Inaugural dissertation

for

obtaining the doctoral degree

of the

Combined Faculty of Mathematics, Engineering and Natural Sciences

of the

Ruprecht - Karls - University

Heidelberg

Presented by

MPharm MBA Donnacha Fitzgerald

born in Cork, Ireland

Oral examination: 6th March 2024

Tumor Evolution through Differentiation in B-cell Lymphomas

Referees: Prof. Dr. Benedikt Brors

Prof. Dr. Jan Korbel

Abstract

Since the dawn of multicellular life, cells have diversified through the differentiation of functionally distinct cell types. These cell types give rise to different cancer types and subtypes, with cell of origin central to *inter*tumor heterogeneity. For example, B-cell non-Hodgkin lymphomas are tied to distinct stages of B-cell maturation. In contrast, *intra*tumor heterogeneity, intrinsic to cancer evolution and treatment response, is typically considered distinct from cell-type differentiation. In this thesis, I challenged this view by investigating whether cell-type differentiation influences intratumor heterogeneity in B-cell lymphomas.

Through joint single-cell transcriptomic and surface epitope profiling of diffuse large B-cell (DLBCL), mantle cell (MCL), follicular (FL), and marginal zone (MZL) lymphomas from 43 patients, alongside 8 reactive lymph nodes, I found that individual tumors were comprised of multiple B-cell maturation states, suggesting ongoing differentiation from the cell of origin. Tumor maturation state composition varied across and within lymphoma entities, blurring entity boundaries. Cell-of-origin classifiers revealed the presence of mixed germinal center (GCB) and activated B-cell (ABC) clinical subtypes within DLBCL and FL tumors. Varying tumor maturation state composition over time indicates tumors evolve through differentiation, maintained by differentially active maturation transcription factors observed between tumor maturation states. Highly multiplexed immunohistochemistry revealed intratumor maturation states occupy distinct spatial niches, differing in immune infiltration while maintaining maturation-associated cellular interactions. Intratumor maturation states were subject to copy number variation and showed varying expression patterns of mutated genes, suggesting that differentiation and genetic variation are interconnected in cancer.

My findings put forward a model of tumor evolution in which cancer cells achieve diversification through cell-type differentiation. Mutability of cell types in cancer poses challenges and opportunities for cancer diagnosis, whereby tumors are not tied to their cell of origin, and cancer treatment, which may need to account for diverse and evolving spectra of cancer cell types. These insights open a wealth of future research avenues in clonal evolution, drug resistance, and precision medicine across cancers.

Zusammenfassung

Seit den Anfängen des mehrzelligen Lebens haben sich die Zellen durch die Differenzierung in funktionell unterschiedlicher Zelltypen diversifiziert. Diese Zelltypen führen zu verschiedenen Krebsarten und Subtypen, wobei die Ursprungszelle für die Heterogenität zwischen den Tumoren von zentraler Bedeutung ist. So sind beispielsweise B-Zell-Non-Hodgkin-Lymphome an unterschiedliche Stadien der B-Zell-Reifung gebunden. Im Gegensatz dazu wird die intratumorale Heterogenität, welchedie sowohl für diemit der Krebsentwicklung als auchund dessendem Ansprechen auf einedessendie Behandlung essentiell istzusammenhängt, in der Regel als von der Zelltypendifferenzierung unabhängig betrachtet. In dieser Arbeit habe ich mich kritisch mit dieser Frage auseinandergesetzthabe ich diese Ansicht in Frage gestellt, indem ich untersucht habe, ob die Zelltypdifferenzierung die Intratumorheterogenität bei B-Zell-Lymphomen beeinflusst.

Durch parallelegemeinsame Einzelzell-Transkriptom- und Oberflächenepitop-Profilierung von diffusen großzelligen B-Zell-Lymphomen (DLBCL), Mantelzell-Lymphomen (MCL), follikulären Lymphomen (FL) und Marginalzonen-Lymphomen (MZL) von 43 Patienten sowie 8 reaktiven Lymphknoten fand ich heraus, dass einzelne Tumore aus mehreren B-Zell-Reifungszuständen bestehenbestanden, was auf eine fortlaufende Differenzierung ausgehend von der Ursprungszelle hindeutet. Die Zusammensetzung der Tumorreifungsstadien variierte zwischen und innerhalb von Lymphom-Entitäten und verwischte die Entitätsgrenzen. Die Klassifizierung der Ursprungszellen lieferte Hinweise aufergab das Vorhandensein von gemischtern Keimzentren (GCB) und aktiviertern B-Zellen (ABC) als klinische Subtypen bei DLBCL und FL-Tumoren. Die variableunterschiedliche Zusammensetzung der Tumorreifungsstadien im Laufe der Zeit deutet darauf hin, dass sich die Tumoren durch Differenzierung weiterentwickeln, welchedie durch unterschiedlich aktive Reifungstranskriptionsfaktoren zwischen den Tumorreifungsstadien aufrechterhalten wird. Hochmultiplex-Immunhistochemie zeigte, dass intratumorale Reifungsstadien verschiedeneunterschiedliche räumliche Nischen besetzen, die sich in der Immuninfiltration unterscheiden, während reifungsassoziierte zelluläre Interaktionen erhalten bleiben. Intratumorale Reifungsstadien unterlagen einer Variation der Kopienzahlvariation und wiesen diverseunterschiedliche Expressionsmuster mutierter Gene auf, was darauf hindeutet, dass Differenzierung und genetische Variation bei Krebs miteinander verbunden sind.

Meine Ergebnisse stellen ein Modell der Tumorevolution vor, bei dem Krebszellen die

Zelldtyp-Differenzierung missbrauchen, um eine Diversifizierung zu erreichen. Daraus ergeben sich Herausforderungen und Chancen sowohl für die Krebsdiagnose, bei der Tumore nicht an ihre Ursprungszelle gebunden sind, als auchund für die Krebsbehandlung, welchedie möglicherweise den vielfältigen und sich entwickelnden Spektren von Tumorzelltypen Rechnung tragen muss. Diese Erkenntnisse eröffnen eine Fülle künftiger Forschungsmöglichkeiten in den Bereichen klonaler Evolution, Arzneimittelresistenz und Präzisionsmedizin bei verschiedenen Krebsarten.

Acknowledgments

Science does not occur in a vacuum, and I have been fortunate to have benefitted from the influence of several people across time and space, shaping the maturation, differentiation, and evolution of this work and myself.

Firstly, I would like to thank my PhD supervisors, Wolfgang Huber and Sascha Dietrich, for putting their trust in me to pursue this work despite transitioning from distant fields. I'm grateful for their mentorship, meticulous feedback, and patience, even when I was headstrong at times. I appreciate that they maintained faith and personal investment in my work over the years, fueling my scientific and personal growth.

I'm grateful to my thesis advisory committee members, Benedikt Brors, Jan Korbel, Christoph Merten, and Judith Zaugg, for their wise counsel and thoughtful suggestions, providing a regular positive and reassuring influence on my PhD.

I would like to thank the many talented collaborators I acknowledged throughout this thesis and the funding sources (BMBF, the ERC, and the EMBL PhD program) who made this work possible. I'm grateful to those who dedicated internships to this project, Linsha Li, Anastasiia Horlova, and Erin Chung, whose enthusiasm helped to sustain mine. I would also like to thank the world-class core facility staff at EMBL Genecore and DKFZ scOpenLab for their crucial technical assistance and advice throughout this project. I'm grateful to Stefan Peidli for providing thoughtful feedback on this thesis.

Being part of two large labs with diverse expertise and skills was an invaluable learning experience, which I owe to my lab mates who also added fun and banter to everyday work. I would especially like to thank Simone Bell for being so generously supportive throughout my PhD, making me always feel at home in the lab, along with Thomas Schwarzl, who has been a valuable mentor to me since I joined the lab.

I would also like to thank past teachers, professors, and research mentors, including Christoph Merten, Christoph Antz, Markus Jeschek, Cormac Gahan, and Christian Weber, whose advice and encouragement helped stimulate me to become a scientist.

I'm grateful to EMBL for creating such an inspiring and supportive scientific and social environment, which I feel privileged to be part of. I would like to thank my PhD cohort and all the friends I've made in my PhD journey, especially Ahmad, Florian, Guido, Julia, Karel,

Luca, Nadine, and Stefan, for their endearing camaraderie and for creating memories that I will cherish for the rest of my life.

I would like to wholeheartedly thank my partner, Luoyan, for being there for me in thick and thin as a regular source of heartfelt support, grounded advice, and fun respite since this odyssey began.

I'm infinitely grateful to my family, who are the best I could have asked for, including my siblings Danielle, Dylan, Adam, and Cecilia, and especially my parents, Catherine and John, for always being behind me in pursuing my ambitions while giving me the space to find myself.

Contents

Introduction

This thesis explores if and how cell-type differentiation influences the variation and evolution of cancers, which I studied in B-cell lymphomas[1]. To address this subject, I first introduce the role of intratumor heterogeneity (ITH) in cancer evolution. I then outline the cell-type differentiation processes responsible for normal cellular diversification, with different cancer subtypes arising from distinct cells of origin. Bridging these topics, I then consider a potential relationship between cell-type differentiation and ITH.

Cancer as an evolutionary process

"For a biologist, the alternative to thinking in evolutionary terms is not to think at all."

Peter Medawar[2]

Cancer can be conceptualized as an evolutionary process, echoing the Darwinian characteristics governing the natural selection of species; Overproduction and struggle for existence, inheritance, variation, natural selection, and speciation[3].

Cancer cells inherently undergo uncontrolled proliferation, a process which is termed clonal expansion[4,5], leading to a struggle for scarce resources such as space, oxygen, and nutrients in the tumor microenvironment. The proliferation rate of cancer cells can far exceed the tumor's growth rate[6], such that the vast majority of cancer cells are unable to survive or continue proliferation[7].

As the genetic code is passed from cell to cell, so too are the somatic mutations that accumulate, including those that give rise to malignancy and mutations that confer a survival advantage over other cancer cells. The heritable nature of these advantageous mutations underpins the pathological and evolutionary progression of cancer[8]. The cell type from which a tumor arises, known as the cell of origin, also shapes the tumor phenotype, with distinct cancer types arising from different cell types[9,10].

Diversification provides evolutionary advantage by creating the foundation for adaptation and enabling the population to better utilize the resources of its environment. Variation arises within tumors in the form of intratumor heterogeneity, where phenotypically distinct subpopulations of cancer cells emerge from a common progenitor. Cancer cells can diversify

through acquiring different genetic mutations[4,5], but also from non-genetic variation on the transcriptomic, epigenetic, and microenvironmental levels[11]. This diversification expands the pool for selection, promoting treatment resistance and cancer progression.

The tumor microenvironment acts as a selective landscape, whereby cancer cells must adapt to evade the immune response and survive through resource scarcity. Clones with advantageous mutations outcompete others, leading to clonal expansion. This selective growth advantage is a cornerstone of the clonal evolution model of cancer[5]. Although treatment may eradicate the majority of cancer cells, selection for treatment-resistant subclones is a frequent cause of cancer relapse and progression[12,13]. In some ways, this is analogous to the development of resistance or vaccine escape in microbes, such as SARS-CoV-2[14].

Natural selection of organisms ultimately leads to speciation, where populations diverge into distinct species uniquely adapted to their environments. An analogous phenomenon occurs in tumors, where selection gives rise to distinct tumor subpopulations uniquely adapted to their microenvironments, including immune pressures, spatial constraints, and treatment exposures, thereby promoting metastasis and progression[15,16].

The characteristics and effects of intratumor heterogeneity

With intratumor heterogeneity (ITH) central to cancer evolution, a growing number of studies have examined its characteristics and effects. ITH was initially described exclusively as genetic variation among cancer cells arising from the accumulation of different mutations throughout their clonal evolution[5]. This is intuitive as cancers arise from genetic variation, through the onset of oncogenic driver mutations such as TP53, KRAS, and EGFR that are critical to their malignant transformation and precipitate further mutations. Tumors can exhibit a range of mutations, from single nucleotide variations to large chromosomal rearrangements, which contribute to the genetic diversity of tumors[17].

In addition to genetic variation, however, numerous studies have shown that non-genetic intratumor heterogeneity is also abundant. This includes variation observed on the epigenetic, transcriptional, morphological, and metabolic levels, which are intricately connected in driving tumor phenotype [11,18–20]. Large-scale transcriptional studies across cancers have highlighted the variable nature of cancer cells in proliferation and stress exposure and

response[21]. Another important axis of ITH is the tumor microenvironment, consisting of variable composition and spatial patterns of stromal cells, immune cells, and the extracellular matrix. Microenvironmental heterogeneity influences the development of ITH by creating spatial niches with differential survival signals and immune infiltration among cells of the same tumor[22].

The abundance of ITH observed across cancer types underscores the key role it plays in cancer pathogenesis, survival, and progression. As ITH leads to a greater diversity of phenotypes, each acquiring different mutations or states promoting evasion of specific therapies, it is a major contributor to treatment resistance[18]. In B-cell lymphomas, transcriptional subpopulations of the same tumor showed differential response to a broad range of drug classes[23]. Conversely, drug treatment can precipitate the emergence of diverse clonal fates[24]. Heterogeneity within the primary tumor can also influence metastatic potential, as subpopulations with specific genetic alterations may be more adept at invading or colonizing distant organs[25]. Therefore, the extent of ITH has been correlated with poor prognosis in various cancer types[26].

The complex and progressive nature of ITH poses significant challenges for cancer diagnosis and treatment. Routine diagnostic methods such as biopsy or bulk genotyping rarely account for ITH, leading to inaccurate prognosis and ineffective therapeutic strategies[27]. The presence of heterogeneous phenotypes in a single patient with different therapeutic sensitivities adds further complexity to the choice of treatment for personalized medicine. Combination therapy (e.g. chemoimmunotherapy combinations like R-CHOP) can help reduce the risk of relapse by targeting multiple survival mechanisms simultaneously but adds to the treatment burden, risk of adverse events, and costs, so drug combinations must be selected carefully[28].

Intratumor heterogeneity is a central aspect of cancer pathogenesis and evolution with profound implications for the understanding, diagnosis, and treatment of cancer. Addressing the challenges posed by ITH requires a better understanding of the processes which govern it.

Cellular diversification via cell-type differentiation

Before studying the processes governing ITH, it is helpful to consider how cells diversify in a healthy context. Cells diversify throughout development, from a unicellular zygote to a multicellular organism with functionally specialized tissues and organs. Here, although

genetic variation can arise from accumulating somatic mutations, this is too unstable to cover the vast multitude of roles that cells need to play in an organism. The most dominant form of cellular diversification is the emergence of different cell types through differentiation, which is intricately regulated by a network of signaling pathways, transcription factors, and epigenetic modifications in response to environmental interactions[29].

The diversification of cell types through differentiation arose with multicellular life itself, and is critical for the complex organization, functionality, and evolutionary success of eukaryotic organisms. The genome contains the necessary genes for all cellular functions of an organism, but these functions are too diverse for a single cell to realize. Therefore, analogous to the division of labor into diverse professional specializations in societies, from architects to zoologists, cells specialize through preferentially activating different elements of the genome to fulfil diverse physiological roles. The human body consists of over 200 functionally distinct cell types making up our tissues and organs[29], with diverse characteristics ranging from neurons for carrying electrical signals across the body to muscle cells specialized for different forms of movement. Cell types are not discrete classes but exist on a continuous tree of cell-type lineages, including many intermediary states and subtypes that are ever being refined due to advancing single-cell technologies and vast single-cell atlases such as *Tabula Sapiens*[30].

Differentiation is not a unidirectional or deterministic process. Cellular plasticity allows differentiated cells under certain conditions to de-differentiate or trans-differentiate, contributing further to cellular diversity[31,32]. For instance, in regenerative medicine, this plasticity is harnessed to reprogram differentiated cells into pluripotent stem cells, which can then be induced to form different cell types[33].

