation of the mechanism behind EBF1 dependency reveled that the EBF1 requirement was most pronounced in CD19+ B cell progenitors. *CD19* expression is linked to PAX5, a TF that has previously been suggested to suppress *Myc* expression [359, 559]. Profiling of PAX5 binding sites in pro-B cells by ChIP-seq reveled that several of the EBF1-responsive elements in the *Myc* locus was also targeted by PAX5. Ectopic expression of PAX5 in EBF1 defi-

cient cells resulted in reduced cell expansion and $G_0$ arrest. Gene expression profiling, reveled reduced *Myc* expression levels and negative enrichment of the MYC-associated transcriptional program, further supporting the idea that *Pax5* functions as a negative regulator of *Myc* expression in normal B cell progenitors. Binding of both EBF1 and PAX5 to distal regulatory elements in the human *MYC* locus in a B-ALL cell line indicates that the EBF1:PAX5:MYC regulatory loop is conserved and may control normal and malignant B cell development also in the human context.

Taken together, out work in *Paper III* found that opposing role of EBF1 and PAX5 in the regulation of *Myc* expression, establishes a regulatory loop that control the expansion of B cell progenitors in mice. Given that these TFs are frequently mutated in B cell leukemias and that the EBF1:PAX5:MYC regulatory loop appears to be conserved in the human setting, the mechanism identified in this study could have relevance in both normal and malignant B cell development.

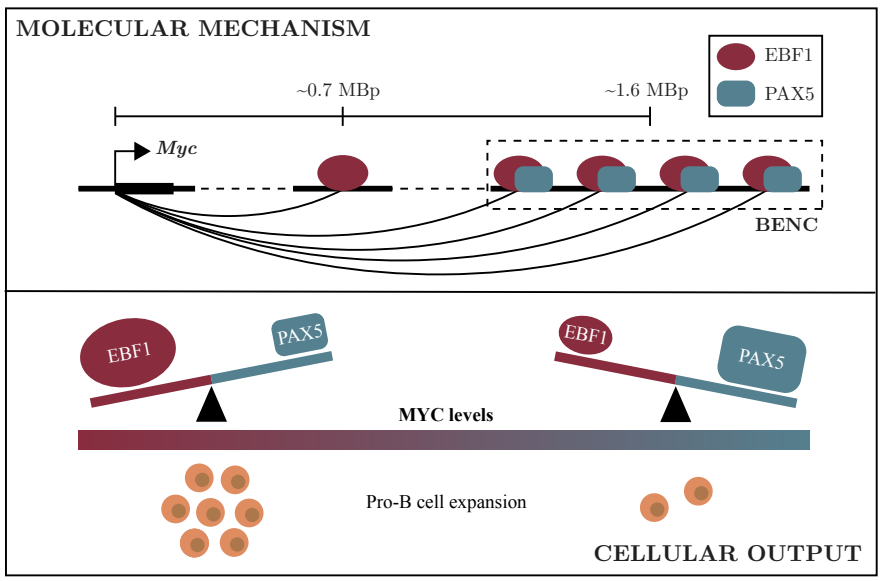



**Figure 4.4: Schematic representation of EBF1 and PAX5 mediated regulation of *Myc*.** Opposing regulation of Myc by EBF1 and PAX5, establishes a regulatory loop controlling pro-B cell expansion.

## 4.4 Paper IV

### B-lineage commitment is dependent on a reversible epigenetic switch

To improve our understanding of the interplay between transcription factor network and the epigenetic landscape in B cell development, we conduced single cell ATAC/RNA-seq of broad overlapping BM progenitor populations to obtain a high resolution trajectory of early B cell development. Trend change analysis reveled two stages associated with dramatic shift in DNA accessibility, indicative of an epigenetic switches (ES). While the second ES (ES2) likely reflect the shift into a proliferative state after successful pre-BCR assembly, the ES1 was found to be linked to B cell commitment, with repression of alternative lineage priming and acquisition of a B cell associated epigenetic landscape.

This "Big bang" of B cell development strongly correlated with the functional activity of the TFs EBF1 and PAX5. Although the increased accessibility could be directly liked to these factors, EBF1 and PAX5 rarely target elements that displayed decreased accessibility. Still these factors have a crucial, dose-dependent role in B cell lineage restriction as indicated by the aberrant expression of T-lineage associated genes and functional plasticity in pro-B cells with combined transheterozygous deletion of *Ebf1* and *Pax5* (*TH*) [333]. Within this plastic *TH* pro-B population, we detected a stable CD25 (*Il2ra*) expressing subpopulation, with increased kinetics in T lineage conversion and a small but significant increase in expression of *Tcf7*. Enforced expression of TCF7 was sufficient to drive the normally committed *Wt* pro-B cells into a T lineage fate in a Notch dependent manner, highlighting the importance of this factor in T lineage development. However, the inability to detect TCF7 protein expression in the *TH* pro-Bs argue against an instructive role of TCF7 in lineage placticity, but rather point toward an inherent epigenetic instability and reduced repressive capacity.

Closer examination of the *Tcf7* locus revealed a bivalent chromatin state in the promoter region, with reduced level of the repressive histone mark H3K27me3 in *TH* cells, suggestive of a role of EBF1 and/or PAX5 in polycomb-mediated repression of alternative lineage fates. Indeed, PAX5 has been shown to interact with EZH2, the catalytic subunit of the PRC2 complex [331, 332], and a general reduction of H3K27me3 levels was found in *TH* cells at element linked to T-lineage genes. While direct targeting of the TFs at EBF1/PAX5 responsive element were rare, a majority of the element were found to interact with chromosomal distal (CD) regions bound by EBF1 and/or PAX5 consistent with a recruitment via looping mechanism for EBF1/PAX5 mediated repression of

alternative lineage genes. Investigation of the EZH2 occupancy patterns reveled that EZH2 binding were far more prominent in dynamic H3K27me3 elements compared to elements displaying comparable H3K27me3 levels between *Wt* and *TH*. The instructive role of B cell TFs in polycomb repression was further substantiated by the discovery that many of the EZH2 binding sites were PAX5 dependent, with an enrichment of dynamic H3K27me3 elements among the PAX5 responsive EZH2 targets, including the *Tcf7* promoter region.

Lastly the functional importance of H3K27me3 deposition on alternative lineage genes was investigated by treatment with the EZH1/2 inhibitor Valemetostat. A rapid decrease in overall H3K27me3 levels was detected along with upregulation of both *Tcf7* and *Gata3* transcription in CD19+ B cell progenitors. Reseeding of cells treated with Valemetostat on OP9-DL1 for 14 days revealed that transient inhibition of EZH activity was sufficient to drive T lineage conversion in response to Notch signaling in normally committed cells.

Collectively the data represented in *Paper IV* show that B cell commitment is dependent on a transcription factor-mediated, dose-dependent epigenetic switch, that involve reversible polymcomb mediated repression of inherent T-lineage potential in early lymphoid progenitors.

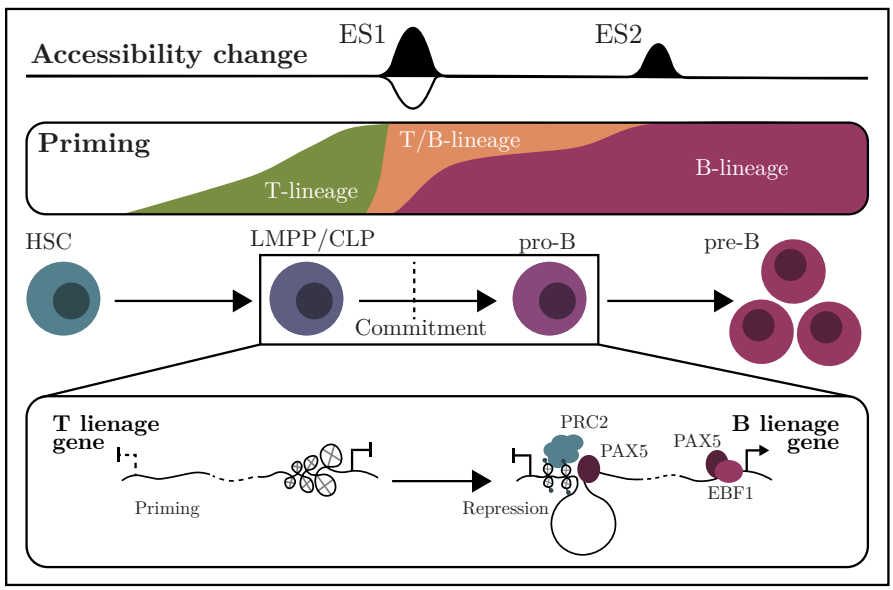



**Figure 4.5: The "Big Bang of B cell development.** Schematic representation of the epigenetic switch that control B cell commitment by TF mediated polycomb repression of alternative lineage potential.

# Chapter 5

# Discussion

***Locking in cellular fate by an epigenetic switch***

The transition from multipotent progenitors to stable lineage commitment is regulated by complex gene regulatory networks, made up of highly intertwined TF circuits. The architecture of these circuits enable temporal and quantitative control of the transcriptional program in a stage dependent manner. The TF networks act in concert with the epigenetic landscape, clearly illustrated for both T (*Paper I*) and B (*Paper IV*) cell development. In the B cell context, we show that this interplay result in an epigenetic switch that establishes stable B cell identity. Although this switch is linked to a global increase in accessibility in regulatory elements of lineage defining genes, the observed polycomb mediated repression seemed to be more highly correlated with functional B cell commitment as indicated by the fact that transient inhibition of the catalytic activity of PRC2 was sufficient to revere lineage commitment in B cell progenitors. The PRC complexes has previously been shown to drive switch like transition in other biological contexts [560, 561], suggesting that this might be a conserved mechanism for cellular processes that require the cell to rapidly adopt another cellular state. Whereas opening of B cell associated genes was almost non existent before *Ebf1* and *Pax5* activation, we detected a epigenetic priming of alternative lineages including T cells in the uncommitted lymphoid progenitors, providing a possible explanation why B cell commitment is so tightly linked to repression. In addition, the fact that B cell commitment take place in the same spatial location as the early development may require a cell intrinsic mechanism for suppression of alternatives fates, in contrast to T cells that can rely on Notch signaling provided by the thymic environment to look in stable cell identity. T cell development, on the other hand, was found to instead depend on the BAF complex to actively open chromatin in order to secure proper identity, although at a later developmental stage. The low number of cells detected in the actual transitional state in B cell commitment indicate that the switch

is rapid once triggered. A putative mechanistic explanation for this would be condensate formation, which has been known to display a switch like behavior once a threshold of protein concentration is reached. For EBF1 a similar mechanisms has been shown for BRG recruited condensate formation [221]. Although condensate formation would be a feasible mechanism, especially considering the lack of direct targeting of EBF1 and PAX5 at the dynamic H3K27me3 elements, further studies is needed to unravel the exact mechanism of the EBF1 and/or PAX5 mediated repression in B cell development, and how similar mechanisms might be at play in other cellular contexts.

***Transcriptional repression beyond closed chromatin***

Although transcriptional activation is of course crucial for lineage defining programs, repression represent a far less studied but equally important mechanism in cell fate determination, as highlighted in *Paper IV*. Often chromatin accessibility is used as a proxy for CRE activity, where an increased accessibility is thought to indicate activation while repression is associated with a closed chromatin state. Although this is of course true in many situations, it is an over simplification that has the potential to limit our understanding of gene regulatory mechanisms. First, transcriptional activity is not necessarily correlated with CRE accessibility, as shown by the many alternative lineage genes retaining an open chromatin state past B cell commitment or the epigenetic priming prior to expression of lineage specific genes. And second, even if a change in chromatin accessibility is in fact observed, its direction does not guarantee a certain transcriptional output, as exampled for EBF1 targets in *Paper II* and *IV*. Although EBF1 can exert both activating and repressive impact on target expression, its impact on local chromatin accessibility is almost exclusively positive. In the case of EBF1 this is likely a result of its pioneering activity and ability to mediate target gene repression via recruitment of co-repressor complexes. However, the fact that negative change in transcription can be mediated either via direct repression or simply through disruption of transcriptional activation make it inherently more complex. In addition to co-repressor recruitment for making the epigenetic landscape less permissible, the TF mediated repression can also be achieved via steric hindrance, competition of co-factors or binding sites or hijacking of regulatory elements via looping to name a few examples. Considering the highly dynamic enhancer-promoter interactome as seen in *Paper II*, with context dependent interaction between lineage specific enhancers and both lineage specific and broadly expressed genes, these genome topology driven mechanisms of gene-regulation could represent a overlooked mechanisms in cell fate determination. Also here lineage specific TFs can have an important regulatory role, to both guide the establishment a 3D landscape permissive for lineage programs by providing topological constraint for control of aberrant

expression patterns. This is exemplified by the altered enhancer–promoter landscape in *Ebf1*-deficient B cell progenitors, which display an interaction profile intermediate that of normal B and T cell progenitors.

***Cell fate determination is coupled to basic cellular processes***

It has previously been described that PAX5 and IKZF1 can regulate energy demand in a stage dependent manner [344, 345]. In *Paper II* this integration of the lineage programs with the basic cellular processes is further explored, with specific focus on EBF1. Indeed, also EBF1 was shown to regulate commonly expressed genes, with a functional implication in B cell expansion and survival. In a subset of these genes, the EBF1 dependence was only observed upon sudden EBF1 removal, indicating that EBF1 can take control of the normal regulatory circuits controlling these genes. Although this TF mediated addiction can provide a vulnerability, the integration of basic processes into developmental networks can couple survival to successful differentiation. A mechanism that may be of extra importance in lymphoid development which if marked by alternating stages of rapid proliferation and recombination, leaving the cell susceptible to potential disease causing genetic alterations. An example of an EBF1 addicted target gene, the proto-oncogene *Myc*, is presented in *Paper III*. Here the opposing regulation by EBF1 and PAX5 in the expression of *Myc*, established an incoherent regulatory loop that control B cell expansion via a distal regulatory enhancer regions. Direct links between critical regulators of lineage fate and expression of a gene in control of survival and proliferation, could provide an important mechanism in preservation of lineage identity and ensure ordered differentiation. Although identified in mice, the regulatory loop appear to be conserved in humans, with potential implications in leukemia. Mutations in *Pax5*, which is frequently detected in B-ALL, could disrupt this regulatory loop causing EBF1 to super-activate *Myc*. Notably, MYC-induced B cell malignancies frequently display lineage plasticity [562], suggesting a disrupted B cell program. The broad occupancy pattern of lineage specific TFs on elements associated with ubiquitous genes (*Paper II*), suggest that this is not an isolated occurrence and that lineage specific TFs integrate into the regulatory circuits of basic cellular processes, with implication both to normal and malignant development.

***TF dose is critical in both normal and malignant development***

Although TF activity is often considered in binary terms, its precise dosage plays a crucial role in shaping the regulatory networks that control biological processes. For lineage specification this is clearly exemplified by the observation that heterozygous loss of several of the critical B cell regulators can cause impaired development [205, 563–566]. Numerous mechanistic explanation for the

sensitivity of TF dose has been proposed, including altered occupancy pattern (based on affinity or partners) or disruption of GRNs. Disruption of regulatory networks seems to at least party explain the dose effect in *Paper IV*, as indicated by the disrupted B cell program and aberrant expression of progenitor and T lineage genes by heterozygous deletion of *Ebf1* and/or *Pax5*. While single heterozygous deletions seems to be sufficient for at least partly disrupting the TFs activating potential, highlighting the distinct roles for EBF1 in cell cycle regulation and PAX5 for B lineage identity, combined dosage reduction was required for disruption of B cell commitment. This may potentially be related to the positive feed-back loop between EBF1 and PAX5 which can the reinforce B cell identity, de-stabilization of the epigenetic state or by mutual requirement for the two TFs in PRC2 mediated repression. In addition to lineage placticity, the GRN disruption by combined *Ebf1* and *Pax5* heterozygous loss can cause malignant transformation in mice [339]. Also in human B-ALL are multiple genetic aberrations a frequent occurrence. In some cases associated with lineage placticity, indicating that TF dose is important for maintaining cell identity not only in normal development but also malignant transformation. Just like EBF1/PAX5, many of these factors are involved in self-enforcing feedback loops. While these circuits are efficient for driving stable cell fates, they can also enhance dose effects with consequences for the entire GRN. *Paper III* provide an potential mechanistic example of how altered lineage specific TF dosage can make cells vulnerable for leukemic transformation. In the EBF1/PAX5/MYC regulatory circuit, reduced dose of PAX5 may weaken the repression of the proto-oncogene, thereby allowing uncontrolled expansion of B cell progenitors. A mechanism that would be consistent with the frequent detection of *Pax5* heterozygous mutations in B-ALL [337, 338]. Another regulatory mechanism that has been shown to be dose dependent is the TFs ability to bypass chromatin barriers [253]. Considering the dramatic shift in chromatin accessibility that EBF1 and PAX5 can induce (*Paper IV*), this could be a contributing mechanism for the observed dosage effect. All these importance dose effects also warrant a reflection regarding the often rather arbitrary cutoff values for fold change used in differential expression analysis, as many small but still biologically relevant changes are most likely excluded. Lastly, it should also be noted that the discussions here are mostly restricted to transcriptional dosage, however, functional dose effect of TF activity can also be mediated by post-translational regulation or modulated degradation kinetics. Irrespective of mechanism, it is clear that TF dosage is an important factor in the intricate regulatory network that control cellular identity, and closer investigation is needed to unravel the exact mechanisms behind individual and combinatorial dosage effects in normal and malignant development, as well as its interplay with the dynamic chromatin landscape.

# Acknowledgments

There are many people that have contributed to making these 6 years as enjoyable as they were, and here I will attempt to mention most of them.

**Micke,** you were the perfect supervisor for me. Under you I learned a lot about the B cell biology and molecular techniques. I want to thank you for all the freedom and trust you gave me in pursuing my projects which gave me a much appreciated understanding of the entire workflow. It is always exciting to hear about your latest ideas and (more or less feasible) experimental strategies. And most importantly, thank you for educating me in the distinguished field of Californian IPAs. **Jonas,** thank you for all your hands-on help in the beginning of my PhD and all the jokes during long days in the lab. Also thanks for forcing some well-needed stress relief on me through all those intense workouts. Although I'm not sure I miss the near eternal Christmas music, I do really miss the scientific discussions that could turn into who can do the most burpees. **Göran,** although we have not interacted much during these years, I've always appreciated our discussions and your help once we do. **Tina and Maria,** thank you both for all the help throughout the years, it was very valuable!