The cell of origin model of *inter*-tumor heterogeneity

Cell-type differentiation has implications for the pathogenesis of cancers as distinct cancer types and subtypes are thought to arise from different cell types and states. This concept of ‘cell of origin’, the cell type from which a tumor originates, has emerged as a fundamental principle of cancer initialization and progression. Cancer cells usually retain characteristics of their cell of origin, which therefore plays a crucial role in dictating the phenotypic characteristics and behavior of the resulting tumor, influencing aspects such as growth rate, invasiveness, response to therapy, and overall prognosis[9]. Due to the vast diversity of cell

types that can undergo malignant transformation, cell of origin is considered central to the heterogeneity of cancer types and subtypes across tissues and organs. Cell of origin patterns have been shown to dominate the molecular classification of tumors and cancer types[10], highlighting the central role it plays in *inter*-tumor heterogeneity.

Distinct cancer subtypes have been associated with cells of origin at different stages of their differentiation lineages. For example, different subtypes of B-cell lymphomas are tied to distinct stages of B-cell development[34–36], acute leukemia subtypes originate from hematopoietic stem cells or differentiated progenitor cells[37], and pancreatic ductal adenocarcinoma subtypes arise from ductal or acinal cells[38].

The concept of the cell of origin is closely tied to the cancer stem cell (CSC) theory. CSCs, a small subpopulation within tumors, possess the ability to self-renew and drive tumor growth[39]. The origin of these CSCs, whether they arise from normal stem cells or from more differentiated cells that acquire stem-like properties, is a subject of intense research. The nature of the cell of origin can profoundly influence the characteristics of these CSCs and, by extension, the behavior of the cancer[40].

The cell of origin does not act in isolation but interacts with its surrounding microenvironment, which can modulate its microenvironment and subsequent malignant progression[41]. For example, in solid tumors, interactions with stromal cells and the extracellular matrix can influence the tumor phenotype that arises from the cell of origin.

The cell of origin has practical implications for cancer treatment. Therapies targeting specific pathways or mutations effective in one subtype of cancer may be ineffective in another, depending on the cell of origin. For example, leukemias originating from hematopoietic stem cells are less sensitive to chemotherapy compared to leukemias derived from myeloid progenitors, which is mediated by differences in apoptotic priming and p53 loss of function[42].

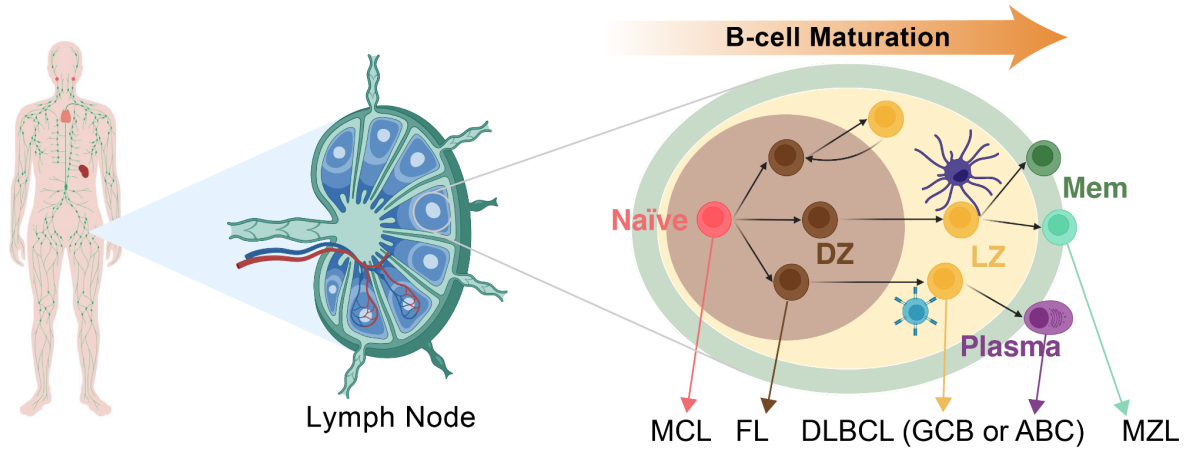
B-cell lymphomas and B-cell maturation

Cell of origin has been well studied in nodal B-cell non-Hodgkin lymphomas (B-NHL), a heterogeneous set of cancers causing over 200,000 deaths annually[43]. B-NHL entities are thought to arise from varying stages of differentiation along the B-cell maturation lineage in the lymph node[34].

The B-cell maturation process in secondary lymphoid organs such as lymph nodes is a key component of B-cell development where B-cell receptor antigen affinity is honed. To facilitate this process, nodal B cells assemble into follicular structures, consisting of a germinal center (GC), containing a dark zone and a light zone, surrounded by a mantle zone. Naïve B cells undergoing T-cell-dependent activation migrate to the GC, where they undergo somatic hypermutation of the B-cell receptor and proliferation in the dark zone (DZ) as centroblasts, followed by selection for improved affinity in the light zone (LZ) as centrocytes by antigen-presenting follicular dendritic (FDC) and follicular helper T-cells (TFH). During this iterative process, termed the GC reaction, selected B cells are stimulated to differentiate into memory B cells (Mem) or plasma cells[44].

B-NHL entities are associated with a cell of origin at different stages of B-cell maturation, including typically pre-GC origin mantle cell lymphoma (MCL) resembling naïve B cells, GC origin follicular lymphoma (FL) resembling DZ or LZ cells, GC origin (GCB) or post-GC activated B cell (ABC) origin diffuse large B-cell lymphoma (DLBCL) resembling LZ or plasma cells respectively, and post-GC origin nodal marginal zone lymphoma (MZL) typically resembling memory B cells (Mem)[34,45,46].

Figure 1: The B-cell maturation process in lymph nodes



Schematics are shown of the human lymphatic system and its large network of lymph nodes (left), a lymph node showing the inner medulla surrounded by the outer cortex containing B-cell follicles (center), and B-cell maturation states in B-cell follicles (right), consisting of a germinal center with a dark zone (brown) and light zone (yellow) surrounded by a mantle zone (green), and their associated B-NHL entities (below). Maturation state annotations: Naïve = Naïve B cells, DZ = Centroblasts from the dark zone of the germinal center, LZ = Centrocytes from the light zone of the germinal center, Mem IgM = IgD+ and IgM+ memory B cells, Mem IgG = class-switched (IgG+ or IgA+) memory B cells, Plasma = plasma cells. Schematics were created with

BioRender[47]. Adapted from Fitzgerald et al. 2023[1], originally produced by myself.

These B-NHL entities show large differences in their clinical courses, ranging from the more indolent FL and MZL to the more aggressive MCL and DLBCL[1]. This emphasizes the clinical impact of cell of origin. In addition, however, high variability in treatment response and survival is observed among patients with the same B-NHL entity[36,48], which has also been associated different cells of origin. For example, DLBCL ABC subtype (from post-GC B cells) shows more frequent relapse to standard chemotherapy treatment than DLBCL GCB subtype (from GC B cells)[36]. This link between the maturation state of DLBCL tumors and their clinical outcomes has been recently recognized as extending beyond the GCB and ABC dichotomy, with poorer outcomes in DZ- vs LZ-associated tumors[49].

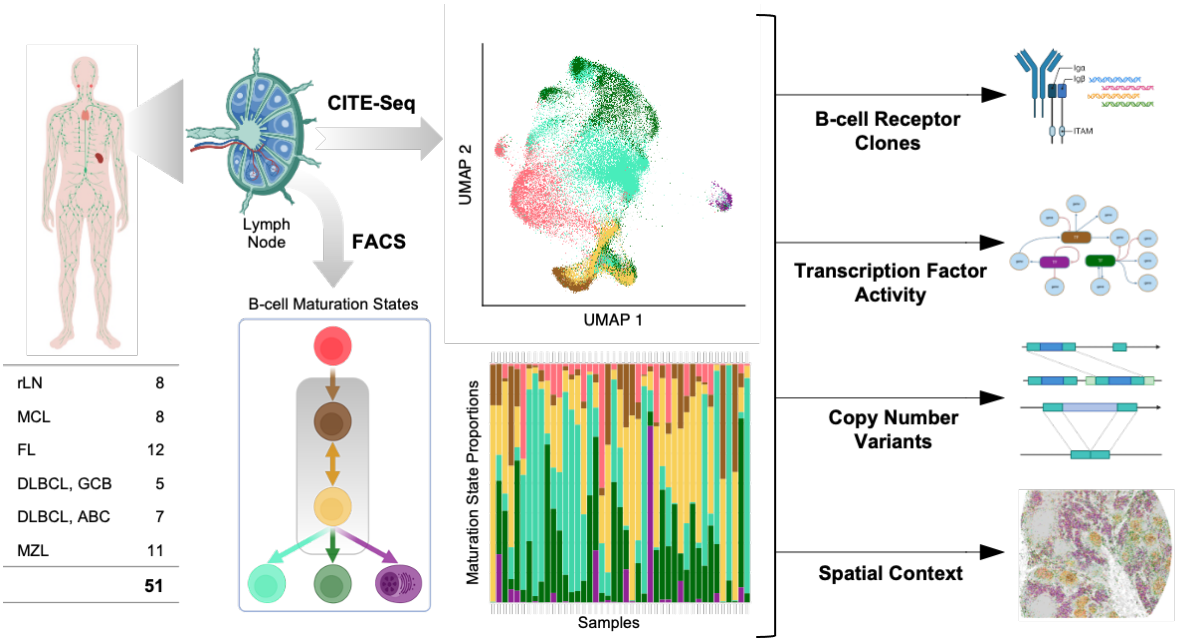
Concurrently, ITH is increasingly recognized as an important phenomenon in B-NHL[23]. Differences in treatment response have been observed among subpopulations within the same tumor. For example, a CD32 high subpopulation of an FL tumor responded better to panabinostat, romidepsin, or vorinostat than its CD32 low counterpart, which was more susceptible to pomalidomide or ibrutinib. Likewise, a CD48/CD62L high DLBCL subpopulation responded better to cladribine or vincristine, while its CD48/CD62L low counterpart was more susceptible to acalabrutinib or dasatinib[23]. This emphasizes the importance of combinatorial therapy that covers distinct susceptibilities within a tumor for successful tumor eradication. To anticipate the different phenotypes that emerge within tumors for devising combinatorial therapy strategies, understanding what drives ITH and tumor evolution is key.

The influence of cell-type differentiation on intratumor heterogeneity

Cell-type differentiation gives rise to cellular diversity among healthy cells, while it is also recognized as central to the diversity of cancers through the cell-of-origin model of *inter*-tumor heterogeneity. As exemplified in B-NHL, distinct cells of origin along differentiation axes leads to substantial biological and clinical heterogeneity between tumors. Meanwhile, ITH is considered a key mediator of treatment resistance and tumor evolution through phenotypic adaptation, although its relationship with the differentiation processes driving cell-type and cancer-type diversity is currently unclear. Studying this relationship could pose

important insights into cancer pathogenesis, variation, and evolution. B-NHL serves as an ideal model to investigate this as it has been well studied with regard to the B-cell maturation differentiation lineage while exhibiting substantial biological and clinical heterogeneity. To this end, I set out to perform a single-cell multi-modal and spatial characterization of B-NHL in relation to B-cell maturation.

Figure 2: Study Design



Overview of the study design, including the distribution of disease entities in the patient cohort and the data types included. Entities: reactive lymph nodes (rLN), mantle cell lymphoma (MCL), follicular lymphoma (FL), germinal center and non-germinal center diffuse large B-cell lymphoma (DLBCL, GCB/non-GCB), and marginal zone lymphoma. FACS = flow-activated cell-sorting. CITE-Seq = cellular indexing of transcriptomes and epitopes by sequencing. Illustrations were created with BioRender[47].

I generated a CITE-Seq (cellular indexing of transcriptomes and epitopes)[50] dataset together with Mareike Knoll and Tobias Roider for joint single-cell transcriptomic and surface protein profiling of reactive lymph nodes (rLN), MCL, FL, DLBCL (GCB and ABC), and MZL samples from 51 patients. I first annotated B-cell maturation states in the rLN samples (n=8). In parallel, Artur Kibler performed flow-activated cell sorting (FACS) of B-cell maturation states in rLN samples (n=5) for RNA sequencing, which I analyzed with Bettina Budeus to generate gene expression profiles for each B-cell maturation state identified and validate the annotations in the CITE-Seq data. I then leveraged the rLN reference to classify B-cell maturation states in cancer cells in the CITE-Seq tumor samples (n=43), which I isolated from non-malignant B cells based on light-chain restriction. Johannes Mammen generated

CITE-Seq data from longitudinal samples in 3 patients, which enabled me to analyze the shifts in B-cell maturation states over time. To understand how gene regulatory networks of maturation states are affected over time, I analyzed transcription factor activity based on gene expression in the CITE-Seq dataset together with Anna Mathioudaki. To link phenotype with genetics, Verena identified tumor variants with DNA sequencing, and I inferred copy number variation on the single-cell level based on gene expression in the CITE-Seq dataset. To profile the spatial distribution of maturation states, Marc-Andrea Bärtsch generated CODEX (co-detection by indexing)[51] highly multiplexed immunohistochemistry data with 52 markers for 18 patients in our cohort. Using the annotated CITE-Seq data for each sample, I classified B-cell maturation states in the CODEX data. Together with Anastasiia Horlova, Erin Chung, and Harald Vöhringer, I characterized the spatial niches and microenvironmental interactions of B-cell maturation states. This large-scale single-cell characterization of intratumor heterogeneity in B-NHL on the transcriptional, protein, genetic, and spatial levels provides unprecedented insights into the role of differentiation in tumor variation and evolution[1].

Results

This section was adapted from Fitzgerald et al. 2023[1], which was originally written by me.

Single cell sequencing of lymph node samples

This study included reactive lymph node (rLN, n = 8), mantle cell lymphoma (MCL, n = 8), follicular lymphoma (FL, n = 12), germinal center (GCB, n = 5) or activated B cell (ABC/non-GCB, n = 7) diffuse large B-cell lymphoma (DLBCL), and marginal zone lymphoma (MZL, n = 11) samples from a total of 51 patients. Of the B-NHL samples, 20 were collected at the time of initial diagnosis and 23 were from patients who had previously undergone one or more lines of systemic treatment. Relapse samples were collected at least 3 months after cessation of systemic treatment. Patient characteristics are summarized in Table 1.

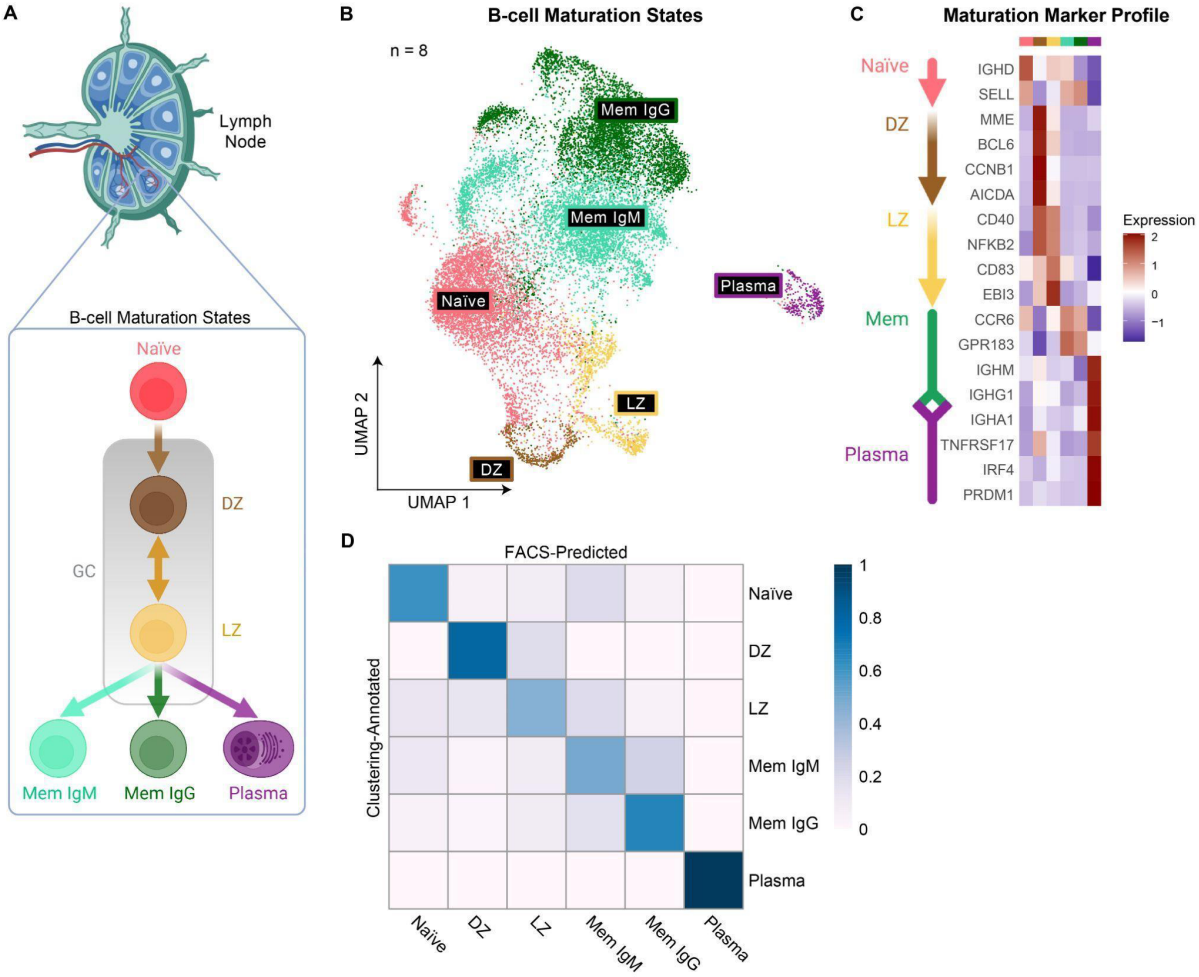
I generated single-cell multi-omic data for each of these samples with CITE-Seq[50], combining single-cell RNA sequencing with surface epitope profiling using oligonucleotide-tagged antibodies for 70 immunological markers (Table 2), in collaboration with Tobias Roider and Mareike Knoll. The alignment of transcripts was carried out against the hg38 reference genome. Post quality control, we successfully acquired data from 154,282 B cells, with a sample median of 2,988 B cells [140-7,868] and a median count of 6,887 transcripts and 2,532 surface protein counts per cell. For 8 of these samples (1 rLN, 1 MCL, 2 FL, 2 GCB DLBCL, and 2 MZL), we also performed joint single-cell RNA sequencing and immune receptor profiling to study clonality.

A single-cell reference map of B-cell maturation states in reactive lymph nodes

Before studying ITH in B-NHL samples, I first sought to profile the B-cell states that arise across the B-cell maturation differentiation lineage (B-cell maturation states) in the non-malignant lymph node context. I performed clustering on the rLN CITE-Seq data (8 samples, 16,625 B cells), and annotated B-cell clusters to known maturation states based on a broad range of maturation state markers from the literature (Table 5, Figure 3B-C)[34,44,49,52–56]. 16 clusters were mapped to 6 distinct B-cell maturation states found at different frequencies

across samples, averaging: 30% pre-germinal center naïve B cells (Naïve), 3% centroblasts from the dark zone of the germinal center (DZ), 6% centrocytes from the light zone of the germinal center (LZ), 30% IgD+/IgM+ memory B cells (Mem IgM), 30% IgG+ or IgA+ class-switched memory B cells (Mem IgG), and 2% plasma cells (Plasma).

Figure 3: A single-cell B-cell maturation reference map in reactive lymph nodes



(A) Schematic of the B-cell maturation trajectory in the lymph node. The labeled populations represent the identified B-cell maturation states. Illustrations were created with BioRender.com[47].

(B) Transcriptomic UMAP of the rLN reference CITE-Seq dataset (8 samples) labeled by the B-cell maturation states in (A). Transcriptomic clusters were assigned to maturation states based on their expression of the maturation markers in Table 5.

(C) Heatmap showing the z-scored average expression of a subset of markers for each maturation state annotated in the reference CITE-Seq dataset.

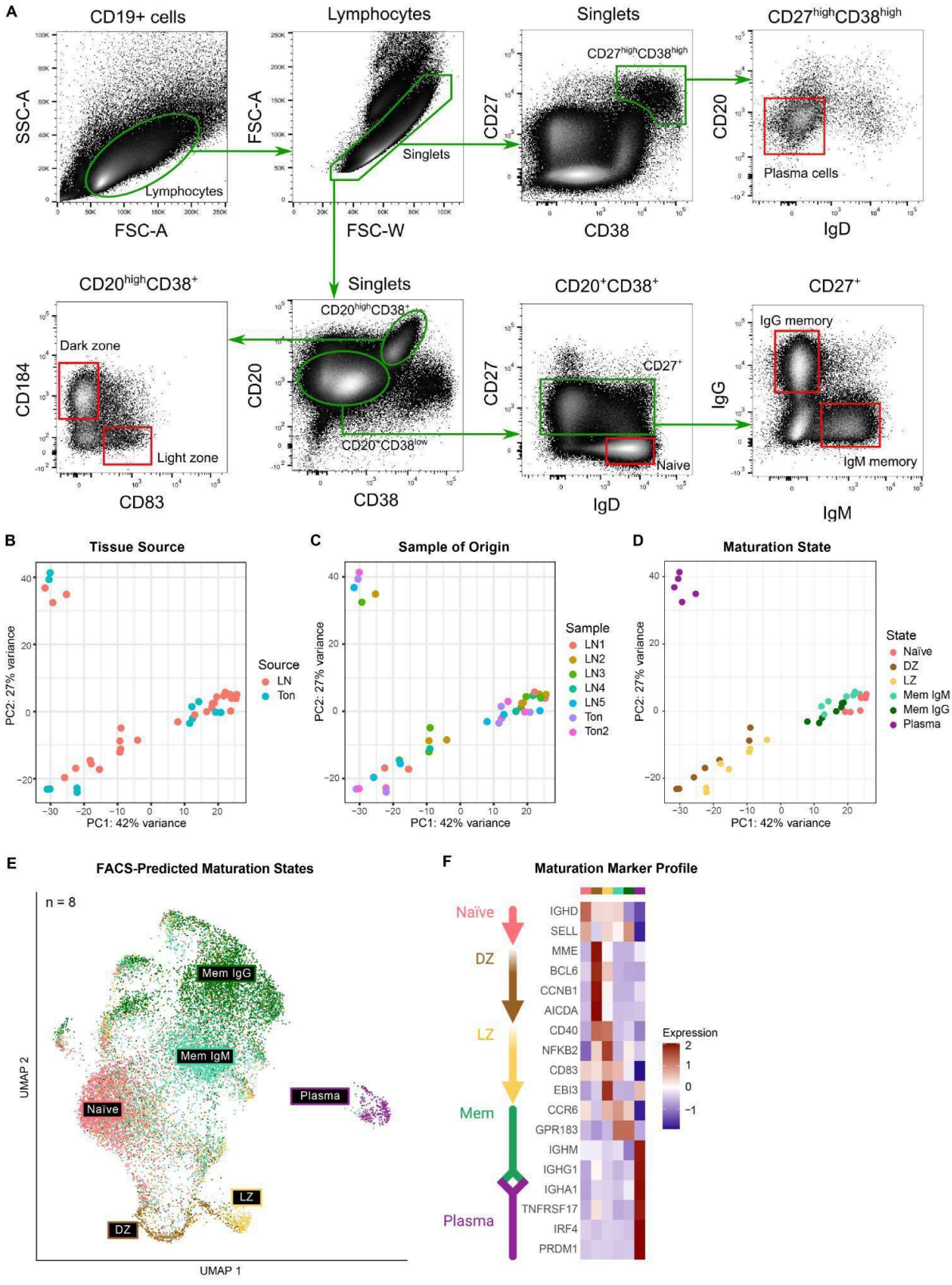
(D) Confusion matrix of cells' maturation state labels annotated by maturation marker profiling of transcriptomic clusters (y-axis) and predicted by a logistic regression[57] classifier trained on RNA-sequencing

data from B-cell maturation states sorted with FACS (x-axis) (Figure 4B-F). The color scale shows probability estimates for each class.

Maturation state annotations: Naïve = Naïve B cells, DZ = Centroblasts from the dark zone of the germinal center, LZ = Centrocytes from the light zone of the germinal center, Mem IgM = IgD+ and IgM+ memory B cells, Mem IgG = class-switched (IgG+ or IgA+) memory B cells, Plasma = plasma cells. From Fitzgerald et al. 2023[1], originally produced by myself.

In parallel, Artur Kibler profiled and sorted B-cell maturation states from 5 rLN samples with flow-activated cell sorting (FACS) using a maturation marker antibody panel (Table 3) and the gating strategy outlined in Figure 4A. I analyzed RNA sequencing data from the sorted B-cell maturation states together with Bettina Budeus to determine their gene expression profiles (Figure 4B-D). I trained a logistic regression classifier to predict maturation states based on gene expression, achieving a balanced accuracy of 94% in cross-validation. I applied this classifier to the single-cell gene expression profiles of B cells in the CITE-Seq dataset (Figure 4E-F). For each annotated maturation state, the majority of cells conformed to the maturation state predicted by the logistic regression model (Figure 3D).

Figure 4: FACS-based classification of B-cell maturation states in reactive lymph nodes



(A) B-cell maturation state gating strategy for flow-activated cell sorting (FACS) employed on B cells isolated from human rLN (n=5) and tonsils (n=2) using the following marker panels: Naïve B cells (CD19+, CD20+,

CD38low, CD27-, IgD high), germinal center dark zone B cells (CD19+, CD20+, CD38+, CD184+, CD83-), germinal center light zone B cells (CD19+, CD20+, CD38+, CD184-, CD83+), IgM memory B cells (CD19+, CD20+, CD38 low, CD27+, IgM+), IgG memory B cells (CD19+, CD20+, CD38 low, CD27+, IgG+), and plasmablasts/plasma cells (CD19+, CD20 low, CD38 high, CD27 high, IgD low).

(B-D) The first two principal components of RNA-seq data from FACS-sorted B-cell maturation states from rLN (n = 5) and tonsils (n = 2) colored by (B) tissue source, (C) the sample of origin, and (D) maturation state.

(E) Transcriptomic UMAP of the integrated CITE-Seq B cells data from 8 rLN labeled by maturation state predicted with logistic regression from the sorted states' RNA-seq data.

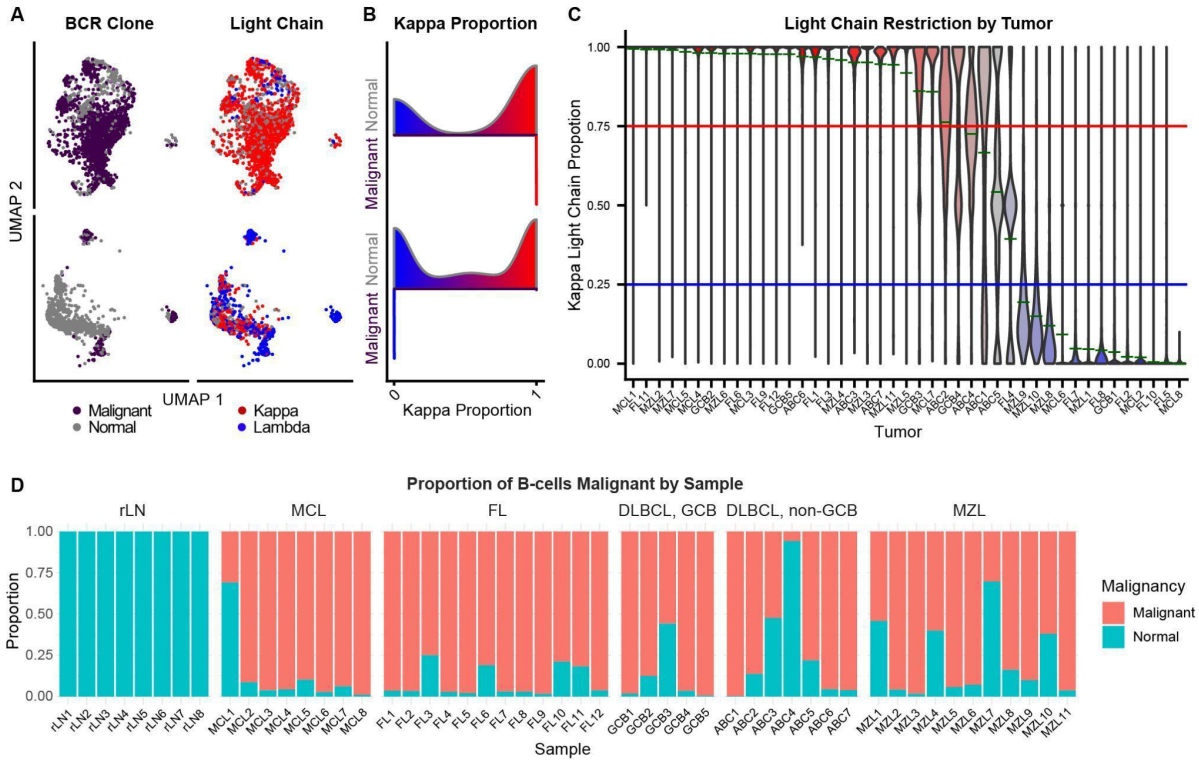
(F) Z-scaled gene expression of a subset of B-cell maturation markers (rows) across predicted maturation states (columns) in the rLN reference.

From Fitzgerald et al. 2023[1], originally produced by myself and Artur Kibler.

Mapping B-cell maturation states in tumors

Having mapped B-cell maturation states in healthy lymph nodes, I then extended this analysis to the 43 tumor samples in the CITE-Seq dataset (8 MCL, 12 FL, 5 GCB DLBCL, 7 non-GCB DLBCL, and 11 MZL). As cancer cells undergo clonal expansion, B-cell malignancies consist of an expanded population of monoclonal B cells restricted to either a kappa or lambda immunoglobulin light chain, a phenomenon known as light chain restriction. I could therefore identify malignant cells in each sample through light chain restriction analysis, whereby I deemed transcriptional clusters with either kappa or lambda light chain proportions >75% as malignant subpopulations. I validated this approach with joint single-cell RNA sequencing and B-cell receptor (BCR) profiling in 8 samples, confirming that each tumor sample harbored an expanded B-cell receptor clone with a restricted immunoglobulin light chain representing the malignant cells, as exemplified in Figure 5A-B. Malignant populations could be identified by light chain restriction in all but 2 tumor samples (ABC5 and FL4) in the CITE-Seq dataset, which showed light chain depletion instead (Figure 5C). Non-malignant B cells, predominantly naïve B cells, accounted for a median of 6% [0%-94%] of all B cells across tumor samples in the CITE-Seq dataset (Figure 5D).

Figure 5: Isolation of malignant B cells based on light chain restriction



(A) Reference-based UMAP labeled by malignant clone as determined by BCR profiling (left) and immunoglobulin light chain (right) for an MCL (top) and MZL (bottom) sample.