**Tobias, Lotta, Rajesh and Josefin,** thank you all for introducing me to the field and helping me realize how interesting it is. Extra thanks to Tobias for his role in supervising me during my master thesis. **Helen, Erik, Kim, Vatsha, Lopa, Jörgen, Dirk and Johanna R,** it's been really nice to be part of a proper group again. Big thanks for all your help with generating the single cell data (especially Helen for the unreasonably long days). This work became a central part of my main project and I couldn't have done it without you. Although I haven't spent so much time in Linköping, I appreciate all our discussions. **Linda,** thank you for all the help with mice, I have often wished that I could clone you and keep one copy in Lund! **LU/LiU FACS cores,** a lot of my work has utilized FACS and I've really appreciated all your help and expertise. **Shamit and Stefan,** it's been great to have some fellow bioinformaticians around, thank you. **David,** thank you for letting me sneak in to your lab meetings and for the (more or less scientific) discussions.

**Julie**, my fellow fireball girl, thanks for being the best possible SSC roommate, for all our late night escapades and for being my companion in beer-fueled 'work' sessions, which were sometimes productive but always enjoyable. **Rasmus,** thank you for being an excellent office mate, your contribution to B12 culture through Whisky Thursdays$^{TM}$ together with **Micke So.** and of course all the fun outside the office. **Sara,** thank you for all the fun times and adventures. It's been great to have someone who can (almost) keep up with my "brilliant" ideas and also contribute with your own! **Elena,** you always light up any occasion with your positive personality and great laugh. Extra thanks for being such a nice travel companion. **Sandro,** thank you for bringing color to BMC and making me never have to try skin Fernet again. **Mo,** the mysterious Mo, you never know when you're gonna show up, but it's always a happy surprise when you do! **Oscar, Niklas, Micke So, Parashar, Hugo, Ania, Ariana, Qinyu, Dang and the rest of B12,** thank you for making BMC what it is, I've had a lot of fun with all of you!

This PhD would not have been the same without the BMC gym and sauna. **Jonas,** thank you for starting the 750/Lion King workouts and for trusting me with taking over the responsibility, they kept me sane during all these years! **The Lion King Gang (Julie, Anna R, Mo, Sandro, Micke So, Hugo, Sara, Elena, Anna P, Roger, Els, Marija, Robi, Theresa, Stefan and the rest),** thank you for all the shared suffering, Roxannes and rehydration in the Sauna. And of course, thank you **Julie, Anna R** and all the other people who helped save the gym, including the 511 people who signed the petition.

**Rugby people,** sometimes when doing a PhD you really want to focus on something other than work, and is there any better way to distract yourself than trying to get a ball to the other side of a field with a lot of people violently attempting to stop you? Well, that would maybe be a rugby party...

**Extended 42 (Elin, Sonen, Linnea, Tove, Hak, Finn, Pigge, Sebastian, Erik, Jonna, David, Johan∗2 and Christoffer),** just like the group's namesake, you have the answers to everything, however, you never know what the questions might be. Such as how does a breast pump feel? Or what's Polands best energy drink for curing hangovers in preparation for less-than-safe water parks? Where to find the best kvirkels or how to mix a semmeldrink? And many many more... You made the Chalmers time amazing and every time we meet, all problems melt away and it feels like we're students again! **Angelo,** like a big brother you showed me the way through academia's messy jungle, gave me amazing opportunities, was an inspiration throughout my studies and all the while we had a lot of fun together! **The rest of "the Gang" (i.e. Gia**

**& Ciffe),** for all the nice quizzes, and games that Ciffe somehow always won. **Maria, Josefin, Linnea, Elma & Adella,** it's always really fun to meet and hear what's new in your lives!

**Carin and Lars,** thanks for taking my mind of failed experiments with outdoor adventures and beers at Ölstugan and Gyllene Prag. For efficient recoveries with good food and drink at Torpet, and for early on preparing me for an academic career through the introduction of attoseconds in kindergarten and photosynthesis as a bed time story. **Mormor,** I really appreciate the times we spent together on all my trips to Östergötland. **Max,** you know what you've done, thanks for all the floff! **Jakob,** thanks for putting up with me and for all our planned or accidental ventures, whether it is sunrise Mario Kart competitions, conquering sketchy dutch playgrounds, camping during a military drill, or enjoying floor lobster, I always have fun with you!

# References


1. Orkin, S. H. *Diversification of haematopoietic stem cells to specific lineages* 2000. doi:10.1038/35049577.

2. Weissman, I. L. *Stem cells: Units of development, units of regeneration, and units in evolution* 2000. doi:10.1016/S0092-8674(00)81692-X.

3. Notta, F. *et al.* Distinct routes of lineage development reshape the human blood hierarchy across ontogeny. *Science* **351.** doi:10.1126/science.aab2116 (2016).

4. Velten, L. *et al.* Human haematopoietic stem cell lineage commitment is a continuous process. *Nature Cell Biology* **19,** 271–281. doi:10.1038/ncb3493 (2017).

5. Laurenti, E. & Göttgens, B. From haematopoietic stem cells to complex differentiation landscapes. *Nature* **553,** 418–426. doi:10.1038/nature25022 (2018).

6. Wilson, A. & Trumpp, A. *Bone-marrow haematopoietic-stem-cell niches* 2006. doi:10.1038/nri1779.

7. Morrison, S. J. & Kimble, J. *Asymmetric and symmetric stem-cell divisions in development and cancer* 2006. doi:10.1038/nature04956.

8. Morrison, S. J., Wandycz, A. M., Hemmati, H. D., Wright, D. E. & Weissman, I. L. Identification of a lineage of multipotent hematopoietic progenitors. *Development* **124,** 1929–1939. doi:10.1242/dev.124.10.1929 (1997).

9. Kiel, M. J. *et al.* SLAM family receptors distinguish hematopoietic stem and progenitor cells and reveal endothelial niches for stem cells. *Cell* **121,** 1109–1121. doi:10.1016/j.cell.2005.05.026 (2005).

10. Kondo, M., Weissman, I. L. & Akashi, K. Identification of clonogenic common lymphoid progenitors in mouse bone marrow. *Cell* **91,** 661–672. doi:10.1016/S0092-8674(00)80453-5 (1997).

11. Akashi, K., Traver, D., Miyamoto, T. & Weissman, I. L. A clonogenic common myeloid progenitor that gives rise to all myeloid lineages. *Nature* **404,** 193–197. doi:10.1038/35004599 (2000).

12. Adolfsson, J. *et al.* Identification of Flt3+ lympho-myeloid stem cells lacking erythro-megakaryocytic potential: A revised road map for adult blood lineage commitment. *Cell* **121,** 295–306. doi:10.1016/j.cell.2005.02.013 (2005).

13. Pronk, C. J. *et al.* Elucidation of the Phenotypic, Functional, and Molecular Topography of a Myeloerythroid Progenitor Cell Hierarchy. *Cell Stem Cell* **1,** 428–442. doi:10.1016/j.stem.2007.07.005 (2007).

14. Wilson, A. *et al.* Hematopoietic Stem Cells Reversibly Switch from Dormancy to Self-Renewal during Homeostasis and Repair. *Cell* **135,** 1118–1129. doi:10.1016/j.cell.2008.10.048 (2008).

15. Pietras, E. M. *et al.* Functionally Distinct Subsets of Lineage-Biased Multipotent Progenitors Control Blood Production in Normal and Regenerative Conditions. *Cell Stem Cell* **17,** 35–46. doi:10.1016/j.stem.2015.05.003 (2015).

16. Iwasaki, A. & Medzhitov, R. *Control of adaptive immunity by the innate immune system* 2015. doi:10.1038/ni.3123.

17. Puhr, S., Lee, J., Zvezdova, E., Zhou, Y. J. & Liu, K. *Dendritic cell development-History, advances, and open questions* 2015. doi:10.1016/j.smim.2016.03.012.

18. Wilson, N. K. *et al.* Combined Single-Cell Functional and Gene Expression Analysis Resolves Heterogeneity within Stem Cell Populations. *Cell Stem Cell* **16,** 712–724. doi:10.1016/j.stem.2015.04.004 (2015).

19. Buenrostro, J. D. *et al.* Integrated Single-Cell Analysis Maps the Continuous Regulatory Landscape of Human Hematopoietic Differentiation. *Cell* **173,** 1535–1548. doi:10.1016/j.cell.2018.03.074 (2018).

20. Giladi, A. *et al.* Single-cell characterization of haematopoietic progenitors and their trajectories in homeostasis and perturbed haematopoiesis. *Nature Cell Biology* **20,** 836–846. doi:10.1038/s41556-018-0121-4 (2018).

21. Weng, C. *et al.* Deciphering cell states and genealogies of human haematopoiesis. *Nature* **627,** 389–398. doi:10.1038/s41586-024-07066-z (2024).

22. Zhang, X. *et al.* An immunophenotype-coupled transcriptomic atlas of human hematopoietic progenitors. *Nature Immunology* **25,** 703–715. doi:10.1038/s41590-024-01782-4 (2024).

23. Safina, K. & van Galen, P. *New frameworks for hematopoiesis derived from single-cell genomics* 2024. doi:10.1182/blood.2024024006.

24. Carrelha, J. *et al.* Hierarchically related lineage-restricted fates of multipotent haematopoietic stem cells. *Nature* **554,** 106–111. doi:10.1038/nature25455 (2018).

25. Drissen, R. *et al.* Distinct myeloid progenitor-differentiation pathways identified through single-cell RNA sequencing. *Nature Immunology* **17,** 666–676. doi:10.1038/ni.3412 (2016).

26. Tusi, B. K. *et al.* Population snapshots predict early haematopoietic and erythroid hierarchies. *Nature* **555,** 54–60. doi:10.1038/nature25741 (2018).

27. Drissen, R., Thongjuea, S., Theilgaard-Mönch, K. & Nerlov, C. Identification of two distinct pathways of human myelopoiesis. *Science Immunology* **4.** doi:10.1126/sciimmunol.aau7148 (2019).

28. Meng, Y. *et al.* Epigenetic programming defines haematopoietic stem cell fate restriction. *Nature Cell Biology* **25,** 812–822. doi:10.1038/s41556-023-01137-5 (2023).

29. Mikkola, H. K. & Orkin, S. H. *The journey of developing hematopoietic stem cells* 2006. doi:10.1242/dev.02568.

30. Galloway, J. L. & Zon, L. I. *3 Ontogeny of hematopoiesis: Examining the emergence of hematopoietic cells in the vertebrate embryo* 2003. doi:10.1016/s0070-2153(03)53004-6.

31. Palis, J. & Yoder, M. C. *Yolk-sac hematopoiesis: The first blood cells of mouse and man* 2001. doi:10.1016/S0301-472X(01)00669-5.

32. Böiers, C. *et al.* Lymphomyeloid contribution of an immune-restricted progenitor emerging prior to definitive hematopoietic stem cells. *Cell stem cell* **13,** 535–548. doi:10.1016/j.stem.2013.08.012 (2013).

33. Medvinsky, A. & Dzierzak, E. Definitive hematopoiesis is autonomously initiated by the AGM region. *Cell* **86,** 897–906. doi:10.1016/S0092-8674(00)80165-8 (1996).

34. Alvarez-Silva, M., Belo-Diabangouaya, P., Salaün, J. & Dieterlen-Lièvre, F. Mouse placenta is a major hematopoietic organ. *Development* **130,** 5437–5444. doi:10.1242/dev.00755 (2003).

35. Maximow, A. Untersuchungen über Blut und Bindegewebe: I. Die frühesten Entwicklungsstadien der Blut- und Bindegewebszellen beim Säugetierembryo, bis zum Anfang der Blutbildung in der Leber. *Archiv für mikroskopische Anatomie* **73,** 444–561. doi:10.1007/BF02979896 (1908).

36. Bowie, M. B. *et al.* Identification of a new intrinsically timed developmental checkpoint that reprograms key hematopoietic stem cell properties. *Proceedings of the National Academy of Sciences of the United States of America* **104,** 5878–5882. doi:10.1073/pnas.0700460104 (2007).

37. Morrison, S. J. & Scadden, D. T. *The bone marrow niche for haematopoietic stem cells* 2014. doi:10.1038/nature12984.

38. Chi, H., Pepper, M. & Thomas, P. G. *Principles and therapeutic applications of adaptive immunity* 2024. doi:10.1016/j.cell.2024.03.037.

39. Alt, F. W. & Baltimore, D. Joining of immunoglobulin heavy chain gene segments: Implications from a chromosome with evidence of three D-J(H) fusions. *Proceedings of the National Academy of Sciences of the United States of America* **79,** 4118–4122. doi:10.1073/pnas.79.13.4118 (1982).

40. Tonegawa, S. Somatic generation of antibody diversity. *Nature* **302,** 575–581. doi:10.1038/302575a0 (1983).

41. Hesslein, D. G. & Schatz, D. G. in *Advances in Immunology* 169–232 (Academic Press, 2001). doi:10.1016/S0065-2776(01)78004-2.

42. Schatz, D. G., Oettinger, M. A. & Baltimore, D. The V(D)J recombination activating gene, RAG-1. *Cell* **59,** 1035–1048. doi:10.1016/0092-8674(89)90760-5 (1989).

43. Oettinger, M. A., Schatz, D. G., Gorka, C. & Baltimore, D. RAG-1 and RAG-2, adjacent genes that synergistically activate V(D)J recombination. *Science* **248,** 1517–1523. doi:10.1126/science.2360047 (1990).

44. Glick, B., Chang, T. S. & Jaap, R. G. The Bursa of Fabricius and Antibody Production. *Poultry Science* **35,** 224–225. doi:10.3382/ps.0350224 (1956).

45. Nossal, G. J. & Lederberg, J. *Antibody production by single cells [35]* 1958. doi:10.1038/1811419a0.

46. Miller, J. F. Immunological function of the thymus. *The Lancet* **278,** 748–749. doi:10.1016/S0140-6736(61)90693-6 (1961).

47. Lam, N., Lee, Y. S. & Farber, D. L. *A guide to adaptive immune memory* 2024. doi:10.1038/s41577-024-01040-6.

48. Kiessling, R., Klein, E. & Wigzell, H. "Natural" killer cells in the mouse. I. Cytotoxic cells with specificity for mouse Moloney leukemia cells. Specificity and distribution according to genotype. *European Journal of Immunology* **5,** 112–117. doi:10.1002/EJI.1830050208 (1975).

49. Naik, S. H. *et al.* Diverse and heritable lineage imprinting of early haematopoietic progenitors. *Nature* **496,** 229–232. doi:10.1038/nature12013 (2013).

50. Rumfelt, L. L., Zhou, Y., Rowley, B. M., Shinton, S. A. & Hardy, R. R. Lineage specification and plasticity in CD19- early B cell precursors. *Journal of Experimental Medicine* **203,** 675–687. doi:10.1084/jem.20052444 (2006).

51. Mårtensson, I. L., Melchers, F. & Winkler, T. H. A transgenic marker for mouse B lymphoid precursors. *Journal of Experimental Medicine* **185,** 653–661. doi:10.1084/jem.185.4.653 (1997).

52. Mansson, R. *et al.* B-lineage commitment prior to surface expression of B220 and CD19 on hematopoietic progenitor cells. *Blood* **112,** 1048–1055. doi:10.1182/blood-2007-11-125385 (2008).

53. Possot, C. *et al.* Notch signaling is necessary for adult, but not fetal, development of ROR$\gamma$+ innate lymphoid cells. *Nature Immunology* **12,** 949–958. doi:10.1038/ni.2105 (2011).

54. Inlay, M. A. *et al.* Ly6d marks the earliest stage of B-cell specification and identifies the branchpoint between B-cell and T-cell development. *Genes and Development* **23,** 2376–2381. doi:10.1101/gad.1836009 (2009).

55. Medina, K. L. *et al.* Separation of plasmacytoid dendritic cells from B-cell-biased lymphoid progenitor (BLP) and pre-pro B cells using PDCA-1. *PLoS ONE* **8.** doi:10.1371/journal.pone.0078408 (2013).

56. Mansson, R. *et al.* Single-cell analysis of the common lymphoid progenitor compartment reveals functional and molecular heterogeneity. *Blood* **115,** 2601–2609. doi:10.1182/BLOOD-2009-08-236398 (2010).

57. Jensen, C. T. *et al.* Dissection of progenitor compartments resolves developmental trajectories in B-lymphopoiesis. *Journal of Experimental Medicine* **215,** 1947–1963. doi:10.1084/JEM.20171384 (2018).

58. Early, P., Huang, H., Davis, M., Calame, K. & Hood, L. An immunoglobulin heavy chain variable region gene is generated from three segments of DNA: VH, D and JH. *Cell* **19,** 981–992. doi:10.1016/0092-8674(80)90089-6 (1980).

59. Sakano, H., Maki, R., Kurosawa, Y., Roeder, W. & Tonegawa, S. Two types of somatic recombination are necessary for the generation of complete immunoglobulin heavy-chain genes. *Nature* **286,** 676–683. doi:10.1038/286676a0 (1980).

60. Sakano, H., Hüppi, K., Heinrich, G. & Tonegawa, S. Sequences at the somatic recombination sites of immunoglobulin light-chain genes. *Nature* **280,** 288–294. doi:10.1038/280288a0 (1979).

61. Seidman, J. G., Max, E. E. & Leder, P. A $\kappa$-immunoglobulin gene is formed by site-specific recombination without further somatic mutation. *Nature* **280,** 370–375. doi:10.1038/280370a0 (1979).

62. Hardy, R. R., Carmack, C. E., Shinton, S. A., Kemp, J. D. & Hayakawa, K. Resolution and characterization of pro-B and pre-pro-B cell stages in normal mouse bone marrow. *Journal of Experimental Medicine* **173,** 1213–1225. doi:10.1084/jem.173.5.1213 (1991).

63. Zandi, S. *et al.* EBF1 Is Essential for B-Lineage Priming and Establishment of a Transcription Factor Network in Common Lymphoid Progenitors. *The Journal of Immunology* **181,** 3364–3372. doi:10.4049/JIMMUNOL.181.5.3364 (2008).

64. Erlandsson, L. *et al.* Both the pre-BCR and the IL-7R$\alpha$ are essential for expansion at the pre-BII cell stage in vivo. *European Journal of Immunology* **35,** 1969–1976. doi:10.1002/eji.200425821 (2005).

65. Grawunder, U. *et al.* Down-regulation of RAG1 and RAG2 gene expression in PreB cells after functional immunoglobulin heavy chain rearrangement. *Immunity* **3,** 601–608. doi:10.1016/1074-7613(95)90131-0 (1995).

66. Nemazee, D. *Mechanisms of central tolerance for B cells* 2017. doi:10.1038/nri.2017.19.

67. Cerutti, A., Cols, M. & Puga, I. *Marginal zone B cells: Virtues of innate-like antibody-producing lymphocytes* 2013. doi:10.1038/nri3383.

68. Nutt, S. L., Hodgkin, P. D., Tarlinton, D. M. & Corcoran, L. M. The generation of antibody-secreting plasma cells. *Nature Reviews Immunology* **15,** 160–171. doi:10.1038/nri3795 (2015).

69. Foss, D. L., Donskoy, E. & Goldschneider, I. The importation of hematogenous precursors by the thymus is a gated phenomenon in normal adult mice. *Journal of Experimental Medicine* **193,** 365–374. doi:10.1084/jem.193.3.365 (2001).

70. 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. *Journal of Experimental Medicine* **205,** 1187–1199. doi:10.1084/jem.20072168 (2008).

71. Saran, N. *et al.* Multiple extrathymic precursors contribute to T-cell development with different kinetics. *Blood* **115,** 1137–1144. doi:10.1182/blood-2009-07-230821 (2010).

72. Zietara, N. *et al.* Multicongenic fate mapping quantification of dynamics of thymus colonization. *Journal of Experimental Medicine* **212,** 1589–1601. doi:10.1084/jem.20142143 (2015).

73. Kyewski, B. A. Seeding of thymic microenvironments defined by distinct thymocyte-stromal cell interactions is developmentally controlled. *Journal of Experimental Medicine* **166,** 520–538. doi:10.1084/jem.166.2.520 (1987).

74. Sambandam, A. *et al.* Notch signaling controls the generation and differentiation of early T lineage progenitors. *Nature Immunology* **6,** 663–670. doi:10.1038/ni1216 (2005).

75. 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. *The Journal of Immunology* **150,** 4244–4252. doi:10.4049/jimmunol.150.10.4244 (1993).

76. Bhandoola, A., von Boehmer, H., 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. doi:10.1016/j.immuni.2007.05.009 (2007).

77. Bell, J. J. & Bhandoola, A. The earliest thymic progenitors for T cells possess myeloid lineage potential. *Nature* **452,** 764–767. doi:10.1038/nature06840 (2008).

78. Yui, M. A., Feng, N. & Rothenberg, E. V. Fine-Scale Staging of T Cell Lineage Commitment in Adult Mouse Thymus. *The Journal of Immunology* **185,** 284–293. doi:10.4049/jimmunol.1000679 (2010).

79. Kueh, H. Y. *et al.* Asynchronous combinatorial action of four regulatory factors activates Bcl11b for T cell commitment. *Nature Immunology* **17,** 956–965. doi:10.1038/ni.3514 (2016).

80. Zhang, J. A., Mortazavi, A., Williams, B. A., Wold, B. J. & Rothenberg, E. V. Dynamic transformations of genome-wide epigenetic marking and transcriptional control establish T cell identity. *Cell* **149,** 467–482. doi:10.1016/j.cell.2012.01.056 (2012).

81. Mingueneau, M. *et al.* The transcriptional landscape of $\alpha\beta$ T cell differentiation. *Nature Immunology* **14,** 619–632. doi:10.1038/ni.2590 (2013).

82. Taghon, T. & Rothenberg, E. V. *Molecular mechanisms that control mouse and human TCR-$\alpha\beta$ and TCR-$\gamma\delta$ T cell development* 2008. doi:10.1007/s00281-008-0134-3.

83. Ashby, K. M. & Hogquist, K. A. A guide to thymic selection of T cells. *Nature Reviews Immunology* **24,** 103–117. doi:10.1038/s41577-023-00911-8 (2024).

84. Singer, A., Adoro, S. & Park, J. H. Lineage fate and intense debate: myths, models and mechanisms of CD4- versus CD8-lineage choice. *Nature reviews. Immunology* **8,** 788–801. doi:10.1038/NRI2416 (2008).

85. Capone, M., Hockett, R. D. & Zlotnik, A. Kinetics of T cell receptor $\beta$, $\gamma$, and $\delta$ rearrangements during adult thymic development: T cell receptor rearrangements are present in CD44+CD25+ Pro-T thymocytes. *Proceedings of the National Academy of Sciences of the United States of America* **95,** 12522–12527. doi:10.1073/pnas.95.21.12522 (1998).

86. Cieslak, S. G. & Shahbazi, R. Gamma delta T cells and their immunotherapeutic potential in cancer. *Biomarker Research* **13,** 1–23. doi:10.1186/s40364-025-00762-6 (2025).

87. Pellin, D. *et al.* A comprehensive single cell transcriptional landscape of human hematopoietic progenitors. *Nature Communications* **10,** 1–15. doi:10.1038/s41467-019-10291-0 (2019).

88. Rothenberg, E. V. Single-cell insights into the hematopoietic generation of T-lymphocyte precursors in mouse and human. *Experimental Hematology* **95,** 1–12. doi:10.1016/j.exphem.2020.12.005 (2021).

89. Cramer, P. Organization and regulation of gene transcription. *Nature* **573,** 45–54. doi:10.1038/s41586-019-1517-4 (2019).

90. Dynan, W. S. & Tjian, R. The promoter-specific transcription factor Sp1 binds to upstream sequences in the SV40 early promoter. *Cell* **35,** 79–87. doi:10.1016/0092-8674(83)90210-6 (1983).

91. Latchman, D. S. Transcription factors: An overview. *International Journal of Biochemistry and Cell Biology* **29,** 1305–1312. doi:10.1016/S1357-2725(97)00085-X (1997).

92. Isbel, L., Grand, R. S. & Schübeler, D. Generating specificity in genome regulation through transcription factor sensitivity to chromatin. *Nature Reviews Genetics* **23,** 728–740. doi:10.1038/s41576-022-00512-6 (2022).

93. Kim, S. & Shendure, J. *Mechanisms of Interplay between Transcription Factors and the 3D Genome* 2019. doi:10.1016/j.molcel.2019.08.010.

94. Gross, D. S., Chowdhary, S., Anandhakumar, J. & Kainth, A. S. *Chromatin* 2015. doi:10.1016/j.cub.2015.10.059.

95. Hafner, A. & Boettiger, A. The spatial organization of transcriptional control. *Nature Reviews Genetics* **24,** 53–68. doi:10.1038/s41576-022-00526-0 (2023).

96. Flemming, W. *Zellsubstanz, Kern und Zelltheilung* doi:10.5962/bhl.title.168645 (von F. C. W. Vogel, Leipzig, 1882).

97. Meisher, F. in (1871).

98. Altmann, R. *Die Elementarorganismen und ihre Beziehungen zu den Zellen* (Leipzig, 1890).

99. Kossel, A. Ueber einen peptonartigen Bestandtheil des Zellkerns. *bchm* **8,** 511–515. doi:10.1515/bchm1.1884.8.6.511 (1884).

100. Wilson, E. B. *The cell in development and inheritance* (The Macmillan company, New York).

101. Sutton, W. S. THE CHROMOSOMES IN HEREDITY. *Biological Bulletin* **4,** 231–251 (1903).

102. Boveri, T. *Über mehrpolige Mitosen als Mittel zur Analyse des Zellkerns - Deutsche Digitale Bibliothek* (1902).

103. Driesch, H. *Analytische theorie der organischen entwicklung* (W. Engelmann, 1894).

104. Morgan, T. H. *Experimental Studies on Echinoderm Eggs* (Harvard University, 1984).

105. Rabl, C. Uber zellthilung. *Morphol Jahrb* **10,** 214–330 (1885).

106. Boveri, T. Die Blastomerenkerne von Ascaris megalocephala und die Theorie der Chromosomenindividualität. *Archiv für Zellforschung* **3,** 181–268 (1909).

107. Nägeli, K. W. *Theorie der Abstammungslehre* (R. Oldenbourg, München/Leipzig, 1884).

108. Sachs, J. *Vorlesungen über Pflanzenphysiologie* (Jena, G. Fisher, 1908).

109. Loeb, J. & Fleming, D. *The mechanistic conception of life: On some facts and principles of physiological morphology* doi:10.4159/HARVARD.9780674864313.C5 (Harvard University Press, 1893).

110. Cremer, T. & Cremer, M. Chromosome territories. *Cold Spring Harbor perspectives in biology* **2,** a003889. doi:10.1101/cshperspect.a003889 (2010).

111. Croft, J. A. *et al.* Differences in the localization and morphology of chromosomes in the human nucleus. *Journal of Cell Biology* **145,** 1119–1131. doi:10.1083/jcb.145.6.1119 (1999).

112. Lieberman-Aiden, E. *et al.* Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326,** 289–293. doi:10.1126/science.1181369 (2009).

113. Peric-Hupkes, D. *et al.* Molecular Maps of the Reorganization of Genome-Nuclear Lamina Interactions during Differentiation. *Molecular Cell* **38,** 603–613. doi:10.1016/j.molcel.2010.03.016 (2010).

114. Zheng, X. *et al.* Lamins Organize the Global Three-Dimensional Genome from the Nuclear Periphery. *Molecular Cell* **71,** 802–815. doi:10.1016/j.molcel.2018.05.017 (2018).

115. Lin, Y. C. *et al.* Global changes in the nuclear positioning of genes and intra-and interdomain genomic interactions that orchestrate B cell fate. *Nature Immunology* **13,** 1196–1204. doi:10.1038/ni.2432 (2012).

116. Boya, R. *et al.* Developmentally regulated higher-order chromatin interactions orchestrate B cell fate commitment. *Nucleic Acids Research* **45,** 11070–11087. doi:10.1093/nar/gkx722 (2017).

117. Dixon, J. R. *et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485,** 376–380. doi:10.1038/nature11082 (2012).

118. Hanssen, L. L. *et al.* Tissue-specific CTCF-cohesin-mediated chromatin architecture delimits enhancer interactions and function in vivo. *Nature Cell Biology* **19,** 952–961. doi:10.1038/ncb3573 (2017).

119. Sexton, T. & Cavalli, G. The role of chromosome domains in shaping the functional genome. *Cell* **160,** 1049–1059. doi:10.1016/J.CELL.2015.02.040/ASSET/9446B3D4-FD1C-49E1-A0FF-C4377536B381/MAIN.ASSETS/GR4.JPG (2015).

120. Dunham, I. *et al.* An Integrated Encyclopedia of DNA Elements in the Human Genome. *Nature* **489,** 57. doi:10.1038/NATURE11247 (2012).

121. Rao, S. S. *et al.* A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping. *Cell* **159,** 1665–1680. doi:10.1016/J.CELL.2014.11.021 (2014).

122. Huang, H. *et al.* CTCF mediates dosage- and sequence-context-dependent transcriptional insulation by forming local chromatin domains. *Nature Genetics* **53,** 1064–1074. doi:10.1038/s41588-021-00863-6 (2021).

123. Laverré, A., Tannier, E. & Necsulea, A. Long-range promoter-enhancer contacts are conserved during evolution and contribute to gene expression robustness. *Genome Research* **32,** gr.275901. bibrangedash 121. doi:10.1101/GR.275901.121 (2021).

124. Lidschreiber, K. *et al.* Transcriptionally active enhancers in human cancer cells. *Molecular Systems Biology* **17,** e9873. doi:10.15252/MSB.20209873 (2021).

125. Clément, Y., Torbey, P., Gilardi-Hebenstreit, P. & Crollius, H. R. Enhancer–gene maps in the human and zebrafish genomes using evolutionary linkage conservation. *Nucleic Acids Research* **48,** 2357–2371. doi:10.1093/NAR/GKZ1199 (2020).

126. Ren, X. *et al.* Parallel characterization of cis-regulatory elements for multiple genes using CRISPRpath. *Science Advances* **7,** 4360–4375. doi:10.1126/SCIADV.ABI4360/SUPPL{_}FILE/SCIADV.ABI4360{_}TABLES{_}S1{_}TO{_}S8.ZIP (2021).

127. Moreau, P. *et al.* The SV40 72 base repair repeat has a striking effect on gene expression both in SV40 and other chimeric recombinants. *Nucleic Acids Research* **9,** 6047–6068. doi:10.1093/nar/9.22.6047 (1981).

128. Travers, A. Chromatin modification by DNA tracking. *Proceedings of the National Academy of Sciences of the United States of America* **96,** 13634–13637. doi:10.1073/pnas.96.24.13634 (1999).

129. Li, G. *et al.* Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation. *Cell* **148,** 84–98. doi:10.1016/j.cell.2011.12.014 (2012).

130. Sanyal, A., Lajoie, B. R., Jain, G. & Dekker, J. The long-range interaction landscape of gene promoters. *Nature* **489,** 109–113. doi:10.1038/nature11279 (2012).

131. Gasperini, M. *et al.* A Genome-wide Framework for Mapping Gene Regulation via Cellular Genetic Screens. *Cell* **176,** 377–390. doi:10.1016/j.cell.2018.11.029 (2019).

132. Yang, J. H. & Hansen, A. S. Enhancer selectivity in space and time: from enhancer–promoter interactions to promoter activation. *Nature Reviews Molecular Cell Biology* **25,** 574–591. doi:10.1038/s41580-024-00710-6 (2024).

133. Ramasamy, S. *et al.* The Mediator complex regulates enhancer-promoter interactions. *Nature Structural and Molecular Biology* **30,** 991–1000. doi:10.1038/s41594-023-01027-2 (2023).

134. Weintraub, A. S. *et al.* YY1 Is a Structural Regulator of Enhancer-Promoter Loops. *Cell* **171,** 1573–1588. doi:10.1016/j.cell.2017.11.008 (2017).

135. Medvedovic, J. *et al.* Flexible long-range loops in the VH gene region of the Igh locus facilitate the generation of a diverse antibody repertoire. *Immunity* **39,** 229–244. doi:10.1016/j.immuni.2013.08.011 (2013).

136. Thomas-Claudepierre, A. S. *et al.* Mediator facilitates transcriptional activation and dynamic long-range contacts at the IgH locus during class switch recombination. *Journal of Experimental Medicine* **213,** 303–312. doi:10.1084/jem.20141967 (2016).

137. Mittag, T. & Pappu, R. V. A conceptual framework for understanding phase separation and addressing open questions and challenges. *Molecular Cell* **82,** 2201–2214. doi:10.1016/j.molcel.2022.05.018 (2022).

138. Malik, S. & Roeder, R. G. The metazoan Mediator co-activator complex as an integrative hub for transcriptional regulation. *Nature Reviews Genetics* **11,** 761–772. doi:10.1038/nrg2901 (2010).

139. Chepelev, I., Wei, G., Wangsa, D., Tang, Q. & Zhao, K. Characterization of genome-wide enhancer-promoter interactions reveals co-expression of interacting genes and modes of higher order chromatin organization. *Cell Research* **22,** 490–503. doi:10.1038/cr.2012.15 (2012).

140. Bahr, C. *et al.* A Myc enhancer cluster regulates normal and leukaemic haematopoietic stem cell hierarchies. *Nature* **553,** 515–520. doi:10.1038/nature25193 (2018).

141. Mateo, L. J. *et al.* Visualizing DNA folding and RNA in embryos at single-cell resolution. *Nature* **568,** 49–54. doi:10.1038/s41586-019-1035-4 (2019).

142. Brückner, D. B., Chen, H., Barinov, L., Zoller, B. & Gregor, T. Stochastic motion and transcriptional dynamics of pairs of distal DNA loci on a compacted chromosome. *Science* **380,** 1357–1362. doi:10.1126/science.adf5568 (2023).

143. Karr, J. P., Ferrie, J. J., Tjian, R. & Darzacq, X. The transcription factor activity gradient (TAG) model: contemplating a contact-independent mechanism for enhancer–promoter communication. *Genes and Development* **36,** 7–16. doi:10.1101/GAD.349160.121 (2022).

144. Thomas, H. F. *et al.* Enhancer cooperativity can compensate for loss of activity over large genomic distances. *Molecular Cell* **85,** 362–375. doi:10.1016/j.molcel.2024.11.008 (2025).

145. Mifsud, B. *et al.* Mapping long-range promoter contacts in human cells with high-resolution capture Hi-C. *Nature Genetics* **47,** 598–606. doi:10.1038/ng.3286 (2015).

146. Schoenfelder, S. *et al.* The pluripotent regulatory circuitry connecting promoters to their long-range interacting elements. *Genome Research* **25,** 582–597. doi:10.1101/gr.185272.114 (2015).

147. Banerji, J., Rusconi, S. & Schaffner, W. Expression of a $\beta$-globin gene is enhanced by remote SV40 DNA sequences. *Cell* **27,** 299–308. doi:10.1016/0092-8674(81)90413-X (1981).

148. Ross, I. L., Browne, C. & Hume, D. A. Transcription of individual genes in eukaryotic cells occurs randomly and infrequently. *Immunology & Cell Biology* **72,** 177–185. doi:10.1038/icb.1994.26 (1994).

149. Raser, J. M. & O'Shea, E. K. Control of stochasticity in eukaryotic gene expression. *Science* **304,** 1811–1814. doi:10.1126/science.1098641 (2004).

150. Eldar, A. & Elowitz, M. B. Functional roles for noise in genetic circuits. *Nature* **467,** 167–173. doi:10.1038/nature09326 (2010).

151. Zhou, Y. & Murre, C. Bursty gene expression and mRNA decay pathways orchestrate B cell activation. *Science Advances* **7,** 819. doi:10.1126/sciadv.abm0819 (2021).

152. Doidy, J. *et al.* "Hit-and-Run" transcription: De novo transcription initiated by a transient bZIP1 "hit" persists after the "run". *BMC Genomics* **17,** 92. doi:10.1186/s12864-016-2410-2 (2016).

153. Park, M., Kang, K. W. & Kim, J. W. *The role and application of transcriptional repressors in cancer treatment* 2023. doi:10.1007/S12272-023-01427-4.

154. Segert, J. A., Gisselbrecht, S. S. & Bulyk, M. L. Transcriptional Silencers: Driving Gene Expression with the Brakes On. *Trends in Genetics* **37,** 514–527. doi:10.1016/j.tig.2021.02.002 (2021).

155. Klemm, S. L., Shipony, Z. & Greenleaf, W. J. Chromatin accessibility and the regulatory epigenome. *Nature Reviews Genetics* **20,** 207–220. doi:10.1038/s41576-018-0089-8 (2019).

156. Allis, C. D. & Jenuwein, T. The molecular hallmarks of epigenetic control. *Nature Reviews Genetics* **17,** 487–500. doi:10.1038/nrg.2016.59 (2016).

157. Thurman, R. E. *et al.* The accessible chromatin landscape of the human genome. *Nature* **489,** 75–82. doi:10.1038/nature11232 (2012).

158. Iwafuchi-Doi, M. & Zaret, K. S. Cell fate control by pioneer transcription factors. *Development (Cambridge)* **143,** 1833–1837. doi:10.1242/dev.133900 (2016).

159. King, H. W. & Klose, R. J. The pioneer factor OCT4 requires the chromatin remodeller BRG1 to support gene regulatory element function in mouse embryonic stem cells. *eLife* **6.** doi:10.7554/eLife.22631 (2017).