(B) Horizontal violin plot depicting the proportion of kappa light chain gene expression (x-axis) in malignant and normal B cells in the samples shown in (A). Red = kappa-positive, blue = lambda-positive.

(C) Vertical violin plot showing the proportion of kappa light chain surface epitope detected in malignant cells isolated from all tumor samples in the CITE-Seq cohort (n = 43). Red = kappa-restricted, blue = lambda-restricted. Malignant B cells were identified as light chain-restricted transcriptional clusters (mean kappa proportion >0.75 or <0.25). A light-chain-restricted tumor population was identified in all samples except ABC5 and FL4, which showed light-chain depletion instead.

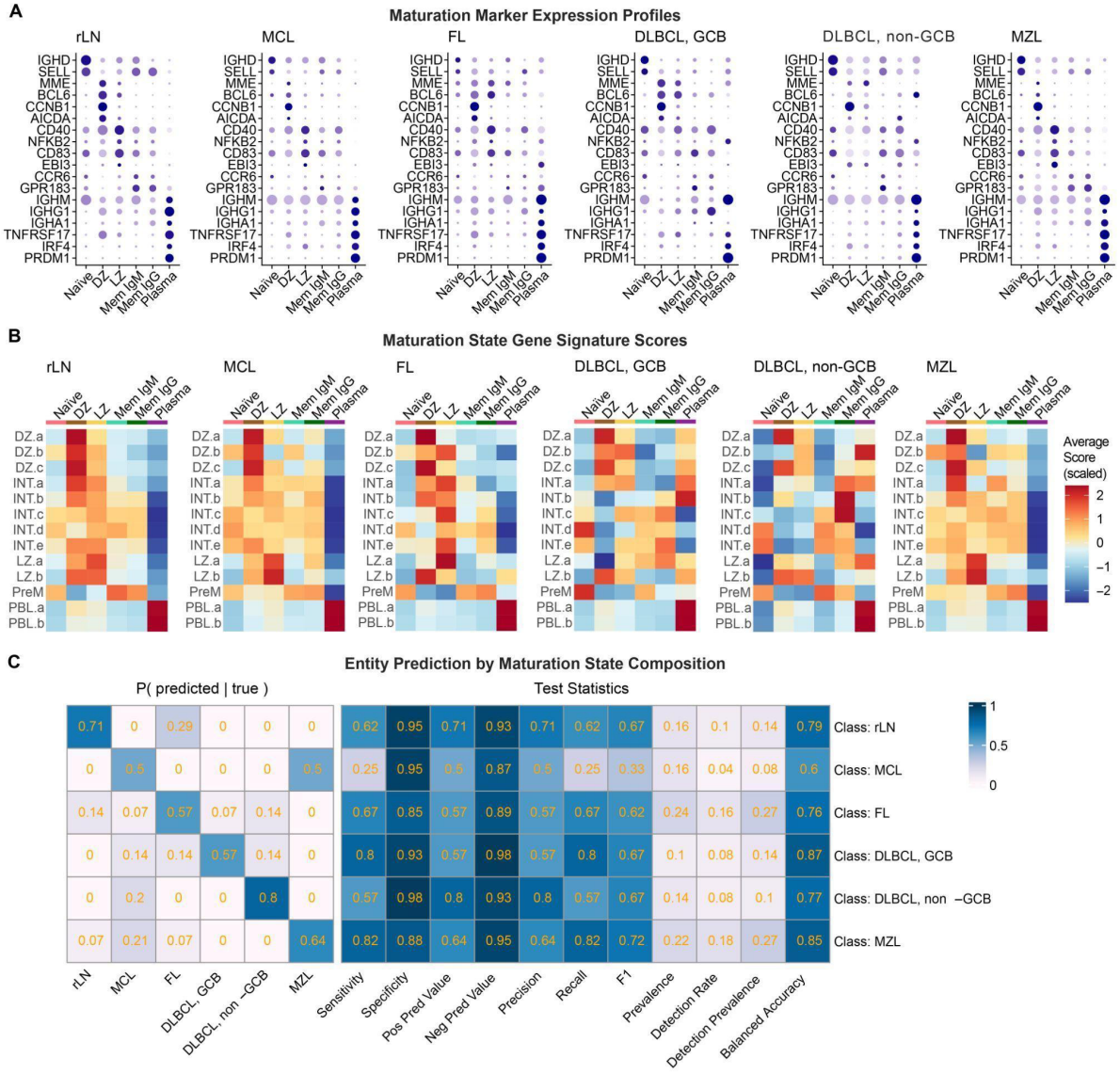
(D) The proportion of B cells that are malignant or non-malignant, based on light chain restriction, in each sample, faceted by entity: reactive lymph nodes (rLN), mantle cell lymphoma (MCL), follicular lymphoma (FL), germinal center and non-germinal center diffuse large B-cell lymphoma (DLBCL, GCB/non-GCB), and marginal zone lymphoma (MZL).

From Fitzgerald et al. 2023[1], originally produced by myself.

I used the annotated rLN samples to map B-cell maturation states in each tumor sample in our CITE-Seq dataset. To this end, I employed a data integration technique utilizing mutual nearest neighbors and canonical correlation analysis for label transfer[58], with the rLN samples as the reference dataset and each tumor sample as a query dataset. Label transfer was performed on the log-normalized transcriptomic counts for each tumor sample separately,

thereby avoiding the risk of introducing artefacts from sample integration or scaling in the transferred annotations. To validate these maturation states, I analyzed the maturation marker profiles of maturation states in each entity and compared them with those of rLN samples. Strong concordance was observed, with key maturation markers differentially expressed in the respective states of each entity as in the non-malignant rLN context (Figure 6A). For further validation, I computed maturation state-specific gene signature scores for each entity's maturation states using the top 50 differentially expressed genes from maturation states annotated in a previously published tonsil dataset[49]. As in the rLN context, higher gene signature scores were observed in their respective maturation states in each entity. Although these differences were better preserved between distant states such as DZ and plasma than between LZ and memory states which are closer in lineage (Figure 6B).

Figure 6: B-cell maturation marker expression and gene signature scores by entity



(A) Dot plots showing the relative gene expression and abundance of maturation markers for maturation states mapped from the reactive lymph node reference (Figure 1B) in each entity.

(B) Heatmap of maturation state scores calculated from annotated maturation states in a published tonsil germinal center scRNA-seq dataset[49]. Each score shown is scaled across all scores in the dataset (mean = 0, SD = 1).

(C) Confusion matrix (left) showing the predicted (x) vs true (y) classes when predicting entity by maturation state proportions with random forest (nested cross-validation), with test statistics (right) for classification of each entity (overall accuracy 63%).

Entities: reactive lymph nodes (rLN), mantle cell lymphoma (MCL), follicular lymphoma (FL), germinal center and non-germinal center diffuse large B-cell lymphoma (DLBCL, GCB/non-GCB), and marginal zone lymphoma. See Figure 3 for maturation state annotations. From Fitzgerald et al. 2023[1], originally produced by myself.

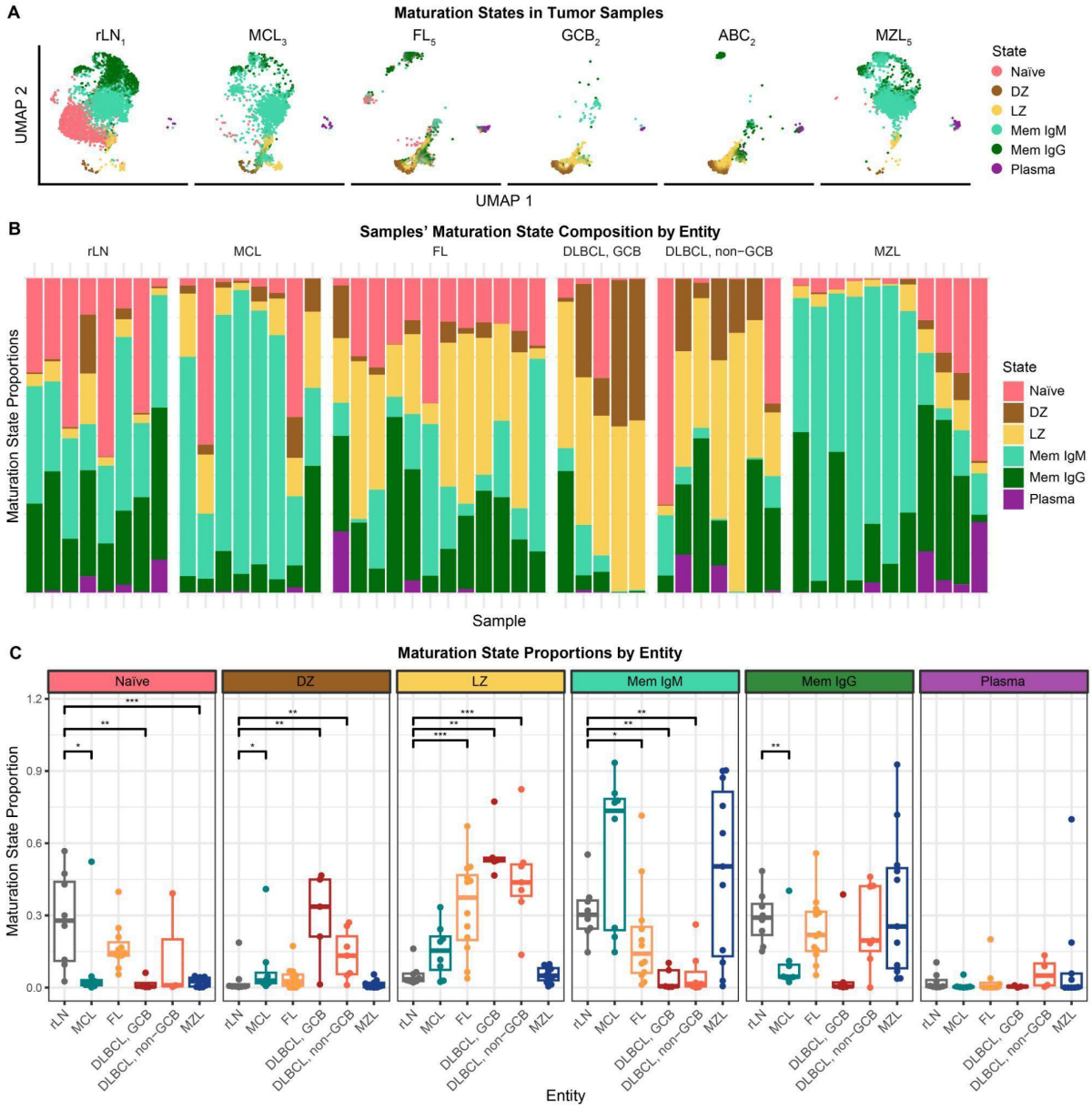
Inter- and intratumor heterogeneity in B-cell maturation states

To study B-cell maturation states among cancer cells, non-malignant B cells were removed for further analysis. The cell-of-origin model of *inter*-tumor heterogeneity postulates that each tumor reflects a single cell type that it originated from, such as naïve B cells in an MCL tumor or DZ cells in an FL tumor. In contrast, I observed a range of B-cell maturation states within individual tumors (Figure 7A-B). These intratumor maturation states were of the same BCR clone, therefore diverging from the same cell of origin. This finding suggests that differentiation in cancer is not static but rather exhibits ongoing plasticity and divergence.

We detected distinct tumor maturation state fingerprints within each entity that predominantly mirrored their cell-of-origin. For example, GC (DZ and LZ) states were prominent in DLBCL GCB and memory states in MZL. Although these states coexisted with others either preceding or succeeding them in the maturation lineage. Notably, there was significant variation in the proportion of these maturation states across different tumors of the same entity (Figure 7B). This *inter*-tumor heterogeneity in maturation state composition blurred entity boundaries. Predicting entity based on maturation state composition using logistic regression achieved a peak accuracy of 63% in nested cross-validation (Figure 6C). These findings highlight cell-type differentiation lineages as a source of variation both across and within tumors.

In both FL and non-GCB DLBCL tumors, GC states (DZ and LZ) co-existed with post-GC states (memory and plasma cells). This indicates a potential for FL, like DLBCL, to transform into post-GC phenotypes. Comparative analysis revealed a significant enrichment of the LZ state in DLBCL and FL compared to reactive lymph node (rLN) controls ($p < 0.01$, Wilcoxon signed-rank test). However, FL did not show similar enrichment for the DZ state (Figure 7C). This suggests that DLBCL is more likely to retain a DZ phenotype, even in non-GCB DLBCL. The ongoing processes of somatic hypermutation and proliferation within the DZ state could contribute to the aggressive nature of DLBCL's disease progression.

Figure 7: Intra- and intertumor heterogeneity in B-cell maturation states



(A) Transcriptomic UMAPs for individual samples from each entity, labeled by B-cell maturation states assigned by label transfer from the reactive lymph node reference (Figure 3B). Only malignant cells, identified based on light-chain restriction as in Figure 5, are shown for tumor samples.

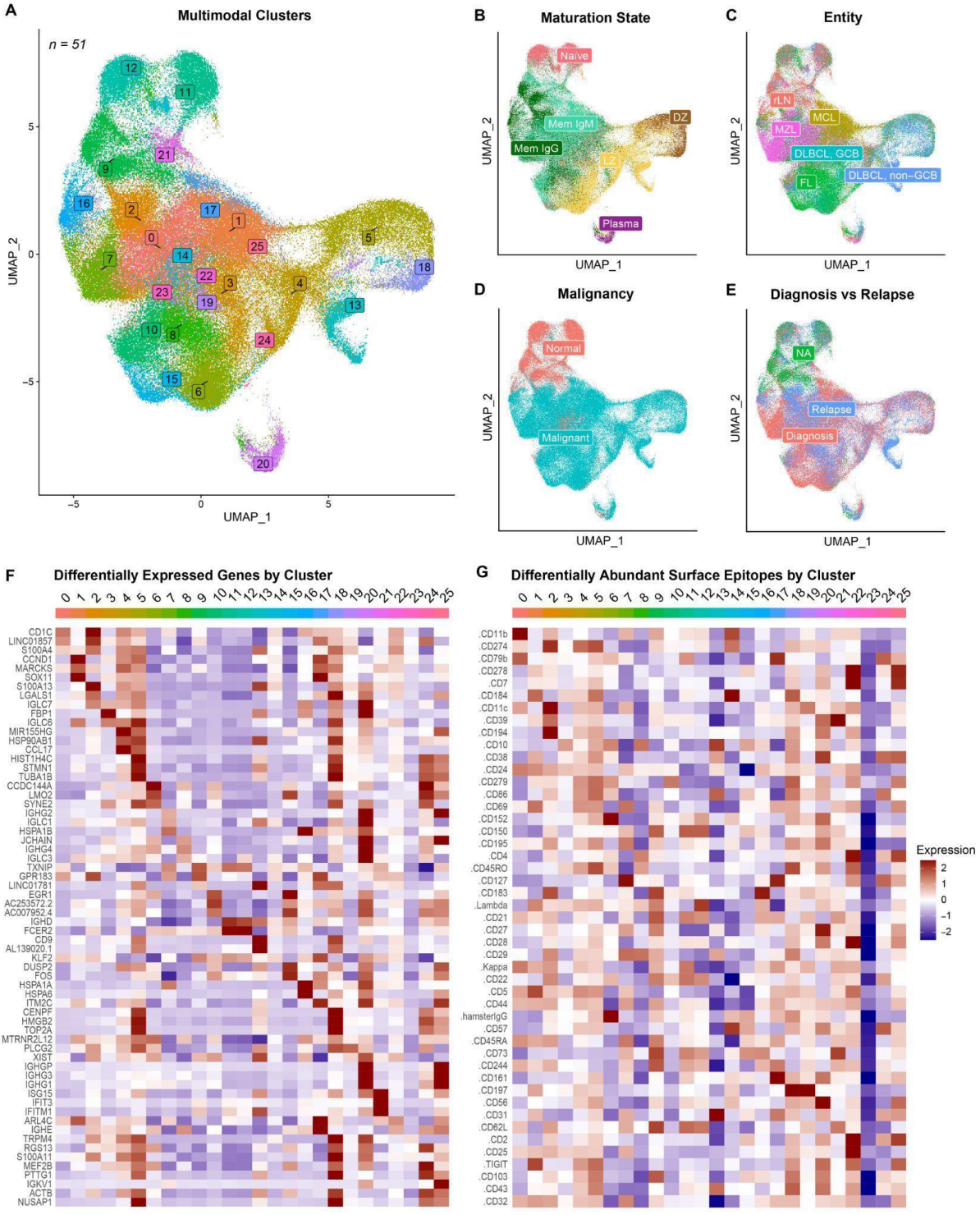
(B) Maturation state composition of all samples (n=51) split by entity and ordered by days since diagnosis.