160. Stern, M., Jensen, R. & Herskowitz, I. Five SWI genes are required for expression of the HO gene in yeast. *Journal of Molecular Biology* **178,** 853–868. doi:10.1016/0022-2836(84)90315-2 (1984).

161. Carlson, M., Osmond, B. C. & Botstein, D. Mutants of yeast defective in sucrose utilization. *Genetics* **98,** 25–40. doi:10.1093/GENETICS/98.1.25 (1981).

162. Neigeborn, L. & Carlson, M. Genes affecting the regulation of SUC2 gene expression by glucose repression in Saccharomyces cerevisiae. *Genetics* **108,** 845–858. doi:10.1093/GENETICS/108.4.845 (1984).

163. Neigeborn, L. & Carlson, M. Mutations causing constitutive invertase synthesis in yeast: genetic interactions with snf mutations. *Genetics* **115,** 247–253. doi:10.1093/GENETICS/115.2.247 (1987).

164. Kwon, H., Imbalzano, A. N., Khavari, P. A., Kingston, R. E. & Green, M. R. Nucleosome disruption and enhancement of activator binding by a human SW1/SNF complex. *Nature* **370,** 477–481. doi:10.1038/370477a0 (1994).

165. Wang, W. *et al.* Diversity and specialization of mammalian SWI/SNF complexes. *Genes & Development* **10,** 2117–2130. doi:10.1101/GAD.10.17.2117 (1996).

166. Centore, R. C., Sandoval, G. J., Soares, L. M. M., Kadoch, C. & Chan, H. M. Mammalian SWI/SNF Chromatin Remodeling Complexes: Emerging Mechanisms and Therapeutic Strategies. *Trends in Genetics* **36,** 936–950. doi:10.1016/J.TIG.2020.07.011 (2020).

167. Turner, B. M. *Decoding the nucleosome* 1993. doi:10.1016/S0092-8674(05)80078-9.

168. Dai, Z., Ramesh, V. & Locasale, J. W. The evolving metabolic landscape of chromatin biology and epigenetics. *Nature Reviews Genetics* **21,** 737–753. doi:10.1038/s41576-020-0270-8 (2020).

169. Millán-Zambrano, G., Burton, A., Bannister, A. J. & Schneider, R. *Histone post-translational modifications — cause and consequence of genome function* 2022. doi:10.1038/s41576-022-00468-7.

170. Vermeulen, M. *et al.* Selective Anchoring of TFIID to Nucleosomes by Trimethylation of Histone H3 Lysine 4. *Cell* **131,** 58–69. doi:10.1016/j.cell.2007.08.016 (2007).

171. Cano-Rodriguez, D. *et al.* Writing of H3K4Me3 overcomes epigenetic silencing in a sustained but context-dependent manner. *Nature Communications* **7,** 1–11. doi:10.1038/ncomms12284 (2016).

172. Newman, J. R. *et al.* Single-cell proteomic analysis of S. cerevisiae reveals the architecture of biological noise. *Nature* **441,** 840–846. doi:10.1038/nature04785 (2006).

173. Benayoun, B. A. *et al.* H3K4me3 breadth is linked to cell identity and transcriptional consistency. *Cell* **158,** 673–688. doi:10.1016/j.cell.2014.06.027 (2014).

174. Talbert, P. B., Meers, M. P. & Henikoff, S. Old cogs, new tricks: the evolution of gene expression in a chromatin context. *Nature Reviews Genetics* **20,** 283–297. doi:10.1038/s41576-019-0105-7 (2019).

175. Creyghton, M. P. *et al.* Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proceedings of the National Academy of Sciences of the United States of America* **107,** 21931–21936. doi:10.1073/pnas.1016071107 (2010).

176. Van Mierlo, G., Veenstra, G. J. C., Vermeulen, M. & Marks, H. The Complexity of PRC2 Subcomplexes. *Trends in Cell Biology* **29,** 660–671. doi:10.1016/J.TCB.2019.05.004 (2019).

177. Lee, C. H. *et al.* Distinct Stimulatory Mechanisms Regulate the Catalytic Activity of Polycomb Repressive Complex 2. *Molecular cell* **70,** 435–448. doi:10.1016/J.MOLCEL.2018.03.019 (2018).

178. Margueron, R. *et al.* Role of the polycomb protein EED in the propagation of repressive histone marks. *Nature* **461,** 762–767. doi:10.1038/NATURE08398 (2009).

179. Macrae, T. A., Fothergill-Robinson, J. & Ramalho-Santos, M. Regulation, functions and transmission of bivalent chromatin during mammalian development. *Nature Reviews Molecular Cell Biology 2022 24:1* **24,** 6–26. doi:10.1038/s41580-022-00518-2 (2022).

180. Allshire, R. C. & Madhani, H. D. Ten principles of heterochromatin formation and function. *Nature Reviews Molecular Cell Biology 2017 19:4* **19,** 229–244. doi:10.1038/nrm.2017.119 (2017).

181. Bannister, A. J. *et al.* Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. *Nature* **410,** 120–124. doi:10.1038/35065138 (2001).

182. Lachner, M., O'Carroll, D., Rea, S., Mechtler, K. & Jenuwein, T. Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. *Nature* **410,** 116–120. doi:10.1038/35065132 (2001).

183. Canzio, D. *et al.* Chromodomain-mediated oligomerization of HP1 suggests a nucleosome-bridging mechanism for heterochromatin assembly. *Molecular Cell* **41,** 67–81. doi:10.1016/j.molcel.2010.12.016 (2011).

184. Strom, A. R. *et al.* Phase separation drives heterochromatin domain formation. *Nature* **547,** 241–245. doi:10.1038/nature22989 (2017).

185. Larson, A. G. *et al.* Liquid droplet formation by HP1$\alpha$ suggests a role for phase separation in heterochromatin. *Nature* **547,** 236–240. doi:10.1038/nature22822 (2017).

186. Pelham, H. R. A regulatory upstream promoter element in the Drosophila Hsp 70 heat-shock gene. *Cell* **30,** 517–528. doi:10.1016/0092-8674(82)90249-5 (1982).

187. Kornberg, T. B. Understanding the homeodomain. *Journal of Biological Chemistry* **268,** 26813–26816. doi:10.1016/s0021-9258(19)74183-0 (1993).

188. Rhodes, D. & Klug, A. Zinc fingers. *Scientific American* **268.** doi:10.1038/scientificamerican\-0293-56 (1993).

189. Parker, M. G. Steroid and related receptors. *Current Opinion in Cell Biology* **5,** 499–504. doi:10.1016/0955-0674(93)90016-J (1993).

190. Sharrocks, A. D., Brown, A. L., Ling, Y. & Yates, P. R. *The ETS-domain transcription factor family* 1997. doi:10.1016/S1357-2725(97)00086-1.

191. O'Dwyer, M. R. *et al.* Nucleosome fibre topology guides transcription factor binding to enhancers. *Nature* **638,** 251–260. doi:10.1038/s41586-024-08333-9 (2025).

192. Wang, X., Lin, P. & Ho, J. W. Discovery of cell-type specific DNA motif grammar in cis-regulatory elements using random Forest. *BMC Genomics* **19,** 929. doi:10.1186/s12864-017-4340-z (2018).

193. Nagy, G. & Nagy, L. Motif grammar: The basis of the language of gene expression. *Computational and Structural Biotechnology Journal* **18,** 2026–2032. doi:10.1016/j.csbj.2020.07.007 (2020).

194. Takaku, M. *et al.* GATA3-dependent cellular reprogramming requires activation-domain dependent recruitment of a chromatin remodeler. *Genome Biology* **17.** doi:10.1186/s13059-016-0897-0 (2016).

195. Balsalobre, A. & Drouin, J. Pioneer factors as master regulators of the epigenome and cell fate. *Nature Reviews Molecular Cell Biology* **23,** 449–464. doi:10.1038/s41580-022-00464-z (2022).

196. Cirillo, L. A. *et al.* Opening of compacted chromatin by early developmental transcription factors HNF3 (FoxA) and GATA-4. *Molecular Cell* **9,** 279–289. doi:10.1016/S1097-2765(02)00459-8 (2002).

197. Boller, S. *et al.* Pioneering Activity of the C-Terminal Domain of EBF1 Shapes the Chromatin Landscape for B Cell Programming. *Immunity* **44,** 527–541. doi:10.1016/j.immuni.2016.02.021 (2016).

198. Larson, E. D., Marsh, A. J. & Harrison, M. M. Pioneering the developmental frontier. *Molecular Cell* **81,** 1640–1650. doi:10.1016/j.molcel.2021.02.020 (2021).

199. Ptashne, M. & Gann, A. Transcriptional activation by recruitment. *Nature* **386,** 569–577. doi:10.1038/386569a0 (1997).

200. Akerblom, I. E., Slater, E. P., Beato, M., Baxter, J. D. & Mellon, P. L. Negative regulation by glucocorticoids through interference with a cAMP responsive enhancer. *Science* **241,** 350–353. doi:10.1126/science.2838908 (1988).

201. Reiter, F., Wienerroither, S. & Stark, A. Combinatorial function of transcription factors and cofactors. *Current Opinion in Genetics & Development* **43,** 73–81. doi:10.1016/J.GDE.2016.12.007 (2017).

202. Furlong, E. E. & Levine, M. Developmental enhancers and chromosome topology. *Science* **361,** 1341–1345. doi:10.1126/science.aau0320 (2018).

203. Gan, P. *et al.* Coactivator condensation drives cardiovascular cell lineage specification. *Science Advances* **10.** doi:10.1126/sciadv.adk7160 (2024).

204. Meyer, K., Huang, B. & Weiner, O. D. Emerging roles of transcriptional condensates as temporal signal integrators. *Nature Reviews Genetics* **26,** 559–570. doi:10.1038/s41576-025-00837-y (2025).

205. DeKoter, R. P. & Singh, H. Regulation of B lymphocyte and macrophage development by graded expression of PU.1. *Science* **288,** 1439–1442. doi:10.1126/science.288.5470.1439 (2000).

206. Lichtinger, M. *et al.* RUNX1 reshapes the epigenetic landscape at the onset of haematopoiesis. *EMBO Journal* **31,** 4318–4333. doi:10.1038/emboj.2012.275 (2012).

207. Hosokawa, H. *et al.* Transcription Factor PU.1 Represses and Activates Gene Expression in Early T Cells by Redirecting Partner Transcription Factor Binding. *Immunity* **48,** 1119–1134. doi:10.1016/j.immuni.2018.04.024 (2018).

208. Yoshida, T., Yao-Ming Ng, S., Zuniga-Pflucker, J. C. & Georgopoulos, K. Early hematopoietic lineage restrictions directed by Ikaros. *Nature Immunology* **7,** 382–391. doi:10.1038/ni1314 (2006).

209. Dias, S., Månsson, R., Gurbuxani, S., Sigvardsson, M. & Kee, B. L. E2A Proteins Promote Development of Lymphoid-Primed Multipotent Progenitors. *Immunity* **29,** 217–227. doi:10.1016/j.immuni.2008.05.015 (2008).

210. Welinder, E. *et al.* The transcription factors E2A and HEB act in concert to induce the expression of FOXO1 in the common lymphoid progenitor. *Proceedings of the National Academy of Sciences of the United States of America* **108,** 17402–17407. doi:10.1073/pnas.1111766108 (2011).

211. Amin, R. H. & Schlissel, M. S. Foxo1 directly regulates the transcription of recombination-activating genes during B cell development. *Nature Immunology* **9,** 613–622. doi:10.1038/ni.1612 (2008).

212. Dengler, H. S. *et al.* Distinct functions for the transcription factor Foxo1 at various stages of B cell differentiation. *Nature Immunology* **9,** 1388–1398. doi:10.1038/ni.1667 (2008).

213. Mansson, R. *et al.* Positive intergenic feedback circuitry, involving EBF1 and FOXO1, orchestrates B-cell fate. *PNAS* **109,** 21028–21033. doi:10.1073/PNAS.1211427109 (2012).

214. Hagman, J., Belanger, C., Travis, A., Turck, C. W. & Grosschedl, R. Cloning and functional characterization of early B-cell factor, a regulator of lymphocyte-specific gene expression. *Genes and Development* **7,** 760–773. doi:10.1101/gad.7.5.760 (1993).

215. Wang, M. M. & Reed, R. R. Molecular cloning of the olfactory neuronal transcription factor Olf-1 by genetic selection in yeast. *Nature* **364,** 121–126. doi:10.1038/364121a0 (1993).

216. Åkerblad, P., Lind, U., Liberg, D., Bamberg, K. & Sigvardsson, M. Early B-Cell Factor (O/E-1) Is a Promoter of Adipogenesis and Involved in Control of Genes Important for Terminal Adipocyte Differentiation. *Molecular and Cellular Biology* **22,** 8015–8025. doi:10.1128/mcb.22.22.8015-8025.2002 (2002).

217. Qian, H. *et al.* Molecular Characterization of Prospectively Isolated Multipotent Mesenchymal Progenitors Provides New Insight into the Cellular Identity of Mesenchymal Stem Cells in Mouse Bone Marrow. *Molecular and Cellular Biology* **33,** 661–677. doi:10.1128/mcb.01287-12 (2013).

218. Lin, H. & Grosschedl, R. Failure of B-cell differentiation in mice lacking the transcription factor EBF. *Nature* **376,** 263–267. doi:10.1038/376263a0 (1995).

219. Gao, H. *et al.* Opposing effects of SWI/SNF and Mi-2/NuRD chromatin remodeling complexes on epigenetic reprogramming by EBF and Pax5. *Proceedings of the National Academy of Sciences of the United States of America* **106,** 11258–11263. doi:10.1073/pnas.0809485106 (2009).

220. Strid, T. *et al.* B Lymphocyte Specification Is Preceded by Extensive Epigenetic Priming in Multipotent Progenitors. *The Journal of Immunology* **206,** 2700–2713. doi:10.4049/jimmunol.2100048 (2021).

221. Wang, Y. *et al.* A Prion-like Domain in Transcription Factor EBF1 Promotes Phase Separation and Enables B Cell Programming of Progenitor Chromatin. *Immunity* **53,** 1151–1167. doi:10.1016/j.immuni.2020.10.009 (2020).

222. Nutt, S. L., Heavey, B., Rolink, A. G. & Busslinger, M. Commitment to the B-lymphoid lineage depends on the transcription factor Pax5. *Journal of Immunology* **195,** 766–772. doi:10.1038/44076 (2015).

223. Rolink, A. G., Nutt, S. L., Melchers, F. & Busslinger, M. Long-term in vivo reconstitution of T-cell development by Pax5-deficient B-cell progenitors. *Nature* **401,** 603–606. doi:10.1038/44164 (1999).

224. Mikkola, I., Heavey, B., Horcher, M. & Busslinger, M. Reversion of B cell commitment upon loss of Pax5 expression. *Science* **297,** 110–113. doi:10.1126/SCIENCE.1067518 (2002).

225. Schaniel, C., Bruno, L., Melchers, F. & Rolink, A. G. Multiple hematopoietic cell lineages develop in vivo from transplanted Pax5-deficient pre-B I-cell clones. *Blood* **99,** 472–478. doi:10.1182/blood.V99.2.472 (2002).

226. Höflinger, S. *et al.* Analysis of Notch1 Function by In Vitro T Cell Differentiation of Pax5 Mutant Lymphoid Progenitors. *The Journal of Immunology* **173,** 3935–3944. doi:10.4049/jimmunol.173.6.3935 (2004).

227. Zandi, S. *et al.* Single-cell analysis of early B-lymphocyte development suggests independent regulation of lineage specification and commitment in vivo. *PNAS* **109,** 15871–15876. doi:10.1073/PNAS.1210144109 (2012).

228. Tagoh, H. *et al.* The mechanism of repression of the myeloid-specific c-fms gene by Pax5 during B lineage restriction. *EMBO Journal* **25,** 1070–1080. doi:10.1038/sj.emboj.7600997 (2006).

229. Bayer, M. *et al.* Tnpo3 enables EBF1 function in conditions of antagonistic Notch signaling. *Genes and Development* **36,** 901–915. doi:10.1101/gad.349696.122 (2022).

230. Germar, K. *et al.* T-cell factor 1 is a gatekeeper for T-cell specification in response to Notch signaling. *Proceedings of the National Academy of Sciences of the United States of America* **108,** 20060–20065. doi:10.1073/pnas.1110230108 (2011).

231. Weber, B. N. *et al.* A critical role for TCF-1 in T-lineage specification and differentiation. *Nature* **476,** 63–68. doi:10.1038/nature10279 (2011).

232. Johnson, J. L. *et al.* Lineage-Determining Transcription Factor TCF-1 Initiates the Epigenetic Identity of T Cells. *Immunity* **48,** 243–257. doi:10.1016/j.immuni.2018.01.012 (2018).

233. Frelin, C. *et al.* GATA-3 regulates the self-renewal of long-term hematopoietic stem cells. *Nature Immunology* **14,** 1037–1044. doi:10.1038/ni.2692 (2013).

234. Tindemans, I., Serafini, N., DiSanto, J. P. & Hendriks, R. W. GATA-3 function in innate and adaptive immunity. *Immunity* **41,** 191–206. doi:10.1016/j.immuni.2014.06.006 (2014).

235. Lim, K. C. *et al.* Gata3 loss leads to embryonic lethality due to noradrenaline deficiency of the sympathetic nervous system. *Nature Genetics* **25,** 209–212. doi:10.1038/76080 (2000).

236. Kouros-Mehr, H., Slorach, E. M., Sternlicht, M. D. & Werb, Z. GATA-3 Maintains the Differentiation of the Luminal Cell Fate in the Mammary Gland. *Cell* **127,** 1041–1055. doi:10.1016/j.cell.2006.09.048 (2006).

237. Hasegawa, S. L. *et al.* Dosage-dependent rescue of definitive nephrogenesis by a distant Gata3 enhancer. *Developmental Biology* **301,** 568–577. doi:10.1016/j.ydbio.2006.09.030 (2007).

238. Wei, G. *et al.* Genome-wide Analyses of Transcription Factor GATA3-Mediated Gene Regulation in Distinct T Cell Types. *Immunity* **35,** 299–311. doi:10.1016/j.immuni.2011.08.007 (2011).

239. Hosoya, T., Maillard, I. & Engel, J. D. From the cradle to the grave: Activities of GATA-3 throughout T-cell development and differentiation. *Immunological Reviews* **238,** 110–125. doi:10.1111/j.1600-065X.2010.00954.x (2010).

240. Li, L., Leid, M. & Rothenberg, E. V. An early T cell lineage commitment checkpoint dependent on the transcription factor Bcl11b. *Science* **329,** 89–93. doi:10.1126/science.1188989 (2010).