(C) Box plot of the proportion of each maturation state in each entity. Each data point is a sample, grouped by entity. The Wilcoxon signed-rank test[59] (two-sided) was performed for each maturation state's proportion between reactive lymph nodes and each entity: $p<0.05$ (*), $p<0.01$ (**), $p<0.001$ (***). Centerline = median, box limits = upper and lower quartiles, whiskers = 1.5x interquartile range.

rLN = reactive lymph nodes, MCL = mantle cell lymphoma, FL = follicular lymphoma, DLBCL = diffuse large B-cell lymphoma (GCB = germinal center, non-GCB/ABC = activated B cell), MZL = marginal zone lymphoma. See Figure 1 for maturation state annotations. From Fitzgerald et al. 2023[1], originally produced by myself.

To investigate the extent to which the B-cell maturation differentiation lineage explains phenotypic diversity among B-NHL tumor cells, I employed an unsupervised technique for multimodal subpopulation mapping across our full CITE-Seq B-cell dataset (51 samples with 154,282 cells). Using Multi-Omic Factor Analysis (MOFA)[60,61], I generated a joint low-dimensional embedding and clustering based on the co-variance of transcriptomic and surface protein features across cells (Figure 8A). This analysis revealed 25 distinct multimodal subpopulations, largely distinguished by their maturation states, highlighting the significant role of maturation in driving tumor variation (Figure 8B-E). Furthermore, within each maturation state, I identified multiple unique subpopulations. Notably, the memory B-cell states exhibited the most pronounced heterogeneity, with distinct subpopulations for each B-NHL entity. Key features distinguishing these subpopulations included entity-associated pathogenic mechanisms like CCND1 overexpression in MCL[62] and CD152 (CTLA-4) overexpression in FL[63](Figure 8F-G). This sheds light on the dual axes of differentiation and oncogenesic mechanisms in tumor variation.

Figure 8: Multimodal subpopulation mapping of B-NHL



(A-E) UMAP visualization of the full CITE-Seq B-cells dataset (n=51) constructed with the latent factors (n = 50) from multi-omic factor analysis (MOFA)[60] based integration of the single-cell RNA and ADT (surface markers) data layers as principle components, labeled by clustering on the multimodal latent factor space (A), maturation states (B) mapped from the reactive lymph node reference in Figure 1B, entity (C), malignancy (D) as determined by light chain restriction (Figure 5C), and samples taken at diagnosis or relapse (E).

(F-G) Z-scaled expression across multi-modal clusters of the 3 most differentially expressed genes (F) and proteins (G) by fold-change per cluster.

From Fitzgerald et al. 2023[1], originally produced by myself.

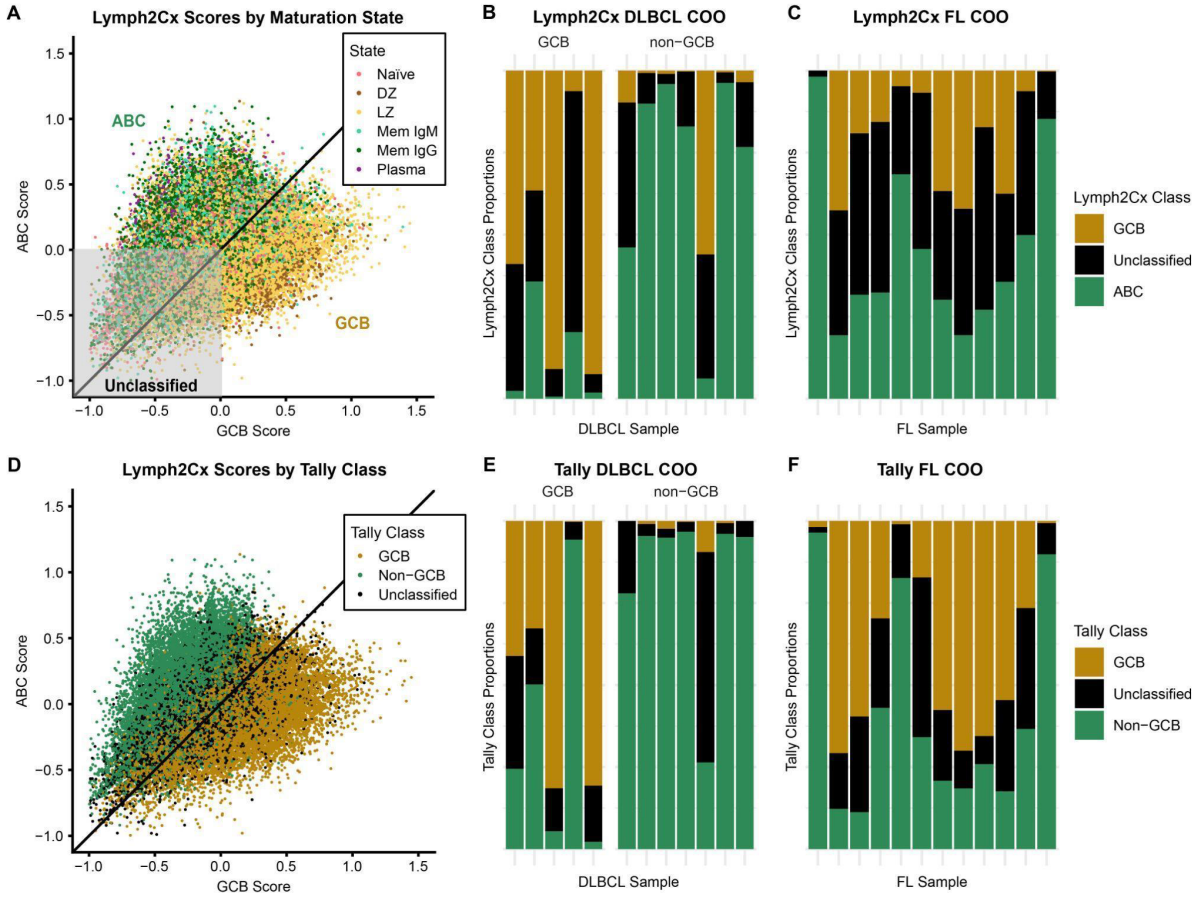
Distinct cell-of-origin subtypes within tumors

Subsequently, I explored how ITH in maturation states affects cell-of-origin (COO) classifiers for B-NHL diagnosis. These classifiers are used to categorize DLBCL into GCB and ABC subtypes according to a GC or post-GC COO, respectively. This however does not account for the presence of multiple maturation states in a tumor. To test this, I applied the Lymph2Cx DLBCL COO classifier, which is based on the expression of 20 genes54 linked to GC and post-GC maturation states[64], to single-cell RNA sequencing data from DLBCL tumor cells. GCB and ABC gene expression signature scores corresponded with germinal center (DZ and LZ) and post-germinal center (Mem and Plasma) states, respectively (Figure 9A). 20-30% of the tumor cells did not predominantly express GCB or ABC signatures and were thus unclassified. Contrary to a single COO subtype per tumor, many tumors displayed a mix of COO subtypes (GCB, unclassified, and ABC), with the prevalent subtype often aligning with the GCB or non-GCB diagnosis (Figure 9B).

This ITH in COO subtypes was verified with the Tally DLBCL COO classifier, a simpler method typically employing immunohistochemistry to determines tumor subtype based on the presence or absence of proteins linked to GCB (CD10, GCET1, LMO2) or non-GCB (MUM1, FOXP1) states[65]. The Tally classifier's results mirrored those obtained with the Lymph2Cx classifier when applied to the scRNA-seq data from DLBCL samples (Figure 9D-E). With both classifiers, GCB DLBCL frequently harbored ABC-like cell populations, indicating the differentiation of some GCB-originating tumor cells into the ABC subtype. The existence of multiple COO subtypes within tumors suggests a more complex pathology of DLBCL than previously understood.

I previously noted that FL exhibits characteristics of both germinal center (GC) and post-germinal center states, despite the absence of an established ABC subtype as seen in DLBCL. Thus, I explored the possibility of detecting both GCB and ABC COO subtypes within FL tumors at the single-cell level. Utilizing the Lymph2Cx and Tally classifiers as previously in DLBCL tumors, I identified the presence of both GCB and ABC classes in 10 of the 12 FL tumors analyzed (Figure 9C+F). This finding indicates the emergence of an additional ABC-like subtype within FL tumors.

Figure 9: Distinct cell-of-origin subtypes coexist within tumors



(A) Normalized GCB and ABC scores for each DLBCL and FL tumor cell determined by the Lymph2Cx cell-of-origin classifier, labeled by B-cell maturation state. The diagonal line divides cells by their classification (above = ABC, below = GCB). Cells without ABC or GCB gene set average expression above the housekeeping genes were unclassified (grey box).

(B-C) Lymph2Cx class proportions among the tumor cells of DLBCL samples faceted by GCB or non-GCB diagnosis (B) and FL samples (C).

(D) Normalized GCB and ABC scores for each DLBCL and FL tumor cell as determined by the Lymph2Cx classifier, as shown in (A), labeled by Tally class.

(E-F) Tally class proportions among the tumor cells of DLBCL samples faceted by GCB or non-GCB diagnosis (E) and FL samples (F).

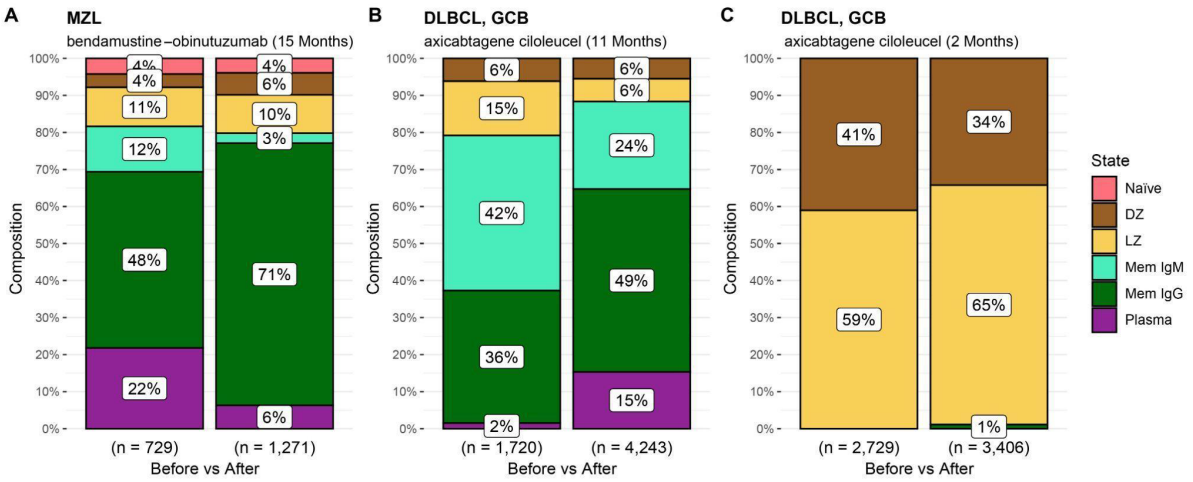
FL = follicular lymphoma, DLBCL = diffuse large B-cell lymphoma (GCB = germinal center, non-GCB = non-germinal center), COO = cell-of-origin. See Figure 3 for maturation state annotations. From Fitzgerald et al. 2023[1], originally produced by myself.

Longitudinal variation in maturation state composition

Discovering ITH in B-cell maturation states, indicating ongoing differentiation within tumors, led me to hypothesize that differentiation influences tumor evolution. A flux in the

composition of maturation states over time could result from ongoing differentiation and/or the selective pressures of treatment. I investigated this through analysing longitudinal tumor samples from three B-NHL patients. In an MZL patient who achieved complete remission after six rounds of obinutuzumab and bendamustine but experienced a relapse after 15 months, I observed a decrease in plasma cells and a surge in the Mem IgG state (Figure 10A). Similarly, a GCB DLBCL patient who relapsed 11 months following CAR-T cell therapy (axicabtagene ciloleucel) showed enrichment of plasma cells (Figure 10B). Shifting maturation state composition implies tumors can evolve through differentiation. These effects seem largely time-dependent, as minimal alteration in the maturation state composition was observed in another GCB DLBCL patient who relapsed just two months after the same CAR-T cell therapy (Figure 10C).

Figure 10: Longitudinal patterns in tumor maturation state composition



(A-C) B-cell maturation state composition of tumor cells in longitudinal samples from 3 patients: an MZL patient who relapsed 15 months following complete response to 6 cycles of obinutuzumab-bendamustine chemo-immunotherapy (A) and two GCB DLBCL patients who relapsed after 11months (B) and 2 months (C) following axicabtagene ciloleucel (CAR-T cell) therapy, respectively. Maturation states were mapped from the reactive lymph node reference dataset in Figure 3B. From Fitzgerald et al. 2023[1], originally produced by myself.

Maturation-associated transcription factor activity in malignant states

Having established differentiation as a driver of inter- and intra-tumor heterogeneity, I next examined the mechanisms by which differentiation continues in cancers. Differentiation is

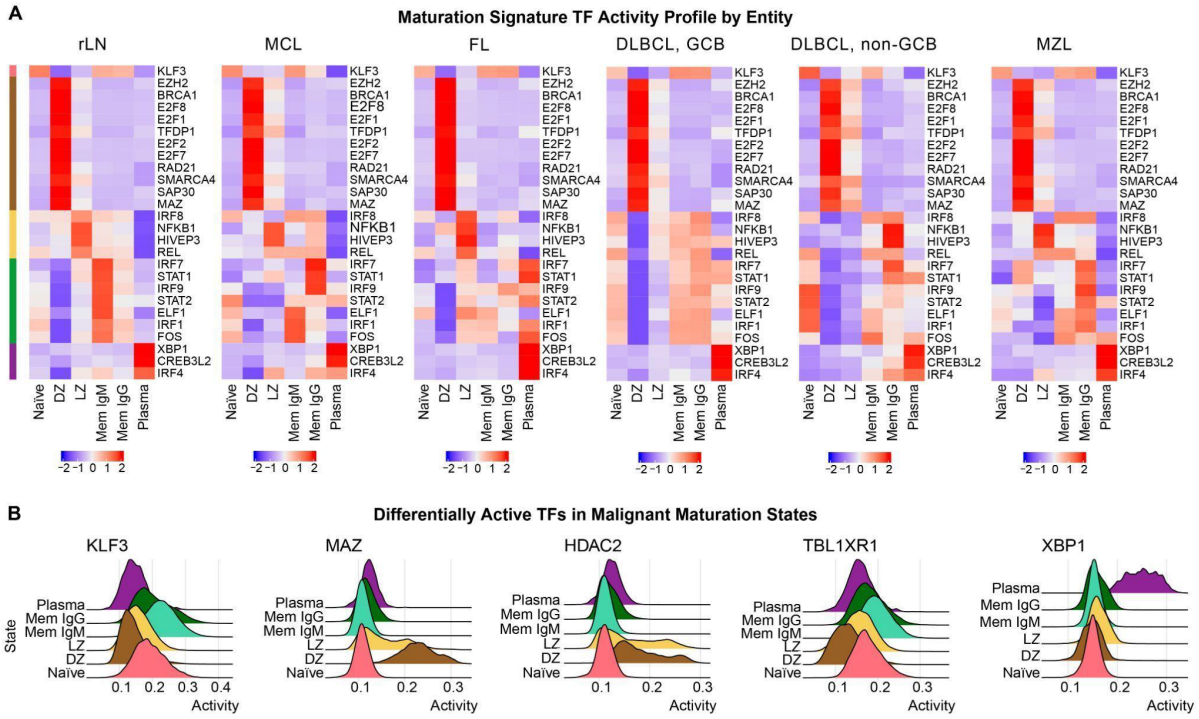
orchestrated by a broad range of connected intrinsic and extrinsic processes, including gene regulation, epigenetic modifications, signal transduction pathways, and cell-cell interactions[66].

Transcription factors (TFs) play an overarching regulatory role in cell-type-differentiation, including in early B-cell development and subsequent B-cell maturation[67]. For example, Pax5, EBF1, and Ikaros are essential for early B-cell lineage commitment and the maintenance of B-cell identity[68], while Blimp-1 (PRDM1) and Bcl-6 play crucial roles in terminal differentiation processes, such as the formation of plasma cells and germinal center B cells, respectively[69,70]. This intricate regulation by TFs not only ensures the effective functioning of the adaptive immune response but also maintains a delicate balance to prevent aberrant B-cell proliferation and the potential for lymphomagenesis. Furthermore, aberrant gene regulatory networks are considered part of the pathogenesis of B-NHL[45,71] with frequent mutations observed in key TFs, including PRDM1 and Bcl-6 [62].

Given the observed ITH in B-cell maturation states, I aimed to determine the extent to which differential TF activities across maturation states persist in cancer despite the epigenomic dysregulation associated with B-NHL. This would shed light on whether TFs' key role in differentiation is preserved in cancer. Anna Mathioudaki computationally inferred TF activities from single-cell RNA sequencing data in the CITE-Seq dataset (51 samples) through the SCENIC workflow[72]. This method identifies co-expression modules of TFs and potential target genes with a common binding motif and then assesses TF activity based on the expression of these modules in each cell. We identified signature maturation TFs as those with target genes differential expressed between maturation states in the reactive lymph node (rLN) reference dataset (Benjamini-Hochberg[73] adjusted p-value < 10e-16, average $\log_2$ fold-change > 0.4). We then evaluated the relative activities of these TFs across different states for each sample. TFs showing differential activity between maturation states in rLN also exhibited similar patterns of activity across states in malignant samples (Figure 11A), indicating that TF signatures associated with maturation states were largely conserved in B-NHL. However, this conservation was partially disrupted in the LZ state in both GCB and non-GCB DLBCL, whereby TFs that were more active in the rLN LZ (e.g., IRF8, NFKB1, HIVEP3, REL) demonstrated increased activity in the DLBCL memory state. We observed considerable variation in the activity of memory signature TFs across different samples, such as IRF7, STAT1, and IRF9. This variability aligns with the diverse phenotypes of memory B cells (Figure 8), which have completed maturation and thus may be less reliant on maturation

gene regulatory networks (GRNs).

Figure 11: TF activity signatures of maturation states are maintained in malignancy



(A) Scaled activity of transcription factors (TFs) across B-cell maturation states in each entity inferred using the *scenic* python package[72]; only TFs significantly enriched (log2 fold-change >0.4, p < 10e-16) for B-cell maturation stages target genes in reactive lymph node samples are shown. TFs (y-axis) are ordered by the maturation state in reactive lymph nodes in which they are enriched. Non-malignant cells in tumor samples were excluded based on light-chain restriction.

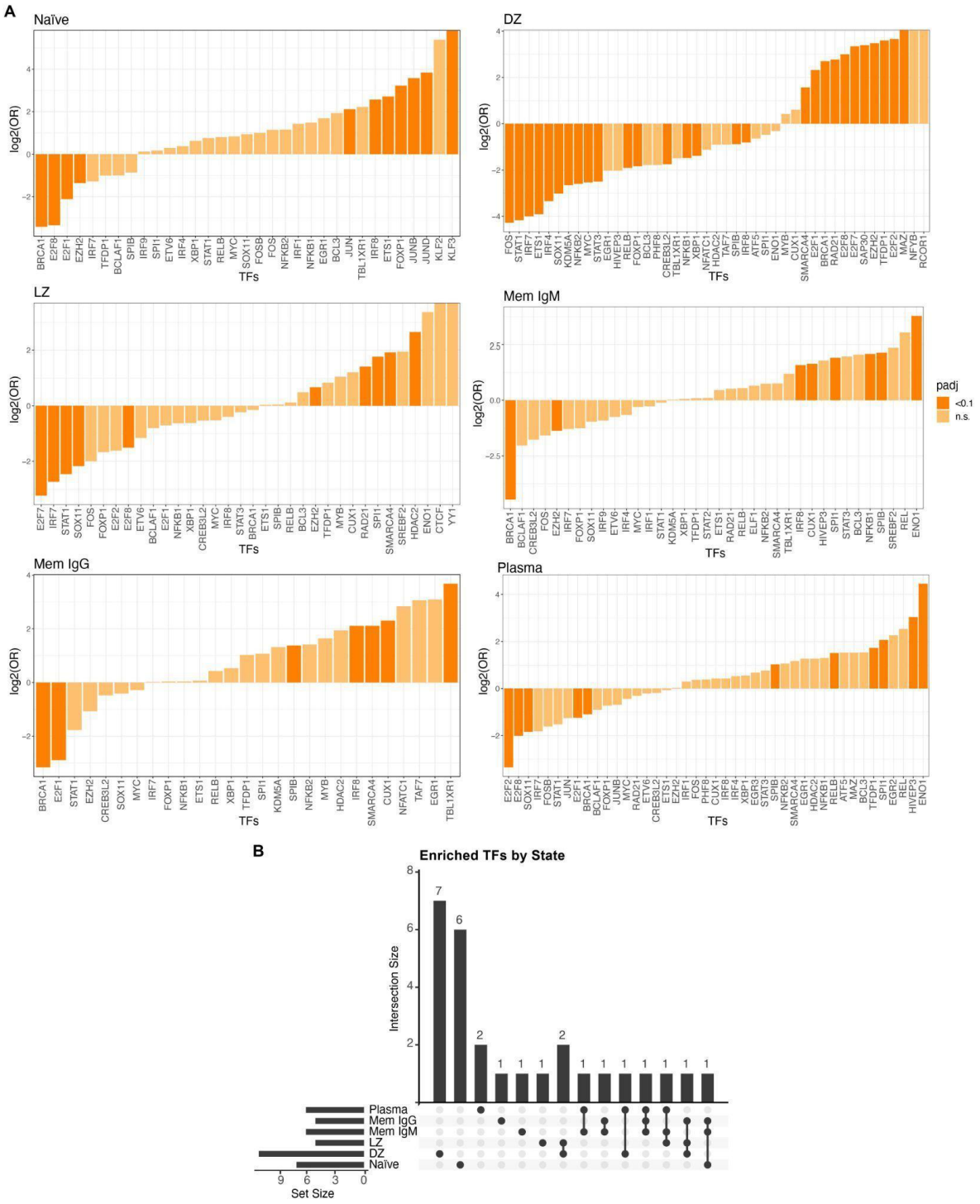
(B) Density plots comparing the activity distribution of selected differentially active transcription factors for each maturation state in malignant cells aggregated from all tumor samples (n = 43). See all differentially active transcription factors in malignant maturation states in Figure 12.

rLN = reactive lymph node, MCL = mantle cell lymphoma, FL = follicular lymphoma, DLBCL = diffuse large B-cell lymphoma (GCB = germinal center, non-GCB = non-germinal center), MZL = marginal zone lymphoma. See Figure 3 for maturation state annotations. From Fitzgerald et al. 2023[1], originally produced by myself and Anna Mathioudaki.

In addition to studying TFs that are associated with maturation states in the non-malignant context, I also analyzed the most differentially active TFs between tumor maturation states in each entity (Figure 12). TFs exhibiting the most significant variation in activity across maturation states were also associated with B-cell development, activation, and differentiation processes. Key examples are KLF3, which is active in naïve and memory B cells, MAZ and HDAC2 in germinal center (GC) B cells, TBL1XR1 in memory B cells, and

XBP1 in plasma cells (Figure 11B). These TFs may play a particularly important role in the initiation or sustenance of the diversity observed in intratumor maturation states.

Figure 12: Differential transcription factor activity between tumor maturation states



(A) Bar charts showing the log2 odds ratio for differentially active transcription factors (TFs) in the tumor cells of each B-cell maturation state. Transcription factor activity was inferred with the *SCENIC* python package[72] using single-cell RNA-sequencing data from the malignant cells of all tumor samples combined (n=43). Only

TFs with differentially expressed target genes (log2 fold-change >0.4, p < 10e-16 as determined with the *MAST* R package[74]) are shown. Bars highlighted in darker orange represent transcription factors with significant differential activity (FDR < 0.1 threshold).

(B) UpSet plot showing the intersections between the differentially active transcription factors (FDR <0.1) for each maturation state.

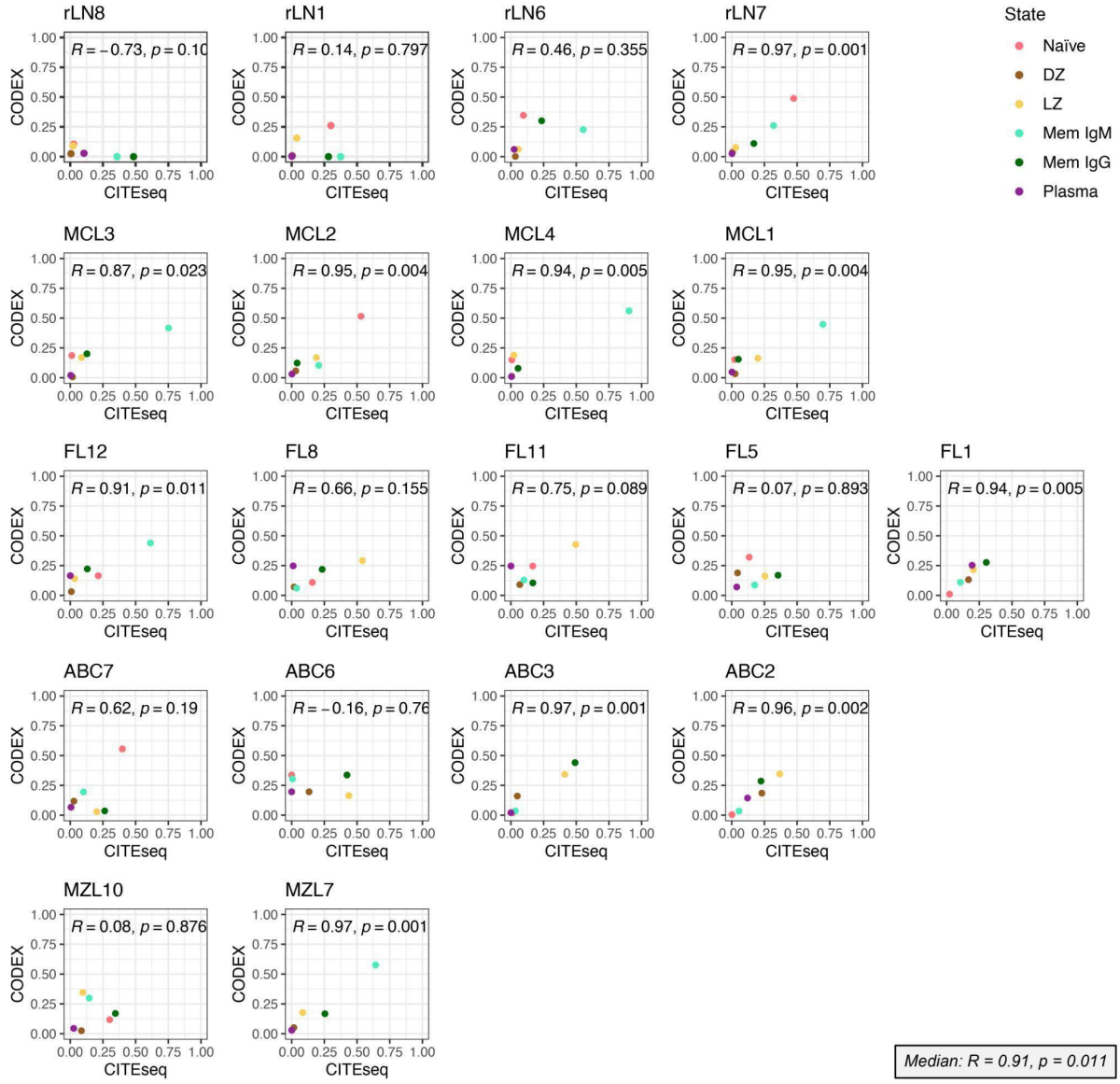
See Figure 3 for maturation state annotations. From Fitzgerald et al. 2023[1], originally produced by myself and Anna Mathioudaki.

Spatial mapping of B-cell maturation states via data integration

The follicular architecture of lymph nodes, with germinal centers (GC) that facilitate the GC reaction[44], is central to the B-cell maturation process. However, these B-cell follicles are often disrupted in malignancy[75]. Given the ITH in B-cell maturation states I found in the CITE-Seq dataset, I aimed to explore the spatial arrangement of maturation states within tumors and determine if malignancies maintain spatial niches that could support ongoing maturation. I analyzed the spatial arrangement of cell types in 18 samples (4 rLN, 4 MCL, 5 FL, 3 non-GCB DLBCL, 2 MZL) from our CITE-Seq cohort using CODEX[51] highly multiplexed immunochemistry data profiling 52 immunological markers (Table 4) generated by Marc-Andrea Bärtsch.

To transfer B-cell maturation state labels from the CITE-Seq to the CODEX dataset, I developed a data integration technique leveraging the shared protein features (n=28) between the two datasets (Tables 2 and 4). For each sample, after log-normalization and z-scaling of the shared protein features in each dataset, I trained a logistic regression classifier for B-cell maturation states on the CITE-Seq features, which I applied to the corresponding features in the CODEX B-cells data. Logistic regression outperformed tree-based (random forest, gradient boosted decision trees) and neural network-based classification methods. Despite employing random sampling for class balancing, I observed a strong correlation in B-cell maturation state proportions between the CITE-Seq and CODEX data across samples (median R=0.91, p = 0.011) (Figure 13).