241. Ikawa, T. *et al.* An essential developmental checkpoint for production of the T cell lineage. *Science* **329,** 93–96. doi:10.1126/SCIENCE.1188995 (2010).

242. Longabaugh, W. J. *et al.* Bcl11b and combinatorial resolution of cell fate in the T-cell gene regulatory network. *Proceedings of the National Academy of Sciences of the United States of America* **114,** 5800–5807. doi:10.1073/pnas.1610617114 (2017).

243. Hosokawa, H. *et al.* Bcl11b sets pro-T cell fate by site-specific cofactor recruitment and by repressing Id2 and Zbtb16. *Nature Immunology* **19,** 1427–1440. doi:10.1038/s41590-018-0238-4 (2018).

244. Cismasiu, V. B. *et al.* BCL11B functionally associates with the NuRD complex in T lymphocytes to repress targeted promoter. *Oncogene* **24,** 6753–6764. doi:10.1038/sj.onc.1208904 (2005).

245. Hosokawa, H. *et al.* Cell type-specific actions of Bcl11b in early T-lineage and group 2 innate lymphoid cells. *Journal of Experimental Medicine* **217.** doi:10.1084/jem.20190972 (2020).

246. Chowdhary, K. & Benoist, C. A variegated model of transcription factor function in the immune system. *Trends in Immunology* **44,** 530–541. doi:10.1016/J.IT.2023.05.001/ASSET/C443A5C6-6371-41EC-8CB6-3E6B9D663846/MAIN.ASSETS/GR3.SML (2023).

247. Jolma, A. *et al.* DNA-dependent formation of transcription factor pairs alters their binding specificity. *Nature* **527,** 384–388. doi:10.1038/nature15518 (2015).

248. Garvie, C. W., Hagman, J. & Wolberger, C. Structural studies of Ets-1/Pax5 complex formation on DNA. *Molecular Cell* **8,** 1267–1276. doi:10.1016/S1097-2765(01)00410-5 (2001).

249. Hollenhorst, P. C. *et al.* DNA Specificity Determinants Associate with Distinct Transcription Factor Functions. *PLOS Genetics* **5,** e1000778. doi:10.1371/JOURNAL.PGEN.1000778 (2009).

250. Hermsen, R., Tans, S. & Ten Wolde, P. R. Transcriptional regulation by competing transcription factor modules. *PLoS Computational Biology* **2,** 1552–1560. doi:10.1371/journal.pcbi.0020164 (2006).

251. Alon, U. Network motifs: Theory and experimental approaches. *Nature Reviews Genetics* **8,** 450–461. doi:10.1038/nrg2102 (2007).

252. Seet, C. S., Brumbaugh, R. L. & Kee, B. L. Early B cell factor promotes B lymphopoiesis with reduced interleukin 7 responsiveness in the absence of E2A. *Journal of Experimental Medicine* **199,** 1689–1700. doi:10.1084/jem.20032202 (2004).

253. Shin, B. & Rothenberg, E. V. *Multi-modular structure of the gene regulatory network for specification and commitment of murine T cells* 2023. doi:10.3389/fimmu.2023.1108368.

254. Nutt, S. L., Morrison, A. M., Dörfler, P., Rolink, A. & Busslinger, M. Identification of BSAP (Pax-5) target genes in early B-cell development by loss- and gain-of-function experiments. *EMBO Journal* **17,** 2319–2333. doi:10.1093/emboj/17.8.2319 (1998).

255. Schebesta, A. *et al.* Transcription Factor Pax5 Activates the Chromatin of Key Genes Involved in B Cell Signaling, Adhesion, Migration, and Immune Function. *Immunity* **27,** 49–63. doi:10.1016/j.immuni.2007.05.019 (2007).

256. Benezra, R., Davis, R. L., Lockshon, D., Turner, D. L. & Weintraub, H. The protein Id: A negative regulator of helix-loop-helix DNA binding proteins. *Cell* **61,** 49–59. doi:10.1016/0092-8674(90)90214-Y (1990).

257. Thompson, E. C. *et al.* Ikaros DNA-Binding Proteins as Integral Components of B Cell Developmental-Stage-Specific Regulatory Circuits. *Immunity* **26,** 335–344. doi:10.1016/j.immuni.2007.02.010 (2007).

258. Sigvardsson, M., O'Riordan, M. & Grosschedl, R. EBF and E47 collaborate to induce expression of the endogenous immunoglobulin surrogate light chain genes. *Immunity* **7,** 25–36. doi:10.1016/S1074-7613(00)80507-5 (1997).

259. Sigvardsson, M. Overlapping Expression of Early B-Cell Factor and Basic Helix-Loop-Helix Proteins as a Mechanism To Dictate B-Lineage-Specific Activity of the $\lambda$ 5 Promoter. *Molecular and Cellular Biology* **20,** 3640–3654. doi:10.1128/mcb.20.10.3640-3654.2000 (2000).

260. Decker, T. *et al.* Stepwise Activation of Enhancer and Promoter Regions of the B Cell Commitment Gene Pax5 in Early Lymphopoiesis. *Immunity* **30,** 508–520. doi:10.1016/J.IMMUNI.2009.01.012 (2009).

261. 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. doi:10.1016/S1074-7613(02)00474-0 (2002).

262. Roessler, S. *et al.* Distinct Promoters Mediate the Regulation of Ebf1 Gene Expression by Interleukin-7 and Pax5. *Molecular and Cellular Biology* **27,** 579–594. doi:10.1128/mcb.01192-06 (2007).

263. Neph, S. *et al.* Circuitry and dynamics of human transcription factor regulatory networks. *Cell* **150,** 1274–1286. doi:10.1016/J.CELL.2012.04.040/ATTACHMENT/12C0874B-E06C-462E-8BEB-D9BDC2D6CB2F/MMC2 (2012).

264. Alpdogan, Ö. & Van Den Brink, M. R. IL-7 and IL-15: Therapeutic cytokines for immunodeficiency. *Trends in Immunology* **26,** 56–64. doi:10.1016/j.it.2004.11.002 (2005).

265. Winer, H. *et al.* IL-7: Comprehensive review. *Cytokine* **160,** 156049. doi:10.1016/j.cyto.2022.156049 (2022).

266. Lodewijckx, I. & Cools, J. Deregulation of the interleukin-7 signaling pathway in lymphoid malignancies. *Pharmaceuticals* **14,** 443. doi:10.3390/ph14050443 (2021).

267. Peschon, J. J. *et al.* Early Lymphocyte Expansion Is Severely Impaired in Interleukin 7 Receptor-deficient Mice. *Journal of Experimental Medicine* **180,** 1955–1960. doi:10.1084/jem.180.5.1955 (1994).

268. Von Freeden-Jeffry, U. *et al.* Lymphopenia in interleukin (IL)-7 gene-deleted mice identifies IL-7 as a nonredundant cytokine. *Journal of Experimental Medicine* **181,** 1519–1526. doi:10.1084/jem.181.4.1519 (1995).

269. Puel, A., Ziegler, S. F., Buckley, R. H. & Leonard, W. J. Defective IL7R expression in T-B+NK+ severe combined immunodeficiency. *Nature Genetics* **20,** 394–397. doi:10.1038/3877 (1998).

270. Dias, S., Silva, H., Cumano, A. & Vieira, P. Interleukin-7 is necessary to maintain the B cell potential in common lymphoid progenitors. *Journal of Experimental Medicine* **201,** 971–979. doi:10.1084/jem.20042393 (2005).

271. Tsapogas, P. *et al.* IL-7 mediates Ebf-1-dependent lineage restriction in early lymphoid progenitors. *Blood* **118,** 1283–1290. doi:10.1182/blood-2011-01-332189 (2011).

272. Kikuchi, K., Kasai, H., Watanabe, A., Lai, A. Y. & Kondo, M. IL-7 Specifies B Cell Fate at the Common Lymphoid Progenitor to Pre-ProB Transition Stage by Maintaining Early B Cell Factor Expression. *The Journal of Immunology* **181,** 383–392. doi:10.4049/jimmunol.181.1.383 (2008).

273. Akashi, K., Kondo, M., Von Freeden-Jeffry, U., Murray, R. & Weissman, I. L. Bcl-2 rescues T lymphopoiesis in Interleukin-7 receptor-deficient mice. *Cell* **89,** 1033–1041. doi:10.1016/S0092-8674(00)80291-3 (1997).

274. Plumb, A. W. *et al.* Interleukin-7 in the transition of bone marrow progenitors to the thymus. *Immunology and Cell Biology* **95,** 916–924. doi:10.1038/icb.2017.68 (2017).

275. Artavanis-Tsakonas, S., Rand, M. D. & Lake, R. J. Notch Signaling: Cell Fate Control and Signal Integration in Development. *Science* **284,** 770–776. doi:10.1126/SCIENCE.284.5415.770 (1999).

276. Yui, M. A. & Rothenberg, E. V. Developmental gene networks: a triathlon on the course to T cell identity. *Nature Reviews Immunology* **14,** 529–545. doi:10.1038/nri3702 (2014).

277. Taniguchi, Y. *et al.* Notch receptor cleavage depends on but is not directly executed by presenilins. *Proceedings of the National Academy of Sciences of the United States of America* **99,** 4014–4019. doi:10.1073/pnas.052017699 (2002).

278. Radtke, F. *et al.* Deficient T cell fate specification in mice with an induced inactivation of Notch1. *Immunity* **10,** 547–558. doi:10.1016/S1074-7613(00)80054-0 (1999).

279. Wilson, A., MacDonald, H. R. & Radtke, F. Notch 1-deficient common lymphoid precursors adopt a B cell fate in the thymus. *Journal of Experimental Medicine* **194,** 1003–1012. doi:10.1084/jem.194.7.1003 (2001).

280. Waddington, C. H. *An introduction to modern genetics* doi:10.5962/bhl.title.6461 (The Macmillan company, New York, 1939).

281. Waddington, C. H. *The strategy of the genes: A discussion of some aspects of theoretical biology* 1–262. doi:10.4324/9781315765471 (Taylor and Francis, 1957).

282. Takahashi, K. & Yamanaka, S. Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors. *Cell* **126,** 663–676. doi:10.1016/j.cell.2006.07.024 (2006).

283. Davis, R. L., Weintraub, H. & Lassar, A. B. Expression of a single transfected cDNA converts fibroblasts to myoblasts. *Cell* **51,** 987–1000. doi:10.1016/0092-8674(87)90585-X (1987).

284. Arinobu, Y. *et al.* Reciprocal Activation of GATA-1 and PU.1 Marks Initial Specification of Hematopoietic Stem Cells into Myeloerythroid and Myelolymphoid Lineages. *Cell Stem Cell* **1,** 416–427. doi:10.1016/j.stem.2007.07.004 (2007).

285. Kueh, H. Y., Champhekhar, A., Nutt, S. L., Elowitz, M. B. & Rothenberg, E. V. Positive feedback between PU.1 and the cell cycle controls myeloid differentiation. *Science* **341,** 670–673. doi:10.1126/science.1240831 (2013).

286. Schlenner, S. M. *et al.* Fate Mapping Reveals Separate Origins of T Cells and Myeloid Lineages in the Thymus. *Immunity* **32,** 426–436. doi:10.1016/j.immuni.2010.03.005 (2010).

287. Spooner, C. J., Cheng, J. X., Pujadas, E., Laslo, P. & Singh, H. A Recurrent Network Involving the Transcription Factors PU.1 and Gfi1 Orchestrates Innate and Adaptive Immune Cell Fates. *Immunity* **31,** 576–586. doi:10.1016/j.immuni.2009.07.011 (2009).

288. Zhang, J. *et al.* Harnessing of the nucleosome-remodeling-deacetylase complex controls lymphocyte development and prevents leukemogenesis. *Nature Immunology* **13,** 86–94. doi:10.1038/ni.2150 (2012).

289. Hu, Y. *et al.* Lineage-specific 3D genome organization is assembled at multiple scales by IKAROS. *Cell* **186,** 5269–5289. doi:10.1016/j.cell.2023.10.023 (2023).

290. Yang, Q. *et al.* E47 Controls the Developmental Integrity and Cell Cycle Quiescence of Multipotential Hematopoietic Progenitors. *The Journal of Immunology* **181,** 5885–5894. doi:10.4049/jimmunol.181.9.5885 (2008).

291. Ng, S. Y. M., Yoshida, T., Zhang, J. & Georgopoulos, K. Genome-wide Lineage-Specific Transcriptional Networks Underscore Ikaros-Dependent Lymphoid Priming in Hematopoietic Stem Cells. *Immunity* **30,** 493–507. doi:10.1016/j.immuni.2009.01.014 (2009).

292. Semerad, C. L., Mercer, E. M., Inlay, M. A., Weissman, I. L. & Murre, C. E2A proteins maintain the hematopoietic stem cell pool and promote the maturation of myelolymphoid and myeloerythroid progenitors. *Proceedings of the National Academy of Sciences of the United States of America* **106,** 1930–1935. doi:10.1073/pnas.0808866106 (2009).

293. Nutt, S. L., Thévenin, C. & Busslinger, M. Essential functions of Pax-5 (BSAP) in pro-B cell development. *Immunobiology* **198,** 227–235. doi:10.1016/S0171-2985(97)80043-5 (1997).

294. Åkerblad, P., Rosberg, M., Leanderson, T. & Sigvardsson, M. The B29 (Immunoglobulin $\beta$-Chain) Gene Is a Genetic Target for Early B-Cell Factor. *Molecular and Cellular Biology* **19,** 392–401. doi:10.1128/mcb.19.1.392 (1999).

295. Gisler, R., Jacobsen, S. E. & Sigvardsson, M. Cloning of human early B-cell factor and identification of target genes suggest a conserved role in B-cell development in man and mouse. *Blood* **96,** 1457–1464. doi:10.1182/blood.v96.4.1457 (2000).

296. Lin, Y. C. *et al.* A global network of transcription factors, involving E2A, EBF1 and Foxo1, that orchestrates B cell fate. *Nature Immunology* **11,** 635–643. doi:10.1038/ni.1891 (2010).

297. Treiber, T. *et al.* Early B cell factor 1 regulates B cell gene networks by activation, repression, and transcription- independent poising of chromatin. *Immunity* **32,** 714–725. doi:10.1016/j.immuni.2010.04.013 (2010).

298. Revilla-I-Domingo, R. *et al.* The B-cell identity factor Pax5 regulates distinct transcriptional programmes in early and late B lymphopoiesis. *EMBO Journal* **31,** 3130–3146. doi:10.1038/emboj.2012.155 (2012).

299. Nutt, S. L., Urbánek, P., Relink, A. & Busslinger, M. Essential functions of Pax5 (BSAP) in pro-B cell development: Difference between fetal and adult B lymphopoiesis and reduced V-to-DJ recombination at the IgH locus. *Genes and Development* **11,** 476–491. doi:10.1101/gad.11.4.476 (1997).

300. Nechanitzky, R. *et al.* Transcription factor EBF1 is essential for the maintenance of B cell identity and prevention of alternative fates in committed cells. *Nature Immunology* **14,** 867–875. doi:10.1038/ni.2641 (2013).

301. Ezine, S., Weissman, I. L. & Rouse, R. V. Bone marrow cells give rise to distinct cell clones within the thymus. *Nature* **309,** 629–631. doi:10.1038/309629a0 (1984).

302. 20, B., Chang, S. J., Macnabb, B. W. & Rothenberg, E. V. Thymus and T Cell Development Focus Transcriptional network dynamics in early T cell development. *Journal of Experimental Medicine* **221.** doi:10.1084/jem.20230893 (2024).

303. Gamble, N. *et al.* PU.1 and BCL11B sequentially cooperate with RUNX1 to anchor mSWI/SNF to poise the T cell effector landscape. *Nature Immunology* **25,** 860–872. doi:10.1038/s41590-024-01807-y (2024).

304. De Obaldia, M. E. *et al.* T cell development requires constraint of the myeloid regulator C/EBP-$\alpha$ by the Notch target and transcriptional repressor Hes1. *Nature Immunology* **14,** 1277–1284. doi:10.1038/ni.2760 (2013).

305. Rothenberg, E. V. Logic and lineage impacts on functional transcription factor deployment for T-cell fate commitment. *Biophysical Journal* **120,** 4162–4181. doi:10.1016/J.BPJ.2021.04.002 (2021).

306. Romero-Wolf, M. *et al.* Notch2 complements Notch1 to mediate inductive signaling that initiates early T cell development. *Journal of Cell Biology* **219.** doi:10.1083/JCB.202005093 (2020).

307. Zhou, W. *et al.* Single-Cell Analysis Reveals Regulatory Gene Expression Dynamics Leading to Lineage Commitment in Early T Cell Development. *Cell Systems* **9,** 321–337. doi:10.1016/j.cels.2019.09.008 (2019).

308. Hattori, N., Kawamoto, H., Fujimoto, S., Kuno, K. & YoshiniotoKatsura. Involvement of transcription factors TCF-1 and GATA-3 in the initiation of the earliest step of T cell development in the thymus. *The Journal of experimental medicine* **184,** 1137–1147. doi:10.1084/JEM.184.3.1137 (1996).

309. Besseyrias, V. *et al.* Hierarchy of Notch-Delta interactions promoting T cell lineage commitment and maturation. *Journal of Experimental Medicine* **204,** 331–343. doi:10.1084/jem.20061442 (2007).

310. Hozumi, K. *et al.* Notch signaling is necessary for GATA3 function in the initiation of T cell development. *European Journal of Immunology* **38,** 977–985. doi:10.1002/eji.200737688 (2008).

311. Harly, C. *et al.* A Shared Regulatory Element Controls the Initiation of Tcf7 Expression During Early T Cell and Innate Lymphoid Cell Developments. *Frontiers in Immunology* **11,** 481089. doi:10.3389/fimmu.2020.00470 (2020).

312. Zhou, W., Gao, F., Romero-Wolf, M., Jo, S. & Rothenberg, E. V. Single-cell deletion analyses show control of pro-T cell developmental speed and pathways by Tcf7, Spi1, Gata3, Bcl11a, Erg, and Bcl11b. *Science Immunology* **7.** doi:10.1126/sciimmunol.abm1920 (2022).

313. Isoda, T. *et al.* Non-coding Transcription Instructs Chromatin Folding and Compartmentalization to Dictate Enhancer-Promoter Communication and T Cell Fate. *Cell* **171,** 103–119. doi:10.1016/j.cell.2017.09.001 (2017).

314. Hu, G. *et al.* Transformation of Accessible Chromatin and 3D Nucleome Underlies Lineage Commitment of Early T Cells. *Immunity* **48,** 227–242. doi:10.1016/j.immuni.2018.01.013 (2018).

315. Yoshida, H. *et al.* The cis-Regulatory Atlas of the Mouse Immune System. *Cell* **176,** 897–912. doi:10.1016/j.cell.2018.12.036 (2019).