Figure 13: Maturation state correlations between CITE-Seq and CODEX by sample



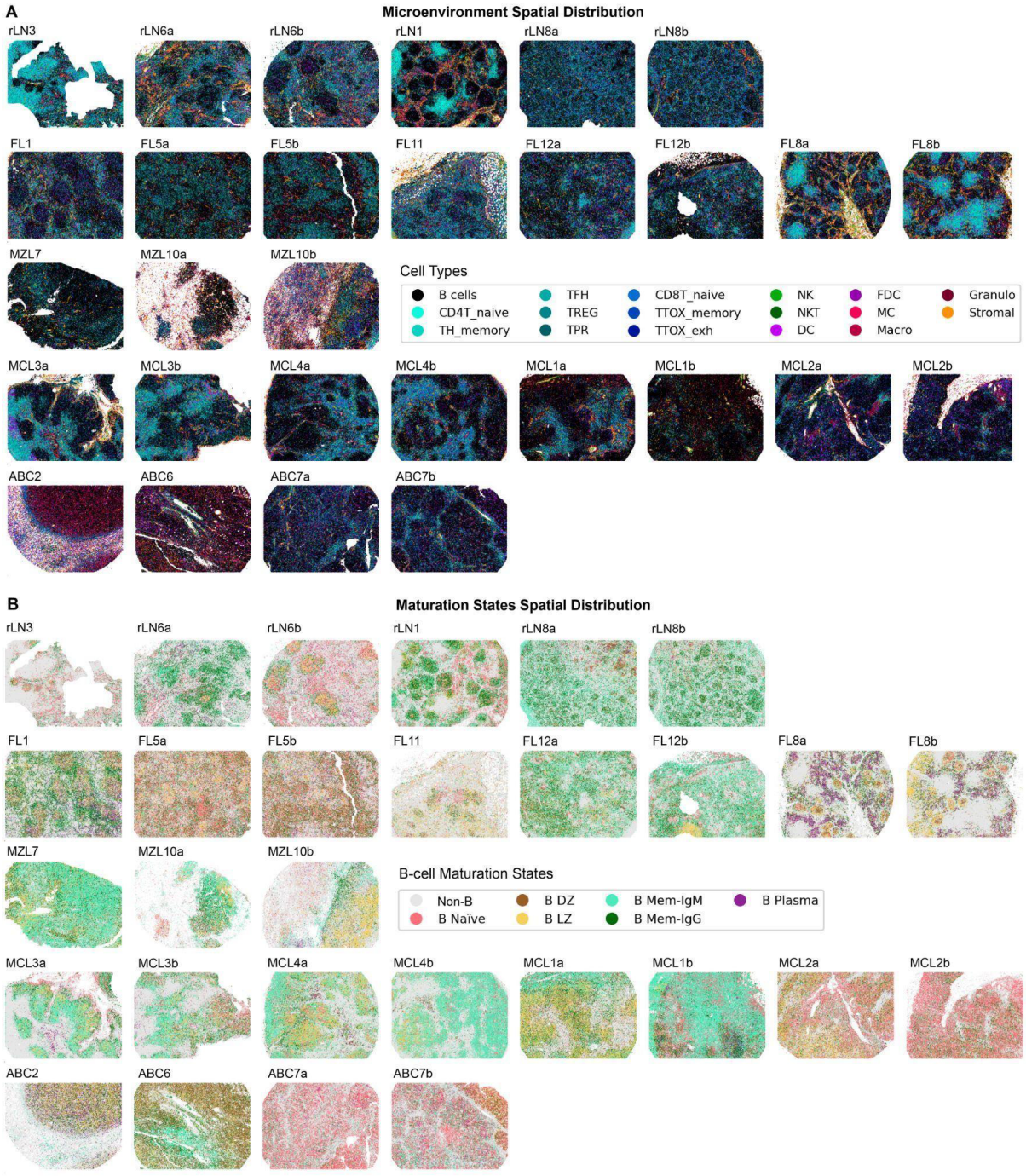
Scatter plots showing the Pearson correlation between B-cell maturation state proportions in the CITE-Seq and CODEX data for each reactive lymph node (rLN), mantle cell lymphoma (MCL), follicular lymphoma (FL), diffuse large B-cell lymphoma (DLBCL) and marginal zone lymphoma (MZL) sample. The median Pearson correlation coefficient across samples is shown (bottom-right). Maturation state labels were transferred from the CITE-Seq to CODEX data for each sample using logistic regression on the shared protein features (n = 28) in Tables 2 and 4. See Fig. 3 for maturation state annotations.

Intratumor maturation states occupy unique spatial niches

Marc-Andrea Bärtsch, Tobias Roider, Felix Czernilowsky, Harald Vöhringer, and I characterized immune and stromal cell types in the tumor microenvironment based on established nodal cell type marker profiles[76]. We observed variability in the tumor

microenvironment's composition and structure across tumors and entities, often marked by a disruption of the typical follicular structure in MCL and DLBCL. Heterogeneity was also observed in immune abundance and infiltration across samples. T-cell zones consisting of varying mixtures of helper, cytotoxic and regulatory T cells were generally more prominent in rLN, MCL, and FL samples than in MZL and DLBCL, although even within entities there was little consistent pattern (Figure 14A).

Figure 14: Spatial mapping of the tumor microenvironment and maturation states



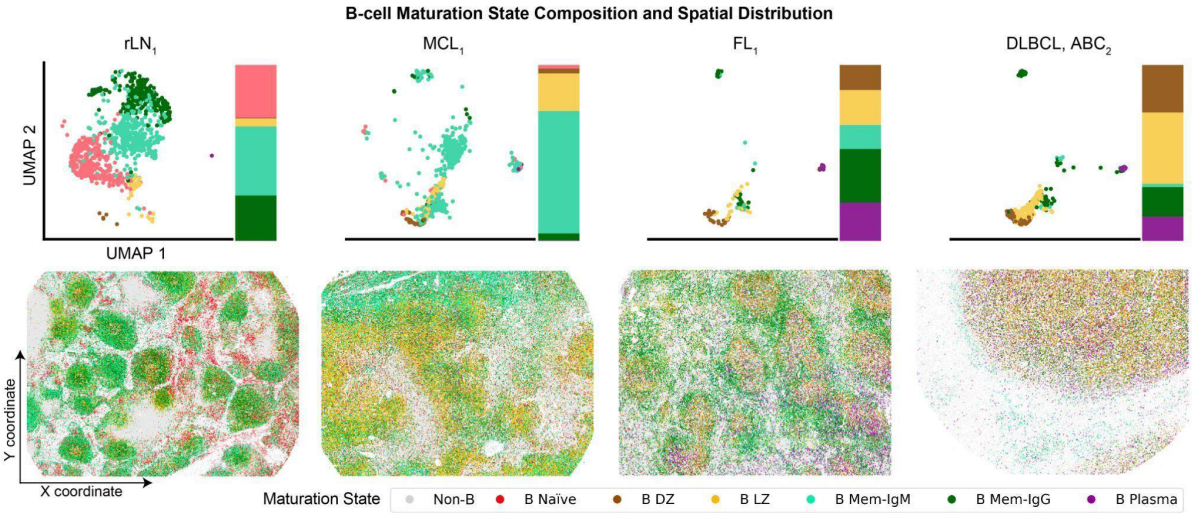
(A) Spatial distribution of cell types on CODEX images on FFPE sections of lymph node samples. Cell types were annotated by clustering on CODEX immunohistochemistry features (n = 52, Table 4). CD4T_naive = naive CD4+ T-cells, CD8T_naive = naive CD8+ T-cells, TH_memory = memory helper T-cells, TTOX_memory = memory cytotoxic T-cells, TTOX_exh = exhausted cytotoxic T-cells, NKT = natural killer T-cells, TFH = follicular helper T-cells, TPR = proliferating T-cells, TREG = regulatory T-cells, FDC = follicular dendritic cells, DC = dendritic cells, Macro = macrophages, Stromal = stromal cells, NK = natural killer cells, MC = monocytes, Granulo = granulocytes. All B cells are labeled black.

(B) Spatial distribution of B-cell maturation states on CODEX images on FFPE sections of lymph node samples. Maturation states were classified with logistic regression from the CITE-Seq maturation state annotations in Figure 7. B Naïve = naïve B cells, B DZ = centroblasts from the dark zone of the germinal center, B LZ = centrocytes from the light zone of the germinal center, B Mem IgM = IgD+ and IgM+ memory B cells, B Mem IgG = class-switched (IgG+ or IgA+) memory B cells, B Plasma = plasma cells,

rLN = reactive lymph node, MCL = mantle cell lymphoma, FL = follicular lymphoma, DLBCL = diffuse large B-cell lymphoma (GCB = germinal center, non-GCB = non-germinal center), MZL = marginal zone lymphoma. From Fitzgerald et al. 2023[1], originally produced by myself and Anastasiia Horlova.

I also observed substantial variation in the spatial distribution of intratumor maturation states among tumors and entities (Figure 14B). For example, FL1 was characterized by enlarged follicles with distinct DZ and LZ states in the germinal center, encircled by memory and plasma states, reflecting the preservation of the follicular structure among tumor cells. MCL1 lacked the usual follicular architecture but still displayed spatial separation between LZ and memory states. In contrast, ABC2 demonstrated a mix of several maturation states with minimal spatial separation, aligning with the diffuse nature of DLBCL (Figure 15). These observations highlight the diverse degrees and patterns of spatial segregation of intratumor maturations among different tumors and entities in B-NHL.

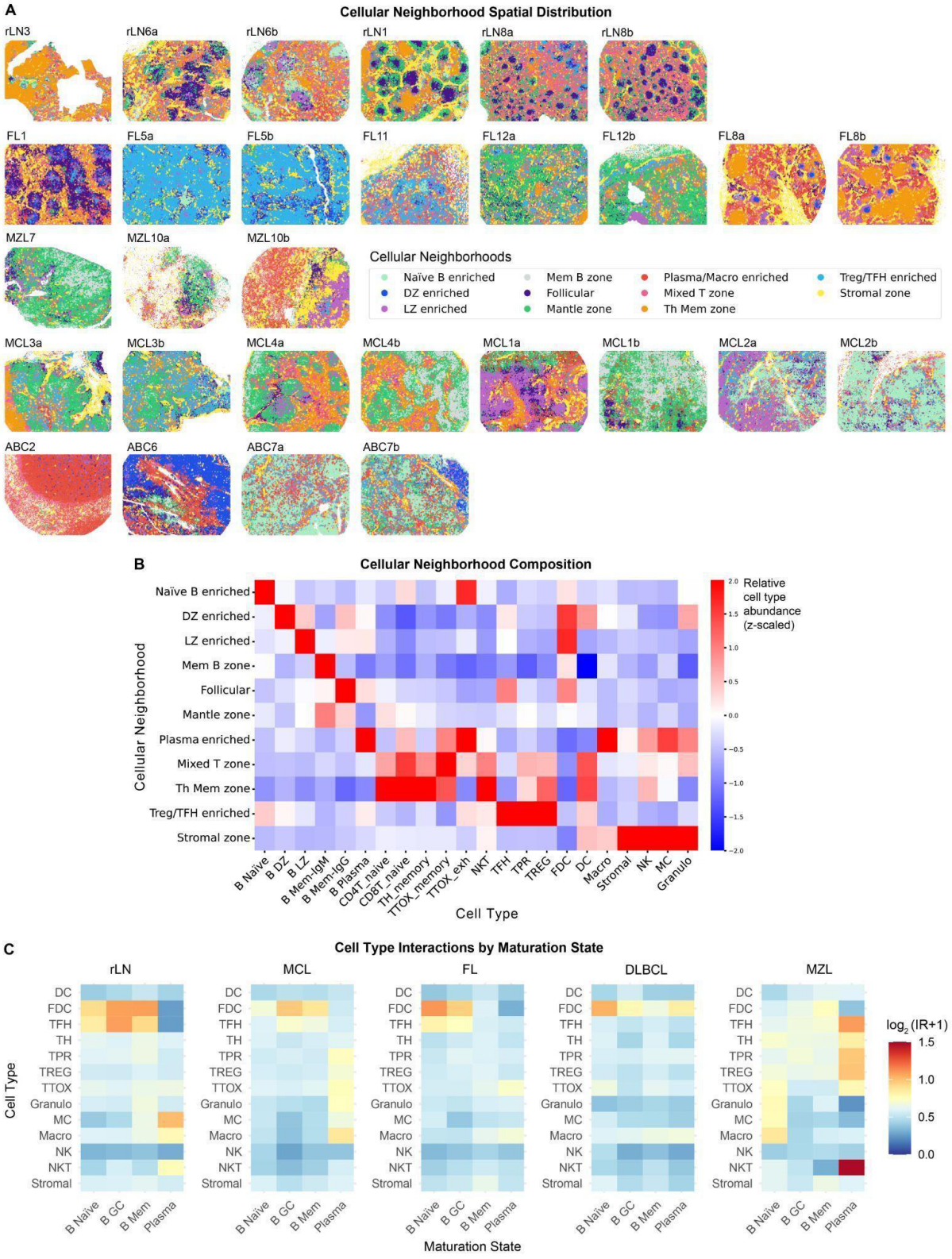
Figure 15: Intratumor maturation states show distinct spatial patterns



UMAP of scRNA-seq data labeled by B-cell maturation states alongside their proportions (top) and spatial distribution of B-cell maturation states on CODEX[51] images (52 markers) of FFPE tissue sections (bottom) from a reactive lymph node (rLN), follicular (FL), mantle cell (MCL), and activated B-cell diffuse large B-cell lymphoma (DLBCL, ABC) sample. See Figure 14B for the distribution of maturation states and cell types on all CODEX slides (n=29). See Figure 3 for maturation state annotations. Adapted from Fitzgerald et al. 2023[1], originally produced by myself and Anastasiia Horlova.

Spatial niches within the CODEX data were uncovered by Anastasiia Horlova using a K nearest neighbor (KNN) graph-based cellular neighborhood (CN) analysis. We annotated 11 CNs (KNN = 20) based on their enriched cell types (Figure 16A-B). Some CNs, such as T-cell zones (Th Mem zone and Mixed T zone) and a stromal zone adjacent to B cells in tumors, were observed in both reactive lymph nodes (rLNs) and tumor samples. Although malignant B cells occupied CNs distinct from those occupied by healthy B cells in rLN samples. Different tumor maturation states occupied distinct CNs, which were often found co-existing within the same sample. For instance, FL1 comprised zones enriched in DZ, LZ, and plasma cells. The infiltration of immune cells like cytotoxic T-cells in tumors is crucial for the efficacy of cellular immunotherapies. We noted that intratumor maturation states were associated with distinct cellular microenvironments with variable immune infiltration, such as greater frequencies of neighboring cytotoxic T-cells (TTOX_memory and TTOX_exh) and macrophages (Macro) in plasma tumor zones (Figure 16A-B).

Figure 16: Cellular neighborhoods and interactions of intratumor maturation states



(A) Spatial distribution of cellular neighborhoods (CNs) in reactive lymph nodes (rLN), mantle cell lymphoma (MCL), follicular lymphoma (FL), diffuse large B-cell lymphoma (DLBCL), and marginal zone lymphoma (MZL) from CODEX images on FFPE sections. CNs (n = 11, KNN = 20) were calculated using all CODEX

images (n=29) and labeled based on their distinguishing features, ie. tumor cells' predominant maturation state (eg. DZ, Mem), B cells' location or function in reactive lymph nodes (eg. Follicular, Mantle zone), or enriched cell type(s) (eg. Mixed T zone, T Mem zone).

(B) Relative abundance of each cell type across cellular neighborhoods (CNs), scaled by cell type frequency.

(C) Log_2 of the pairwise cell-cell observed over expected interaction ratios (IR) between B-cell maturation states and all other cell types in each entity, with a pseudocount of 1. Spatial interactions were determined with Delaunay triangulation. A higher ratio is associated with increased proximity.

B Naïve = naïve B cells, B DZ = centroblasts from the dark zone of the germinal center, B LZ = centrocytes from the light zone of the germinal center, B Mem IgM = IgD+ and IgM+ memory B cells, B Mem IgG = class-switched (IgG+ or IgA+) memory B cells, B Plasma = plasma cells, CD4T_naive = naive CD4+ T-cells, CD8T_naive = naive CD8+ T-cells, TH_memory = memory helper T-cells, TTOX_memory = memory cytotoxic T-cells, TTOX_exh = exhausted cytotoxic T-cells, NKT = natural killer T-cells, TFH = follicular helper T-cells, TPR = proliferating T-cells, TREG = regulatory T-cells, FDC = follicular dendritic cells, DC = dendritic cells, Macro = macrophages, Stromal = stromal cells, NK = natural killer cells, MC = monocytes, Granulo = granulocytes. Adapted from Fitzgerald et al. 2023[1], originally produced by myself, Anastasiia Horlova, and Erin Chung.