316. Wang, W. *et al.* TCF-1 promotes chromatin interactions across topologically associating domains in T cell progenitors. *Nature Immunology* **23,** 1052–1062. doi:10.1038/s41590-022-01232-z (2022).

317. Miyazaki, M. *et al.* The opposing roles of the transcription factor E2A and its antagonist Id3 that orchestrate and enforce the naive fate of T cells. *Nature Immunology* **12,** 992–1001. doi:10.1038/ni.2086 (2011).

318. Miyazaki, K. *et al.* The transcription factor E2A activates multiple enhancers that drive Rag expression in developing T and B cells. *Science Immunology* **5.** doi:10.1126/sciimmunol.abb1455 (2020).

319. Engel, I. & Murre, C. E2A proteins enforce a proliferation checkpoint in developing thymocytes. *EMBO Journal* **23,** 202–211. doi:10.1038/sj.emboj.7600017 (2004).

320. Arenzana, T. L., Schjerven, H. & Smale, S. T. Regulation of gene expression dynamics during developmental transitions by the ikaros transcription factor. *Genes and Development* **29,** 1801–1816. doi:10.1101/gad.266999.115 (2015).

321. Shin, B. *et al.* Runx1 and Runx3 drive progenitor to T-lineage transcriptome conversion in mouse T cell commitment via dynamic genomic site switching. *Proceedings of the National Academy of Sciences of the United States of America* **118,** e2019655118. doi:10.1073/pnas.2019655118 (2021).

322. Ng, K. K. *et al.* A stochastic epigenetic switch controls the dynamics of T-cell lineage commitment. *eLife* **7.** doi:10.7554/eLife.37851 (2018).

323. Pease, N. A. *et al.* Tunable, division-independent control of gene activation timing by a polycomb switch. *Cell Reports* **34.** doi:10.1016/j.celrep.2021.108888 (2021).

324. Goldman, N. *et al.* Intrinsically disordered domain of transcription factor TCF-1 is required for T cell developmental fidelity. *Nature Immunology* **24,** 1698–1710. doi:10.1038/s41590-023-01599-7 (2023).

325. Li, P. *et al.* Reprogramming of T cells to natural killer-like cells upon Bcl11b deletion. *Science* **329,** 85–89. doi:10.1126/science.1188063 (2010).

326. Rolink, A. G., Schaniel, C., Bruno, L. & Melchers, F. In vitro and in vivo plasticity of Pax5-deficient pre-B I cells. *Immunology Letters* **82,** 35–40. doi:10.1016/S0165-2478(02)00016-0 (2002).

327. Souabni, A., Cobaleda, C., Schebesta, M. & Busslinger, M. Pax5 promotes B lymphopoiesis and blocks T cell development by repressing Notch1. *Immunity* **17,** 781–793. doi:10.1016/S1074-7613(02)00472-7 (2002).

328. Kim, J. *et al.* Ikaros DNA-binding proteins direct formation of chromatin remodeling complexes in lymphocytes. *Immunity* **10,** 345–355. doi:10.1016/S1074-7613(00)80034-5 (1999).

329. McManus, S. *et al.* The transcription factor Pax5 regulates its target genes by recruiting chromatin-modifying proteins in committed B cells. *EMBO Journal* **30,** 2388–2404. doi:10.1038/emboj.2011.140 (2011).

330. Okuyama, K. *et al.* PAX5 is part of a functional transcription factor network targeted in lymphoid leukemia. *PLoS Genetics* **15.** doi:10.1371/journal.pgen.1008280 (2019).

331. Hill, L. *et al.* Wapl repression by Pax5 promotes V gene recombination by Igh loop extrusion. *Nature* **584,** 142–147. doi:10.1038/s41586-020-2454-y (2020).

332. Astori, A. *et al.* ARID1a Associates with Lymphoid-Restricted Transcription Factors and Has an Essential Role in T Cell Development. *Journal of Immunology* **205.** doi:10.4049/jimmunol.1900959 (2020).

333. Ungerbäck, J., Åhsberg, J., Strid, T., Somasundaram, R. & Sigvardsson, M. Combined heterozygous loss of Ebf1 and Pax5 allows for T-lineage conversion of B cell progenitors. *The Journal of Experimental Medicine* **212,** 1109–1123. doi:10.1084/jem.20132100 (2015).

334. Mandal, M. *et al.* Ras orchestrates exit from the cell cycle and light-chain recombination during early B cell development. *Nature Immunology* **10,** 1110–1117. doi:10.1038/ni.1785 (2009).

335. Johnson, K. *et al.* Regulation of Immunoglobulin Light-Chain Recombination by the Transcription Factor IRF-4 and the Attenuation of Interleukin-7 Signaling. *Immunity* **28,** 335–345. doi:10.1016/j.immuni.2007.12.019 (2008).

336. Sigvardsson, M. Transcription factor networks link B-lymphocyte development and malignant transformation in leukemia. *Genes and Development* **37,** 703–723. doi:10.1101/gad.349879.122 (2023).

337. Mullighan, C. G. *et al.* Genome-wide analysis of genetic alterations in acute lymphoblastic leukaemia. *Nature* **446,** 758–764. doi:10.1038/nature05690 (2007).

338. Kuiper, R. P. *et al.* High-resolution genomic profiling of childhood ALL reveals novel recurrent genetic lesions affecting pathways involved in lymphocyte differentiation and cell cycle progression. *Leukemia 2007 21:6* **21,** 1258–1266. doi:10.1038/sj.leu.2404691 (2007).

339. Prasad, M. A. *et al.* Ebf1 heterozygosity results in increased DNA damage in pro-B cells and their synergistic transformation by Pax5 haploinsufficiency. *Blood* **125,** 4052–4059. doi:10.1182/BLOOD-2014-12-617282 (2015).

340. Arber, D. A. *et al.* International Consensus Classification of Myeloid Neoplasms and Acute Leukemias: integrating morphologic, clinical, and genomic data. *Blood* **140,** 1200–1228. doi:10.1182/blood.2022015850 (2022).

341. Jennings, C. D. & Foon, K. A. *Recent advances in flow cytometry: Application to the diagnosis of hematologic malignancy* 1997. doi:10.1182/blood.v90.8.2863.

342. Craig, F. E. & Foon, K. A. Flow cytometric immunophenotyping for hematologic neoplasms. *Blood* **111,** 3941–3967. doi:10.1182/BLOOD-2007-11-120535 (2008).

343. Heltemes-Harris, L. M. *et al.* Ebf1 or Pax5 haploinsufficiency synergizes with STAT5 activation to initiate acutelymphoblastic leukemia. *Journal of Experimental Medicine* **208,** 1135–1149. doi:10.1084/jem.20101947 (2011).

344. Chan, L. N. *et al.* Metabolic gatekeeper function of B-lymphoid transcription factors. *Nature* **542,** 479–483. doi:10.1038/nature21076 (2017).

345. Xiao, G. *et al.* B-Cell-Specific Diversion of Glucose Carbon Utilization Reveals a Unique Vulnerability in B Cell Malignancies. *Cell* **173,** 470–484. doi:10.1016/j.cell.2018.02.048 (2018).

346. Aster, J. C., Blacklow, S. C. & Pear, W. S. Notch signalling in T-cell lymphoblastic leukaemia/lymphoma and other haematological malignancies. *Journal of Pathology* **223,** 263–274. doi:10.1002/path.2789 (2011).

347. Baldus, C. D. *et al.* High expression of the ETS transcription factor ERG predicts adverse outcome in acute T-lymphoblastic leukemia in adults. *Journal of Clinical Oncology* **24,** 4714–4720. doi:10.1200/JCO.2006.06.1580 (2006).

348. Haydu, J. E. & Ferrando, A. A. *Early T-cell precursor acute lymphoblastic leukaemia* 2013. doi:10.1097/MOH.0b013e3283623c61.

349. Hoang, T., Lambert, J. A. & Martin, R. in *Current Topics in Developmental Biology* 163–204 (Academic Press, 2016). doi:10.1016/bs.ctdb.2016.01.004.

350. McCormack, M. P. *et al.* Requirement for Lyl1 in a model of Lmo2-driven early T-cell precursor ALL. *Blood* **122,** 2093–2103. doi:10.1182/blood-2012-09-458570 (2013).

351. Smith, S. *et al.* LIM Domain Only-2 (LMO2) induces T-cell leukemia by two distinct pathways. *PLoS ONE* **9,** e85883. doi:10.1371/journal.pone.0085883 (2014).

352. 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. doi:10.1182/blood-2004-10-3900 (2005).

353. Zhang, J. *et al.* The genetic basis of early T-cell precursor acute lymphoblastic leukaemia. *Nature* **481,** 157–163. doi:10.1038/nature10725 (2012).

354. Tottone, L. *et al.* A Tumor Suppressor Enhancer of PTEN in T-cell Development and Leukemia. *Blood Cancer Discovery* **2,** 92–109. doi:10.1158/2643-3230.BCD-20-0201 (2021).

355. Montefiori, L. E. *et al.* Enhancer hijacking drives oncogenic bcl11b expression in lineage-ambiguous stem cell leukemia. *Cancer Discovery* **11,** 2846–2867. doi:10.1158/2159-8290.CD-21-0145 (2021).

356. Mullighan, C. G. *et al.* JAK mutations in high-risk childhood acute lymphoblastic leukemia. *Proceedings of the National Academy of Sciences of the United States of America* **106,** 9414–9418. doi:10.1073/pnas.0811761106 (2009).

357. Brady, S. W. *et al.* The genomic landscape of pediatric acute lymphoblastic leukemia. *Nature Genetics* **54,** 1376–1389. doi:10.1038/s41588-022-01159-z (2022).

358. Heltemes-Harris, L. M. *et al.* Identification of mutations that cooperate with defects in B cell transcription factors to initiate leukemia. *Oncogene* **40,** 6166–6179. doi:10.1038/s41388-021-02012-z (2021).

359. Ramamoorthy, S. *et al.* EBF1 and Pax5 safeguard leukemic transformation by limiting IL-7 signaling, Myc expression, and folate metabolism. *Genes & development* **34,** 1503–1519. doi:10.1101/GAD.340216.120 (2020).

360. Schebesta, M., Pfeffer, P. L. & Busslinger, M. Control of pre-BCR signaling by Pax5-dependent activation of the BLNK gene. *Immunity* **17,** 473–485. doi:10.1016/S1074-7613(02)00418-1 (2002).

361. Liu, H., Lee, S., Zhang, Q., Chen, Z. & Zhang, G. The potential underlying mechanism of the leukemia caused by MLL-fusion and potential treatments. *Molecular Carcinogenesis* **59,** 839–851. doi:10.1002/mc.23204 (2020).

362. Aldoss, I. & Song, J. Y. *Extramedullary relapse of kmt2a(mll)-rearranged acute lymphoblastic leukemia with lineage switch following blinatumomab* 2018. doi:10.1182/blood-2018-02-834911.

363. Haddox, C. L. *et al. Blinatumomab-induced lineage switch of B-ALL with t(4:11) (q21;q23) KMT2A/AFF1 into an aggressive AML: pre-and post-switch phenotypic, cytogenetic and molecular analysis* 2017. doi:10.1038/BCJ.2017.89.

364. He, R. R. *et al.* Immunotherapy- (Blinatumomab-) Related Lineage Switch of KMT2A/AFF1 Rearranged B-Lymphoblastic Leukemia into Acute Myeloid Leukemia/Myeloid Sarcoma and Subsequently into B/Myeloid Mixed Phenotype Acute Leukemia. *Case Reports in Hematology* **2019,** 1–4. doi:10.1155/2019/7394619 (2019).

365. Rossi, J. G. *et al.* Lineage switch in childhood acute leukemia: An unusual event with poor outcome. *American Journal of Hematology* **87,** 890–897. doi:10.1002/ajh.23266 (2012).

366. Tirtakusuma, R. *et al.* Epigenetic regulator genes direct lineage switching in MLL/AF4 leukemia. *Blood* **140,** 1875–1890. doi:10.1182/blood.2021015036 (2022).

367. Boeckx, N. *et al.* An inv(16)(p13q22) positive acute myeloid leukaemia relapsing as acute precursor B-cell lymphoblastic leukaemia. *Haematologica* **89** (2004).

368. Jacoby, E. *et al.* CD19 CAR immune pressure induces B-precursor acute lymphoblastic leukaemia lineage switch exposing inherent leukaemic plasticity. *Nature Communications* **7.** doi:10.1038/ncomms12320 (2016).

369. Somasundaram, R. *et al.* Clonal conversion of B lymphoid leukemia reveals cross-lineage transfer of malignant states. *Genes and Development* **30,** 2486–2499. doi:10.1101/gad.285536.116 (2016).

370. Park, J. H., Geyer, M. B. & Brentjens, R. J. *CD19-targeted CAR T-cell therapeutics for hematologic malignancies: Interpreting clinical outcomes to date* 2016. doi:10.1182/blood-2016-02-629063.

371. Gardner, R. *et al.* Acquisition of a CD19-negative myeloid phenotype allows immune escape of MLL-rearranged B-ALL from CD19 CAR-T-cell therapy. *Blood* **127,** 2406–2410. doi:10.1182/blood-2015-08-665547 (2016).

372. Orlando, E. J. *et al.* Genetic mechanisms of target antigen loss in CAR19 therapy of acute lymphoblastic leukemia. *Nature Medicine* **24,** 1504–1506. doi:10.1038/s41591-018-0146-z (2018).

373. Triana, S. *et al.* Single-cell proteo-genomic reference maps of the hematopoietic system enable the purification and massive profiling of precisely defined cell states. *Nature Immunology* **22,** 1577–1589. doi:10.1038/s41590-021-01059-0 (2021).

374. Koretzky, G. A., Kohmetscher, M. A., Kadleck ' And, T. & Weissi, A. Restoration of T cell receptor-mediated signal transduction by transfection of CD45 cDNA into a CD45-deficient variant of the Jurkat T cell line. *The Journal of Immunology* **149,** 1138–1142. doi:10.4049/JIMMUNOL.149.4.1138 (1992).

375. Altin, J. G. & Sloan, E. K. The role of CD45 and CD45-associated molecules in T cell activation. *Immunology and Cell Biology* **75,** 430–445. doi:10.1038/ICB.1997.68 (1997).

376. Huntington, N. D. *et al.* CD45 links the B cell receptor with cell survival and is required for the persistence of germinal centers. *Nature Immunology 2005 7:2* **7,** 190–198. doi:10.1038/ni1292 (2005).

377. Ralph, S. J. & Berridge, M. V. Expression of antigens of the 'T200' family of glycoproteins on hemopoietic stem cells: evidence that thymocyte cell lineage antigens are represented on 'T200'. *The Journal of Immunology* **132,** 2510–2514. doi:10.4049/JIMMUNOL.132.5.2510 (1984).

378. Miller, B. A., Antognetti ' And, G. & Springer, T. A. Identification of cell surface antigens present on murine hematopoietic stem cells (1985).

379. Komuro, K., Itakura, K., Boyse, E. A. & John, M. Ly-5: A new T-lymphocyte antigen system. *Immunogenetics* **1,** 452–456. doi:10.1007/BF01564083/METRICS (1974).

380. Scheid, M. P. & Triglia, D. Further description of the Ly-5 system. *Immunogenetics* **9,** 423–433. doi:10.1007/BF01570435/METRICS (1979).

381. Spangrude, G. J., Heimfeld, S. & Weissman, I. L. Purification and Characterization of Mouse Hematopoietic Stem Cells. *Science* **241,** 58–62. doi:10.1126/SCIENCE.2898810 (1988).

382. Sykes, M. *et al.* Effects of T cell depletion in radiation bone marrow chimeras. III. Characterization of allogeneic bone marrow cell populations that increase allogeneic chimerism independently of graft-vs-host disease in mixed marrow recipients. *The Journal of Immunology* **143,** 3503–3511. doi:10.4049/JIMMUNOL.143.11.3503 (1989).

383. Jinek, M. *et al.* A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337,** 816–821. doi:10.1126/SCIENCE.1225829 (2012).

384. Cong, L. *et al.* Multiplex genome engineering using CRISPR/Cas systems. *Science* **339,** 819–823. doi:10.1126/SCIENCE.1231143 (2013).

385. Shrivastav, M., De Haro, L. P. & Nickoloff, J. A. Regulation of DNA double-strand break repair pathway choice. *Cell research* **18,** 134–147. doi:10.1038/CR.2007.111 (2008).

386. Zhu, H. *et al.* The Lin28/let-7 axis regulates glucose metabolism. *Cell* **147,** 81–94. doi:10.1016/J.CELL.2011.08.033 (2011).

387. Gao, X. *et al.* ES cell pluripotency and germ-layer formation require the SWI/SNF chromatin remodeling component BAF250a. *PNAS* **105,** 6656–6661. doi:10.1073/PNAS.0801802105 (2008).

388. Wilson, T. J., Cowdery, H. E., Xu, D., Kola, I. & Hertzog, P. J. A human CD2 minigene directs CRE-mediated recombination in T cells in vivo. *Genesis* **33,** 181–184. doi:10.1002/GENE.10107 (2002).

389. Siegemund, S., Shepherd, J., Xiao, C. & Sauer, K. hCD2-iCre and Vav-iCre mediated gene recombination patterns in murine hematopoietic cells. *PloS one* **10.** doi:10.1371/JOURNAL.PONE.0124661 (2015).

390. Urbanek, P., Wang, Z.-Q., Fetka, L., Wagner, E. F. & Busslinger, M. Complete Block of Early B Cell Differentiation and Altered Patterning of the Posterior Midbrain in Mice Lacking PaxWB-SAP. *Cell* **79,** 901–912 (1994).

391. Rosenberg, N. & Baltimore, D. A quantitative assay for transformation of bone marrow cells by abelson murine leukemia virus*. *Journal of Experimental Medicine* **143,** 1453–1463. doi:10.1084/jem.143.6.1453 (1976).

392. Alt, F., Rosenberg, N., Lewis, S., Thomas, E. & Baltimore, D. Organization and reorganization of immunoglobulin genes in A-MuLV-transformed cells: Rearrangement of heavy but not light chain genes. *Cell* **27,** 381–390. doi:10.1016/0092-8674(81)90421-9 (1981).

393. Siden, E. J., Baltimore, D., Clark, D. & Rosenberg, N. E. Immunoglobulin synthesis by lymphoid cells transformed in vitro by Abelson murine leukemia virus. *Cell* **16,** 389–396. doi:10.1016/0092-8674(79)90014-X (1979).

394. Burrows, P., Lejeune, M. & Kearney, J. F. *Evidence that murine pre-B cells synthesise mu; heavy chains but no light chains [11]* 1979. doi:10.1038/280838a0.

395. Levitt, D. & Cooper, M. D. Mouse pre-B cells synthesize and secrete $\mu$ heavy chains but not light chains. *Cell* **19,** 617–625. doi:10.1016/S0092-8674(80)80038-9 (1980).

396. Dionne, C. J. *et al.* Subversion of T lineage commitment by PU.1 in a clonal cell line system. *Developmental Biology* **280,** 448–466. doi:10.1016/j.ydbio.2005.01.027 (2005).

397. Graham, F. L., Smiley, J., Russell, W. C. & Nairn, R. Characteristics of a human cell line transformed by DNA from human adenovirus type 5. *Journal of General Virology* **36,** 59–72. doi:10.1099/0022-1317-36-1-59/CITE/REFWORKS (1977).

398. Life Technologies Corporation. *Growth and maintenance of the Flp-In T-REx cell line User Guide* tech. rep. (2020), 5–21.

399. Minowada, J. *et al.* Human leukemia cell lines. Evidence for differentiation toward T- and B-cell axis within a leukemia. *Proc. Am. Assoc. Cancer Res.* **19,** 109–109 (1978).

400. Ghandi, M. *et al.* Next-generation characterization of the Cancer Cell Line Encyclopedia. *Nature* **569,** 503–508. doi:10.1038/s41586-019-1186-3 (2019).

401. Li, H. *et al.* The landscape of cancer cell line metabolism. *Nature Medicine* **25,** 850–860. doi:10.1038/s41591-019-0404-8 (2019).

402. Nusinow, D. P. *et al.* Quantitative Proteomics of the Cancer Cell Line Encyclopedia. *Cell* **180,** 387–402. doi:10.1016/j.cell.2019.12.023 (2020).

403. Molday, R. S., Yen, S. P. & Rembaum, A. Application of magnetic microspheres in labelling and separation of cells. *Nature* **268,** 437–438. doi:10.1038/268437a0 (1977).

404. Guruprasad, K. P. *et al.* in *Flow Cytometry: Principles, Applications, and Clinical Perspectives* 1–25 (Springer Singapore, 2024). doi:10.1007/978-981-97-4553-1{_}1.

405. Coulter, W. H. *Means for counting particles suspended in a fluid* 1949.

406. Dittrich, W. & Göhde, W. *Impulsfluorometrie bei Einzelzellen in Suspensionen* 1969. doi:10.1515/znb-1969-0326.

407. Fulwyler, M. J. Electronic separation of biological cells by volume. *Science* **150,** 910–911. doi:10.1126/science.150.3698.910 (1965).

408. Hulett, H. R., Bonner, W. A., Barrett, J. & Herzenberg, L. A. Cell sorting: Automated separation of mammalian cells as a function of intracellular fluorescence. *Journal of Immunology* **193,** 2045–2047. doi:10.1126/science.166.3906.747 (2014).

409. Herzenberg, L. A., Sweet, R. G. & Herzenberg, L. A. Fluorescence-activated cell sorting. *Scientific American* **234,** 108–117. doi:10.1038/scientificamerican0376-108 (1976).

410. Till, J. E. & A., M. E. A direct measurement of the radiation sensitivity of normal mouse bone marrow cells. *Radiation research* **14,** 213–222. doi:10.2307/3570892 (1961).

411. Becker, A. J., McCulloch, E. A. & Till, J. E. Cytological demonstration of the clonal nature of spleen colonies derived from transplanted mouse marrow cells. *Nature* **197,** 452–454. doi:10.1038/197452a0 (1963).

412. Siminovitch, L., McCulloch, E. A. & Till, J. E. The distribution of colony-forming cells among spleen colonies. *Journal of cellular and comparative physiology* **62,** 327–336. doi:10.1002/jcp.1030620313 (1963).

413. Schofield, R. The relationship between the spleen colony-forming cell and the haemopoietic stem cell. *Blood cells* **4,** 7–25 (1978).

414. Tomita, Y., Sachs, D. H. & Sykes, M. Myelosuppressive Conditioning Is Required to Achieve Engraftment of Pluripotent Stem Cells Contained in Moderate Doses of Syngeneic Bone Marrow. *Blood* **83,** 939–948. doi:10.1182/BLOOD.V83.4.939.939 (1994).

415. Czechowicz, A., Kraft, D., Weissman, I. L. & Bhattacharya, D. Efficient transplantation via antibody-based clearance of hematopoietic stem cell niches. *Science* **318,** 1296–1299. doi:10.1126/science.1149726 (2007).

416. Nakano, T., Kodama, H. & Honjo, T. Generation of Lymphohematopoietic Cells from Embryonic Stem Cells in Culture. *Science* **265,** 1098–1101. doi:10.1126/science.8066449 (1994).

417. Cho, S. K. *et al.* Functional characterization of B lymphocytes generated in vitro from embryonic stem cells. *Proceedings of the National Academy of Sciences of the United States of America* **96,** 9797–9802. doi:10.1073/pnas.96.17.9797 (1999).

418. Vieira, P. & Cumano, A. Differentiation of B lymphocytes from hematopoietic stem cells. *Methods in molecular biology (Clifton, N.J.)* **271,** 67–76. doi:10.1385/1-59259-796-3:067 (2004).

419. Pui, J. C. *et al.* Notch1 expression in early lymphopoiesis influences B versus T lineage determination. *Immunity* **11,** 299–308. doi:10.1016/S1074-7613(00)80105-3 (1999).

420. Jung, T., Schauer, U., Heusser, C., Neumann, C. & Rieger, C. Detection of intracellular cytokines by flow cytometry. *Journal of Immunological Methods* **159,** 197–207. doi:10.1016/0022-1759(93)90158-4 (1993).

421. Howard, A. & Pelc, S. R. Synthesis of desoxyribonucleic acid in normal and irradiated cells and its relation to chromosome breakage. *Heredity* **6,** 261–273. doi:10.1080/09553008514552501 (1953).

422. Zelenin, A. V. *et al.* 7-amino-actinomycin D as a specific fluorophore for DNA content analysis by laser flow cytometry. *Cytometry* **5,** 348–354. doi:10.1002/cyto.990050410 (1984).

423. Schwarting, R., Gerdes, J., Niehus, J., Jaeschke, L. & Stein, H. Determination of the growth fraction in cell suspensions by flow cytometry using the monoclonal antibody Ki-67. *Journal of Immunological Methods* **90,** 65–70. doi:10.1016/0022-1759(86)90384-4 (1986).

424. Vignon, C. *et al.* Flow Cytometric Quantification of All Phases of the Cell Cycle and Apoptosis in a Two-Color Fluorescence Plot. *PLoS ONE* **8,** e68425. doi:10.1371/journal.pone.0068425 (2013).

425. Gurley, L. R., D'Aanna, J. A., Barham, S. S., Deaven, L. L. & Tobey, R. A. Histone Phosphorylation and Chromatin Structure during Mitosis in Chinese Hamster Cells. *European Journal of Biochemistry* **84,** 1–15. doi:10.1111/J.1432-1033.1978.TB12135.X (1978).

426. Andreassen, P. R., Skoufias, D. A. & Margolis, R. L. Analysis of the spindle-assembly checkpoint in HeLa cells. *Methods in molecular biology (Clifton, N.J.)* **281,** 213–225. doi:10.1385/1-59259-811-0:213 (2004).

427. Zhang, G., Gurtu, V., Kain, S. R. & Yan, G. Early detection of apoptosis using a fluorescent conjugate of annexin V. *BioTechniques* **23,** 525–531. doi:10.2144/97233pf01 (1997).

428. Miller, E. Apoptosis measurement by annexin v staining. *Methods in molecular medicine* **88,** 191–202. doi:10.1385/1-59259-406-9:191/FIGURES/8 (2004).

429. Lehn, P., Fabrega, S., Oudrhiri, N. & Navarro, J. *Gene delivery systems: Bridging the gap between recombinant viruses and artificial vectors* 1998. doi:10.1016/S0169-409X(97)00102-6.

430. Coffin, J. M. in *The Retroviridae* 19–49 (Springer, Boston, MA, 1992). doi:10.1007/978-1-4615-3372-6{_}2.

431. Coffin, J. M., Hughes, S. H. & Varmus, H. E. in *Retroviruses* (eds Coffin JM, Hughes SH & Varmus HE) (Cold Spring Harbor Laboratory Press, 1997).

432. Naldini, L. *et al.* In vivo gene delivery and stable transduction of nondividing cells by a lentiviral vector. *Science* **272,** 263–267. doi:10.1126/science.272.5259.263 (1996).

433. Ishino, Y., Shinagawa, H., Makino, K., Amemura, M. & Nakatura, A. Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isoenzyme conversion in Escherichia coli, and identification of the gene product. *Journal of Bacteriology* **169,** 5429–5433. doi:10.1128/jb.169.12.5429-5433.1987 (1987).

434. Barrangou, R. *et al.* CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315,** 1709–1712. doi:10.1126/science.1138140 (2007).

435. Garneau, J. E. *et al.* The CRISPR/cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature 2010 468:7320* **468,** 67–71. doi:10.1038/nature09523 (2010).

436. Meyenberg, M., Ferreira da Silva, J. & Loizou, J. I. *Tissue Specific DNA Repair Outcomes Shape the Landscape of Genome Editing* 2021. doi:10.3389/fgene.2021.728520.

437. Webster, N. J., Green, S., Rui Jin, J. & Chambon, P. The hormone-binding domains of the estrogen and glucocorticoid receptors contain an inducible transcription activation function. *Cell* **54,** 199–207. doi:10.1016/0092-8674(88)90552-1 (1988).

438. Louvion, J. F., Havaux-Copf, B. & Picard, D. Fusion of GAL4-VP16 to a steroid-binding domain provides a tool for gratuitous induction of galactose-responsive genes in yeast. *Gene* **131,** 129–134. doi:10.1016/0378-1119(93)90681-R (1993).

439. Danielian, P. S., White, R., Hoare, S. A., Fawell, S. E. & Parker, M. G. Identification of residues in the estrogen receptor that confer differential sensitivity to estrogen and hydroxytamoxifen. *Molecular Endocrinology* **7,** 232–240. doi:10.1210/mend.7.2.8469236 (1993).

440. Feil, R. *et al.* Ligand-activated site-specific recombination in mice. *Proceedings of the National Academy of Sciences* **93,** 10887–10890. doi:10.1073/PNAS.93.20.10887 (1996).

441. Feil, R., Wagner, J., Metzger, D. & Chambon, P. Regulation of Cre recombinase activity by mutated estrogen receptor ligand-binding domains. *Biochemical and Biophysical Research Communications* **237,** 752–757. doi:10.1006/bbrc.1997.7124 (1997).

442. Gossen, M. & Bujard, H. Tight control of gene expression in mammalian cells by tetracycline-responsive promoters. *Proceedings of the National Academy of Sciences of the United States of America* **89,** 5547–5551. doi:10.1073/pnas.89.12.5547 (1992).

443. Das Atze, T., Tenenbaum, L. & Berkhout, B. Tet-On Systems For Doxycycline-inducible Gene Expression. *Current Gene Therapy* **16,** 156–167. doi:10.2174/1566523216666160524144041 (2016).

444. Gossen, M. *et al.* Transcriptional activation by tetracyclines in mammalian cells. *Science* **268,** 1766–1769. doi:10.1126/science.7792603 (1995).

445. Head, S. R. *et al.* Library construction for next-generation sequencing: Overviews and challenges. *BioTechniques* **56,** 61–77. doi:10.2144/000114133 (2014).

446. Adey, A. *et al.* Rapid, low-input, low-bias construction of shotgun fragment libraries by high-density in vitro transposition. *Genome Biology* **11,** 1–17. doi:10.1186/gb-2010-11-12-r119 (2010).

447. Picelli, S. *et al.* Tn5 transposase and tagmentation procedures for massively scaled sequencing projects. *Genome Research* **24,** 2033–2040. doi:10.1101/gr.177881.114 (2014).

448. Metzker, M. L. *Sequencing technologies the next generation* 2010. doi:10.1038/nrg2626.

449. Hu, T., Chitnis, N., Monos, D. & Dinh, A. Next-generation sequencing technologies: An overview. *Human Immunology* **82,** 801–811. doi:10.1016/j.humimm.2021.02.012 (2021).

450. Heid, C. A., Stevens, J., Livak, K. J. & Williams, P. M. Real time quantitative PCR. *Genome Research* **6,** 986–994. doi:10.1101/gr.6.10.986 (1996).

451. Holland, P. M., Abramson, R. D., Watson, R. & Gelfand, D. H. Detection of specific polymerase chain reaction product by utilizing the 5 $\rightarrow$ 3 exonuclease activity of Thermus aquaticus DNA polymerase. *Proceedings of the National Academy of Sciences of the United States of America* **88,** 7276–7280. doi:10.1073/pnas.88.16.7276 (1991).

452. Hamrouni, A., Aublin, A., Guillaume, P. & Maryanski, J. L. T cell receptor gene rearrangement lineage analysis reveals clues for the origin of highly restricted antigen-specific repertoires. *Journal of Experimental Medicine* **197,** 601–614. doi:10.1084/jem.20021945 (2003).

453. Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L. & Wold, B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature Methods* **5,** 621–628. doi:10.1038/nmeth.1226 (2008).

454. Gertz, J. *et al.* Transposase mediated construction of RNA-seq libraries. *Genome Research* **22,** 134–141. doi:10.1101/gr.127373.111 (2012).

455. Picelli, S. *et al.* Full-length RNA-seq from single cells using Smart-seq2. *Nature Protocols* **9,** 171–181. doi:10.1038/nprot.2014.006 (2014).

456. Qi, L. S. *et al.* Repurposing CRISPR as an RNA-$\gamma$uided platform for sequence-specific control of gene expression. *Cell* **152,** 1173–1183. doi:10.1016/j.cell.2013.02.022 (2013).

457. Gilbert, L. A. *et al.* CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell* **154,** 442. doi:10.1016/j.cell.2013.06.044 (2013).

458. De Wet, J. R., Wood, K. V., DeLuca, M., Helinski, D. R. & Subramani, S. Firefly Luciferase Gene: Structure and Expression in Mammalian Cells. *Molecular and Cellular Biology* **7,** 725–737. doi:10.1128/mcb.7.2.725-737.1987 (1987).

459. Majumder, S., Zhao, Z., Kaneko, K. & DePamphilis, M. L. Developmental acquisition of enhancer function requires a unique coactivator activity. *EMBO Journal* **16,** 1721–1731. doi:10.1093/emboj/16.7.1721 (1997).

460. Roux, K. J., Kim, D. I., Raida, M. & Burke, B. A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells. *Journal of Cell Biology* **196,** 801–810. doi:10.1083/jcb.201112098 (2012).

461. Choi-Rhee, E., Schulman, H. & Cronan, J. E. Promiscuous protein biotinylation by Escherichia coli biotin protein ligase. *Protein Science* **13,** 3043–3050. doi:10.1110/ps.04911804 (2004).

462. Cronan, J. E. *Targeted and proximity-dependent promiscuous protein biotinylation by a mutant Escherichia coli biotin protein ligase* in *Journal of Nutritional Biochemistry* **16** (Elsevier, 2005), 416–418. doi:10.1016/j.jnutbio.2005.03.017.

463. Garner, M. M. & Revzin, A. A gel electrophoresis method for quantifying the binding of proteins to specific DNA regions: Application to components of the Escherichia coli lactose operon regulatory system. *Nucleic Acids Research* **9,** 3047–3060. doi:10.1093/nar/9.13.3047 (1981).

464. Fried, M. & Crothers, D. M. Equilibria and kinetics of lac repressor-operator interactions by polyacrylamide gel electrophoresis. *Nucleic Acids Research* **9,** 6505–6525. doi:10.1093/nar/9.23.6505 (1981).

465. Fried, M. G. Measurement of protein-DNA interaction parameters by electrophoresis mobility shift assay. *Electrophoresis* **10,** 366–376. doi:10.1002/elps.1150100515 (1989).

466. Carey, M. F., Peterson, C. L. & Smale, S. T. Experimental Strategies for the Identification of DNA-Binding Proteins. *Cold Spring Harbor Protocols* **2012,** pdb.top067470. doi:10.1101/PDB.TOP067470 (2012).

467. Johnson, D. S., Mortazavi, A., Myers, R. M. & Wold, B. Genome-wide mapping of in vivo protein-DNA interactions. *Science* **316,** 1497–1502. doi:10.1126/science.1141319 (2007).

468. Solomon, M. J. & Varshavsky, A. Formaldehyde-mediated DNA-protein crosslinking: A probe for in vivo chromatin structures. *Proceedings of the National Academy of Sciences of the United States of America* **82,** 6470–6474. doi:10.1073/pnas.82.19.6470 (1985).

469. Tian, B., Yang, J. & Brasier, A. R. Two-step cross-linking for analysis of protein-chromatin interactions. *Methods in Molecular Biology* **809,** 105–120. doi:10.1007/978-1-61779-376-9{_}7 (2012).

470. Singh, A. A. *et al.* Optimized ChIP-seq method facilitates transcription factor profiling in human tumors. *Life Science Alliance* **2,** e201800115. doi:10.26508/lsa.201800115 (2019).

471. Solomon, M. J., Larsen, P. L. & Varshavsky, A. Mapping proteinDNA interactions in vivo with formaldehyde: Evidence that histone H4 is retained on a highly transcribed gene. *Cell* **53,** 937–947. doi:10.1016/S0092-8674(88)90469-2 (1988).

472. Schmidl, C., Rendeiro, A. F., Sheffield, N. C. & Bock, C. ChIPmentation: Fast, robust, low-input ChIP-seq for histones and transcription factors. *Nature Methods* **12,** 963–965. doi:10.1038/nmeth.3542 (2015).

473. Skene, P. J. & Henikoff, S. An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. *eLife* **6.** doi:10.7554/eLife.21856 (2017).

474. Skene, P. J., Henikoff, J. G. & Henikoff, S. Targeted in situ genome-wide profiling with high efficiency for low cell numbers. *Nature Protocols* **13,** 1006–1019. doi:10.1038/nprot.2018.015 (2018).

475. Kaya-Okur, H. S. *et al.* CUT&Tag for efficient epigenomic profiling of small samples and single cells. *Nature Communications* **10,** 1–10. doi:10.1038/s41467-019-09982-5 (2019).

476. Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nature Methods* **10,** 1213–1218. doi:10.1038/nmeth.2688 (2013).

477. Goryshin, I. Y. & Reznikoff, W. S. Tn5 in vitro transposition. *Journal of Biological Chemistry* **273,** 7367–7374. doi:10.1074/jbc.273.13.7367 (1998).

478. Gangadharan, S., Mularoni, L., Fain-Thornton, J., Wheelan, S. J. & Craig, N. L. DNA transposon Hermes inserts into DNA in nucleosome-free regions in vivo. *Proceedings of the National Academy of Sciences of the United States of America* **107,** 21966–21972. doi:10.1073/pnas.1016382107 (2010).

479. Corces, M. R. *et al.* An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. *Nature Methods* **14,** 959–962. doi:10.1038/nmeth.4396 (2017).

480. Jerkovic´, I. & Cavalli, G. *Understanding 3D genome organization by multidisciplinary methods* 2021. doi:10.1038/s41580-021-00362-w.

481. Langer-Safer, P. R., Levine, M. & Ward, D. C. Immunological methods for mapping genes on Drosophila polytene chromosomes. *Proceedings of the National Academy of Sciences of the United States of America* **79,** 4381–4385. doi:10.1073/pnas.79.14.4381 (1982).

482. Su, J. H., Zheng, P., Kinrot, S. S., Bintu, B. & Zhuang, X. Genome-Scale Imaging of the 3D Organization and Transcriptional Activity of Chromatin. *Cell* **182,** 1641–1659. doi:10.1016/j.cell.2020.07.032 (2020).

483. Sati, S. & Cavalli, G. *Chromosome conformation capture technologies and their impact in understanding genome function* 2017. doi:10.1007/s00412-016-0593-6.

484. Quinodoz, S. A. *et al.* Higher-Order Inter-chromosomal Hubs Shape 3D Genome Organization in the Nucleus. *Cell* **174,** 744–757. doi:10.1016/j.cell.2018.05.024 (2018).

485. Beagrie, R. A. *et al.* Complex multi-enhancer contacts captured by genome architecture mapping. *Nature* **543,** 519–524. doi:10.1038/nature21411 (2017).

486. Chen, Y. *et al.* Mapping 3D genome organization relative to nuclear compartments using TSA-Seq as a cytological ruler. *Journal of Cell Biology* **217,** 4025–4048. doi:10.1083/jcb.201807108 (2018).

487. Dekker, J., Rippe, K., Dekker, M. & Kleckner, N. Capturing chromosome conformation. *Science* **295,** 1306–1311. doi:10.1126/science.1067799 (2002).

488. Hsieh, T. H. S. *et al.* Mapping Nucleosome Resolution Chromosome Folding in Yeast by Micro-C. *Cell* **162,** 108–119. doi:10.1016/j.cell.2015.05.048 (2015).

489. Mumbach, M. R. *et al.* HiChIP: Efficient and sensitive analysis of protein-directed genome architecture. *Nature Methods* **13,** 919–922. doi:10.1038/nmeth.3999 (2016).

490. Fang, R. *et al. Mapping of long-range chromatin interactions by proximity ligation-assisted ChIP-seq* 2016. doi:10.1038/cr.2016.137.

491. Tang, F. *et al.* mRNA-Seq whole-transcriptome analysis of a single cell. *Nature Methods* **6,** 377–382. doi:10.1038/nmeth.1315 (2009).

492. Jaitin, D. A. *et al.* Massively parallel single-cell RNA-seq for marker-free decomposition of tissues into cell types. *Science* **343,** 776–779. doi:10.1126/science.1247651 (2014).

493. Macosko, E. Z. *et al.* Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* **161,** 1202–1214. doi:10.1016/j.cell.2015.05.002 (2015).

494. Zhu, C., Preissl, S. & Ren, B. *Single-cell multimodal omics: the power of many* 2020. doi:10.1038/s41592-019-0691-5.

495. Reyes, M., Billman, K., Hacohen, N. & Blainey, P. C. Simultaneous Profiling of Gene Expression and Chromatin Accessibility in Single Cells. *Advanced Biosystems* **3.** doi:10.1002/adbi.201900065 (2019).

496. Ma, S. *et al.* Chromatin Potential Identified by Shared Single-Cell Profiling of RNA and Chromatin. *Cell* **183,** 1103–1116. doi:10.1016/j.cell.2020.09.056 (2020).

497. 10X Genomics. *Chromium Next GEM Single Cell Multiome ATAC + Gene Expression Reagent Kits User Guide -User Guide -Library Prep -Single Cell Multiome ATAC + Gene Exp. -Official 10x Genomics User Guide* 2021.

498. Bentley, D. R. *et al.* Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* **456,** 53–59. doi:10.1038/nature07517 (2008).

499. Babraham Bioinformatics. *FastQC A Quality Control tool for High Throughput Sequence Data* 2020.

500. Bioinformatics, B. *Trim Galore* 2017.

501. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nature Methods* **9,** 357–359. doi:10.1038/nmeth.1923 (2012).

502. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25,** 1754–1760. doi:10.1093/bioinformatics/btp324 (2009).

503. Dobin, A. *et al.* STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29,** 15–21. doi:10.1093/bioinformatics/bts635 (2013).

504. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25,** 2078–2079. doi:10.1093/bioinformatics/btp352 (2009).

505. Broad Institute. *Picard Tools* 2009.

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

507. Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biology* **9,** 1–9. doi:10.1186/gb-2008-9-9-r137 (2008).

508. Meers, M. P., Tenenbaum, D. & Henikoff, S. Peak calling by Sparse Enrichment Analysis for CUT&RUN chromatin profiling. *Epigenetics and Chromatin* **12,** 1–11. doi:10.1186/s13072-019-0287-4 (2019).

509. Li, Q., Brown, J. B., Huang, H. & Bickel, P. J. Measuring reproducibility of high-throughput experiments. *Annals of Applied Statistics* **5,** 1752–1779. doi:10.1214/11-AOAS466 (2011).

510. Stark, R. & Brown, G. *DiffBind: Differential binding analysis of ChIPSeq peak.* 2011.

511. Servant, N. *et al.* HiC-Pro: An optimized and flexible pipeline for Hi-C data processing. *Genome Biology* **16,** 1–11. doi:10.1186/s13059-015-0831-x (2015).

512. Imakaev, M. *et al.* Iterative correction of Hi-C data reveals hallmarks of chromosome organization. *Nature Methods* **9,** 999–1003. doi:10.1038/nmeth.2148 (2012).

513. Ay, F., Bailey, T. L. & Noble, W. S. Statistical confidence estimation for Hi-C data reveals regulatory chromatin contacts. *Genome Research* **24,** 999–1011. doi:10.1101/gr.160374.113 (2014).

514. Bhattacharyya, S., Chandra, V., Vijayanand, P. & Ay, F. Identification of significant chromatin contacts from HiChIP data by FitHiChIP. *Nature Communications* **10,** 1–14. doi:10.1038/s41467-019-11950-y (2019).

515. Hao, Y. *et al.* Dictionary learning for integrative, multimodal and scalable single-cell analysis. *Nature Biotechnology* **42,** 293–304. doi:10.1038/s41587-023-01767-y (2024).

516. Stuart, T., Srivastava, A., Madad, S., Lareau, C. A. & Satija, R. Single-cell chromatin state analysis with Signac. *Nature Methods* **18,** 1333–1341. doi:10.1038/s41592-021-01282-5 (2021).

517. Bellman, R. *Adaptive control processes: A guided tour* 1–255. doi:10.2307/2343225 (Princeton University Press, 1961).

518. Altman, N. & Krzywinski, M. *The curse(s) of dimensionality this-month* 2018. doi:10.1038/s41592-018-0019-x.

519. Becht, E. *et al.* Dimensionality reduction for visualizing single-cell data using UMAP. *Nature Biotechnology* **37,** 38–47. doi:10.1038/nbt.4314 (2019).

520. Jolliffe, I. Principal Component Analysis. *International Encyclopedia of Statistical Science,* 1094–1096. doi:10.1007/978-3-642-04898-2{_}455 (2011).

521. Pearson, K. LIII. On lines and planes of closest fit to systems of points in space. *Philosophical Magazine Series 1* **2,** 559–572. doi:10.1080/14786440109462720 (1901).

522. Hotelling, H. Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology* **24,** 417–441. doi:10.1037/h0071325 (1933).

523. Ringnér, M. *What is principal component analysis?* 2008. doi:10.1038/nbt0308-303.

524. Leek, J. T. *et al. Tackling the widespread and critical impact of batch effects in high-throughput data* 2010. doi:10.1038/nrg2825.

525. Luecken, M. D. & Theis, F. J. Current best practices in single-cell RNA-seq analysis: a tutorial. *Molecular Systems Biology* **15.** doi:10.15252/msb.20188746 (2019).

526. McInnes, L., Healy, J. & Melville, J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. *arXiv:* **1802.03426** (2018).

527. James & others, M. Some methods for classification and analysis of multivariate observations. *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability* **1,** 281–297 (1967).

528. Köhn, H.-F. & Hubert, L. J. in *Wiley StatsRef: Statistics Reference Online* 1–13 (John Wiley & Sons, Ltd, 2015). doi:10.1002/9781118445112.stat02449.pub2.

529. Blondel, V. D., Guillaume, J. L., Lambiotte, R. & Lefebvre, E. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment* **2008.** doi:10.1088/1742-5468/2008/10/P10008 (2008).

530. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* **15,** 1–21. doi:10.1186/s13059-014-0550-8 (2014).

531. Wald, A. Tests of Statistical Hypotheses Concerning Several Parameters When the Number of Observations is Large. *Transactions of the American Mathematical Society* **54,** 426. doi:10.2307/1990256 (1943).

532. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society Series B: Statistical Methodology* **57,** 289–300. doi:10.1111/j.2517-6161.1995.tb02031.x (1995).

533. Drăghici, S., Khatri, P., Martins, R. P., Ostermeier, G. C. & Krawetz, S. A. Global functional profiling of gene expression. *Genomics* **81,** 98–104. doi:10.1016/S0888-7543(02)00021-6 (2003).

534. Subramanian, A. *et al.* Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences of the United States of America* **102,** 15545–15550. doi:10.1073/pnas.0506580102 (2005).

535. Bailey, T. L. & Gribskov, M. Combining evidence using p-values: Application to sequence homology searches. *Bioinformatics* **14,** 48–54. doi:10.1093/bioinformatics/14.1.48 (1998).

536. Heinz, S. *et al.* Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities. *Molecular Cell* **38,** 576–589. doi:10.1016/J.MOLCEL.2010.05.004 (2010).

537. Stormo, G. D. *DNA binding sites: Representation and discovery* 2000. doi:10.1093/bioinform\-atics/16.1.16.

538. Corces, M. R. *et al.* The chromatin accessibility landscape of primary human cancers. *Science* **362.** doi:10.1126/science.aav1898 (2018).

539. Granja, J. M. *et al.* ArchR is a scalable software package for integrative single-cell chromatin accessibility analysis. *Nature Genetics* **53,** 403–411. doi:10.1038/s41588-021-00790-6 (2021).

540. Lara-Astiaso, D. *et al.* Chromatin state dynamics during blood formation. *Science* **345,** 943–949. doi:10.1126/science.1256271 (2014).

541. Sahin, M. *et al.* HiC-DC+ enables systematic 3D interaction calls and differential analysis for Hi-C and HiChIP. *Nature Communications* **12,** 1–11. doi:10.1038/s41467-021-23749-x (2021).

542. Fulco, C. P. *et al.* Activity-by-contact model of enhancer–promoter regulation from thousands of CRISPR perturbations. *Nature Genetics* **51,** 1664–1669. doi:10.1038/s41588-019-0538-0 (2019).

543. Russell, W. M. S. & Burch, R. L. *The Principles of Humane Experimental Technique* **13,** 500–500. doi:10.5694/j.1326-5377.1960.tb73127.x (John Wiley & Sons, Ltd, 1959).

544. Ryme, J., Asp, P., Böhm, S., Cavellán, E. & Farrants, A. K. Ö. Variations in the composition of mammalian SWI/SNF chromatin remodelling complexes. *Journal of cellular biochemistry* **108,** 565–576. doi:10.1002/JCB.22288 (2009).

545. Raab, J. R., Resnick, S. & Magnuson, T. Genome-Wide Transcriptional Regulation Mediated by Biochemically Distinct SWI/SNF Complexes. *PLoS genetics* **11.** doi:10.1371/JOURNAL.PGEN.1005748 (2015).

546. Taghon, T., Yui, M. A., Pant, R., Diamond, R. A. & Rothenberg, E. V. Developmental and molecular characterization of emerging $\beta$- and $\gamma\delta$-selected pre-T cells in the adult mouse thymus. *Immunity* **24,** 53–64. doi:10.1016/J.IMMUNI.2005.11.012 (2006).

547. Osipovich, O. *et al.* Essential function for SWI-SNF chromatin-remodeling complexes in the promoter-directed assembly of Tcrb genes. *Nature immunology* **8,** 809–816. doi:10.1038/NI1481 (2007).

548. Popay, T. M. & Dixon, J. R. Coming full circle: On the origin and evolution of the looping model for enhancer–promoter communication. *Journal of Biological Chemistry* **298,** 102117. doi:10.1016/J.JBC.2022.102117 (2022).

549. Hnisz, D. *et al.* Activation of proto-oncogenes by disruption of chromosome neighborhoods. *Science* **351,** 1454–1458. doi:10.1126/SCIENCE.AAD9024 (2016).

550. Kragesteen, B. K. *et al.* Dynamic 3D chromatin architecture contributes to enhancer specificity and limb morphogenesis. *Nature Genetics 2018 50:10* **50,** 1463–1473. doi:10.1038/s41588-018-0221-x (2018).

551. Lupiáñez, D. G. *et al.* Disruptions of Topological Chromatin Domains Cause Pathogenic Rewiring of Gene-Enhancer Interactions. *Cell* **161,** 1012–1025. doi:10.1016/J.CELL.2015.04.004 (2015).

552. Stadhouders, R., Filion, G. J. & Graf, T. Transcription factors and 3D genome conformation in cell-fate decisions. *Nature* **569,** 345–354. doi:10.1038/s41586-019-1182-7 (2019).

553. Chen, P. B. *et al.* Systematic discovery and functional dissection of enhancers needed for cancer cell fitness and proliferation. *Cell Reports* **41,** 111630. doi:10.1016/J.CELREP.2022.111630 (2022).

554. Somasundaram, R., Prasad, M. A., Ungerbäck, J. & Sigvardsson, M. Transcription factor networks in B-cell differentiation link development to acute lymphoid leukemia. *Blood* **126,** 144–152. doi:10.1182/BLOOD-2014-12-575688 (2015).

555. Shah, S. *et al.* A recurrent germline PAX5 mutation confers susceptibility to pre-B cell acute lymphoblastic leukemia. *Nature Genetics 2013 45:10* **45,** 1226–1231. doi:10.1038/ng.2754 (2013).

556. Gu, Z. *et al.* PAX5-driven subtypes of B-progenitor acute lymphoblastic leukemia. *Nature Genetics 2019 51:2* **51,** 296–307. doi:10.1038/s41588-018-0315-5 (2019).

557. Vilagos, B. *et al.* Essential role of EBF1 in the generation and function of distinct mature B cell types. *Journal of Experimental Medicine* **209,** 775–792. doi:10.1084/JEM.20112422 (2012).

558. Györy, I. *et al.* Transcription factor Ebf1 regulates differentiation stage-specific signaling, proliferation, and survival of B cells. *Genes & Development* **26,** 668–682. doi:10.1101/GAD.187328.112 (2012).

559. Liu, G. J. *et al.* Pax5 loss imposes a reversible differentiation block in B-progenitor acute lymphoblastic leukemia. *Genes & Development* **28,** 1337–1350. doi:10.1101/GAD.240416.114 (2014).

560. Hemming, S. *et al.* EZH2 and KDM6A act as an epigenetic switch to regulate mesenchymal stem cell lineage specification. *Stem Cells* **32,** 802–815. doi:10.1002/stem.1573 (2014).

561. Qin, K. *et al.* Molecular basis of polycomb group protein–mediated fetal hemoglobin repression. *Blood* **141,** 2756–2770. doi:10.1182/blood.2022019578 (2023).

562. Yu, D. *et al.* Oscillation between B-lymphoid and myeloid lineages in Myc-induced hematopoietic tumors following spontaneous silencing/reactivation of the EBF/Pax5 pathway. *Blood* **101,** 1950–1955. doi:10.1182/blood-2002-06-1797 (2003).

563. Ferreir´ os-Vidal, I. *et al.* Genome-wide identification of Ikaros targets elucidates its contribution to mouse B-cell lineage specification and pre-B-cell differentiation. *Blood* **121,** 1769–1782. doi:10.1182/blood-2012-08-450114 (2013).

564. O'Riordan, M. & Grosschedl, R. Coordinate regulation of B cell differentiation by the transcription factors EBF and E2A. *Immunity* **11,** 21–31. doi:10.1016/S1074-7613(00)80078-3 (1999).

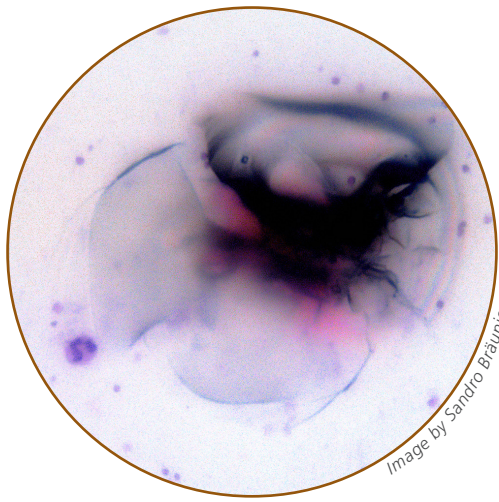
565. Åhsberg, J. *et al.* Early B-cell factor 1 regulates the expansion of B-cell progenitors in a dose-dependent manner. *Journal of Biological Chemistry* **288,** 33449–33461. doi:10.1074/jbc.M113.506261 (2013).

566. Yu, Y. *et al.* Bcl11a is essential for lymphoid development and negatively regulates p53. *Journal of Experimental Medicine* **209,** 2467–2483. doi:10.1084/jem.20121846 (2012).

# Part II

# RESEARCH PAPERS

Printed by Media-Tryck, Lund 2025 NORDIC SWAN ECOLABEL 3041 0903




Just as Hamlet, on his restless search for meaning and truth, was confronted by a ghost, I too encountered a ghost during my PhD. Not in a dramatic Shakespearean fashion, but rather on a Hematoxylin & Eosin stained slide of my own blood, prepared for a public outreach event. Unlike Hamlet's conveniently informative ghost, my "ghost of blood" brought no revelation and is, by all accounts, nothing more than a blob of dye. Yet, I let it stand as a modest symbol of this thesis (and maybe science at large), always puzzling, often leading nowhere, but at times nudging us toward insight.

LUND UNIVERSITY

FACULTY OF MEDICINE

Department of Laboratory Medicine

Lund University, Faculty of Medicine
Doctoral Dissertation Series 2025:122
ISBN 978-91-8021-775-0
ISSN 1652-8220