The diversity of microenvironments within tumors may facilitate the divergence of maturation states by influencing various cellular interactions, which are essential for different stages of the B-cell maturation process. Follicular dendritic cells (FDCs) and T follicular helper cells (TFH) were predominantly found in GC-rich tumor CNs (DZ and LZ), which could shed light on the differentiation of malignant GC cells into post-GC memory or plasma cells (Figure 16B).

To quantify cell-cell interactions of maturation states within each entity, Erin Chung calculated the pairwise cell-cell observed over expected interaction ratios (IR) between B-cell maturation states and other cell types in each entity. Interactions between GC B cells and FDCs/TFH were partially preserved in tumors, offering an explanation as to how the GC reaction is sustained in B-NHL to promote ongoing maturation (Figure 16C). These interactions were less pronounced in the more diffuse DLBCL and generally post-GC MZL, potentially explaining their reduced diversity in maturation states compared to FL (Figure 7B).

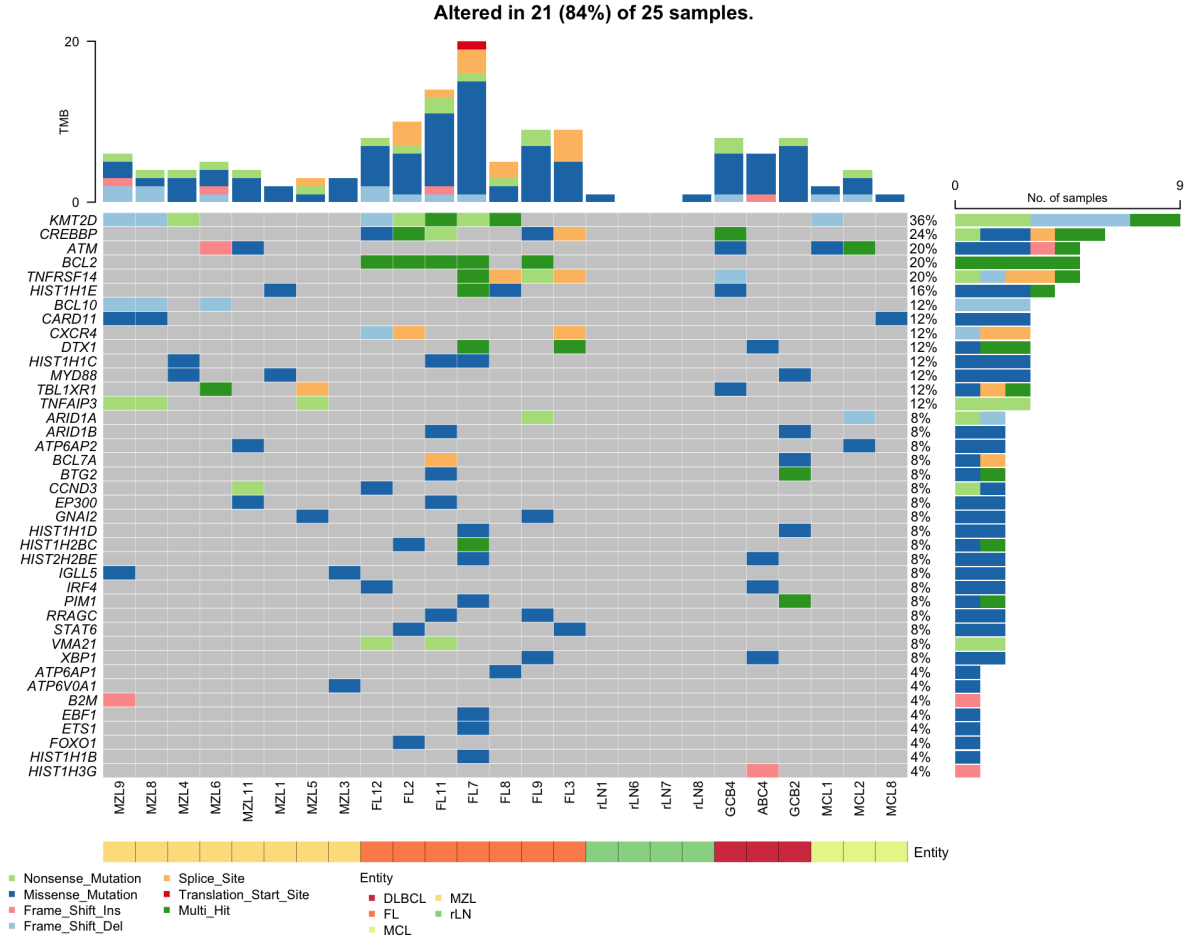
Genetic variation among intratumor maturation states

The processes that govern the differentiation and activation of healthy B cells are often aberrated in B-cell lymphomas to support their unchecked growth and survival. Germinal center B cells have a particularly high susceptibility to malignant transformation because

somatic hypermutation for antibody diversification can lead to oncogenic chromosomal translocations and mutations. Therefore, aberrations are frequently observed in characteristic features of B-cell maturation states, such as *BCL6*, *IRF4*, and *PRDM1*[44,62]. I therefor aimed to explore how intratumor maturation states relate to subclonal genetic variation. To set the stage for this, Verena Passerini conducted genetic variant profiling, including single-nucleotide variants (SNVs), insertions and deletions, and copy number variants (CNVs) through targeted DNA sequencing of 25 samples (4 rLN, 3 MCL, 7 FL, 2 DLBCL GCB, 1 DLBCL ABC, 8 MZL) from the CITE-Seq sample cohort (Figures 17-18).

All B-NHL entities showed increased mutation frequency compared to the rLN samples, with the highest tumor mutational burden (TMB) in FL (Figure 17). Missense mutations were most common, with many genes harboring multi-hit variants. *KMT2D* methyltransferase and *CREBBP* acetyltransferase were the most frequently mutated genes across all B-NHL samples. Apoptosis regulator *BCL2* and immune regulator *TNFRSF14* were commonly aberrated in FL samples, while *ATM* and *BCL10* harbored frequent mutations in MCL and MZL, respectively. Several mutated genes are involved directly or indirectly in the B-cell maturation process. *KMT2D* and *CREBBP* are both critical for the delivery of immune signals while *BCL2* helps maintain survival during selection in the GC LZ[44,77]. *IRF4* and *XBP1*, which are associated with plasma cells[44], were also mutated in a minority of FL and DLBCL samples.

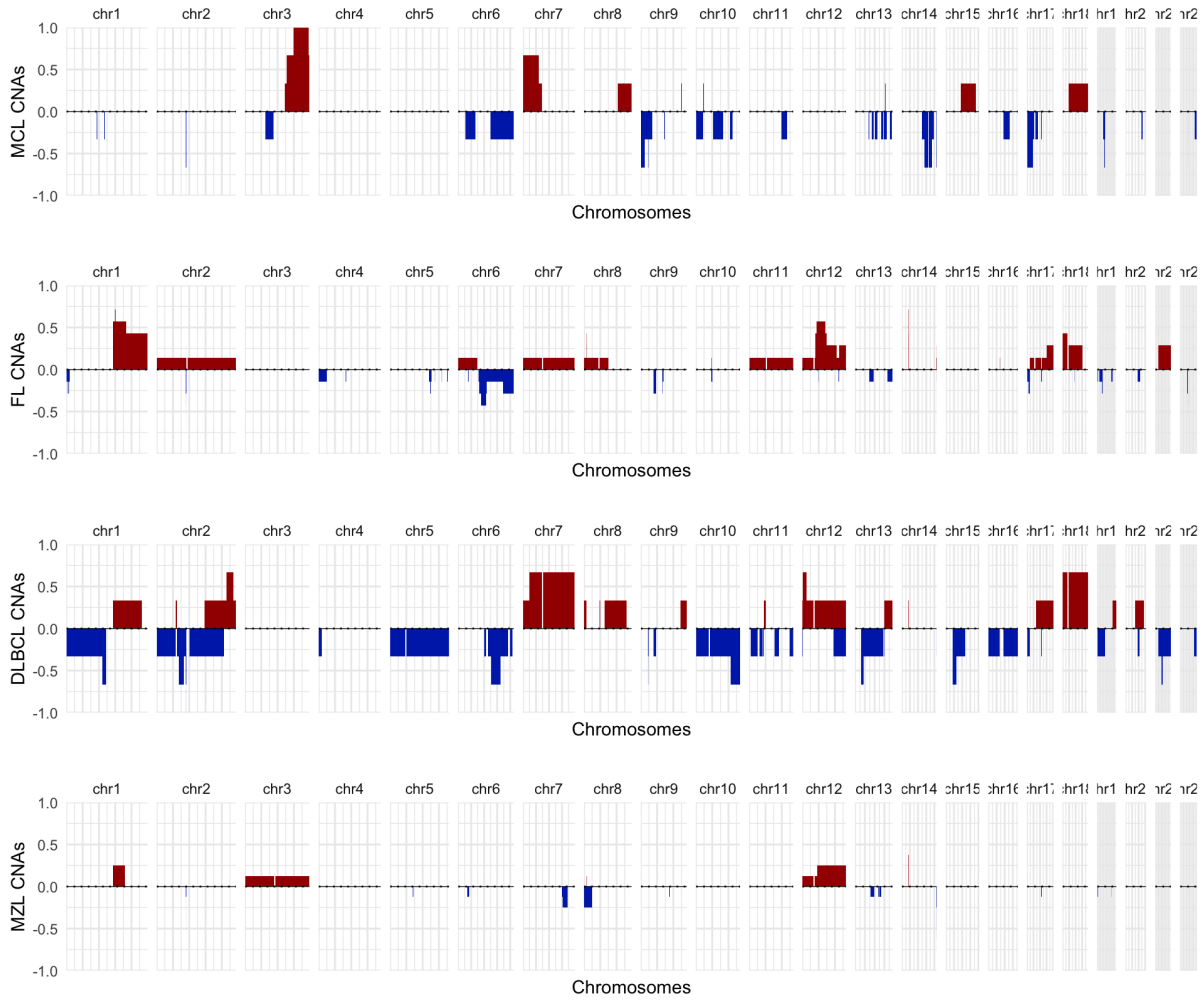
Figure 17: Non-silent mutation profile of tumor samples



Mutation landscape of 25 samples included in the CITE-Seq cohort, as determined by targeted DNA sequencing. All non-silent mutations with VAF >= 10% are depicted. Mutation frequency is shown for samples (top) and genes (right). Adapted from Fitzgerald et al. 2023[1], originally produced by Verena Passerini.

DNA copy number variation (CNV) is also a common feature of B-NHL pathogenesis[78], with implications for disease classification and outcomes[79,80]. B-NHL entities are associated with different hallmark CNVs, such as 6q loss in FL and DLBCL which affects tumor suppressor genes[81]. We observed entity-specific patterns in CNV (Figure 18). All 3 MCL samples showed copy number gain in 3q, while 2 samples showed gain in 7p and loss of 9q. In FL, copy number gain was widespread access the genome, especially in 1q (~50%), and various regions of chromosome 12, while some samples showed 6q loss. DLBCL samples showed the most abundant CNV, with copy number gain in chromosomes 7 and 18 and copy number loss in 6q in 2 out of 3 samples. MZL showed the least CNV, although copy number gain was detected in chromosomes 3 and 12 in a minority of samples (<25%).

Figure 18: Copy number variant profile by entity

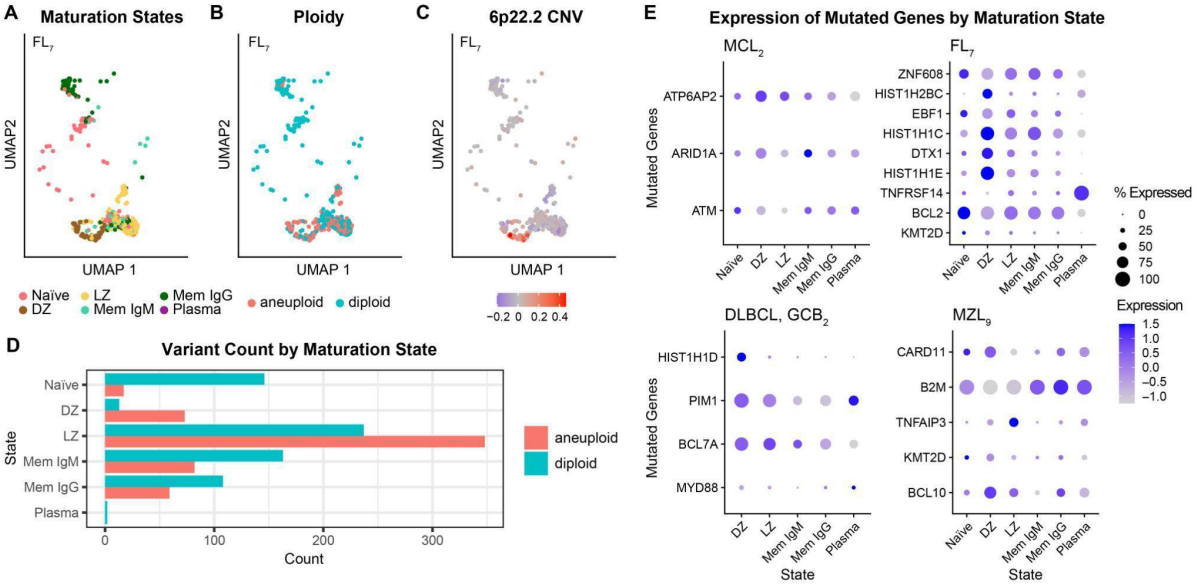


Average genome-wide copy number variation across samples for each entity. Copy numbers were calculated for each sample using CNVkit (v0.9.9), which takes advantage of both on- and off-target sequencing reads and estimates the copy number using a pooled normal reference to compare binned read depths. Log2 change from a pool of normal control of ±0.2 was used as an indication of chromosomal gain or loss. Chromosome X and Y are excluded from the analysis. Adapted from Fitzgerald et al. 2023[1], originally produced by Verena Passerini.

To probe the relationship between intratumor maturation states and genetic variation, I inferred CNV at the single-cell level from the gene expression data in the CITE-Seq dataset (51 samples) using the *copykat* R package[82]. The results revealed CNV among the cells of individual tumors, indicating that CNV in a tumor may be limited to specific groups of tumor cells. Notably, I identified CNV across B-cell maturation states. For instance, in FL7, the majority of germinal center (GC) tumor cells exhibited aneuploidy, unlike its mostly diploid non-GC tumor cells (Figure 19A-D). This tumor demonstrated copy number gain in the 6p22.2 region, which is linked to FL progression[83]. Remarkably, this gain in the 6p22.2

region was predominantly found in the DZ tumor cells, suggesting that DZ cells predisposed to acquiring this specific CNV or its acquisition might either induce a maturation lock at the DZ stage.

Figure 19: Genetic variants among intratumor maturation states



(A-C) UMAP of scRNA-seq data from malignant cells of follicular lymphoma sample FL7 labeled by (A) B-cell maturation state, (B) diploid or aneuploid status, and (C) copy number variation (CNV) in cytogenetic band 6p22.2 (chromosomal position 6-26329011) among tumor cells. CNV was inferred from gene expression with the *copykat* R package[82]. Copy number loss was also detected in this chromosomal region with targeted DNA sequencing (Figure 18).

(D) Frequencies of the aneuploid and diploid variants in each intratumor maturation state in follicular lymphoma sample FL7.

(E) The average expression (z-scaled) and percentage of cells expressing genes with non-silent mutations detected with targeted DNA sequencing (Supplemental DNA Sequencing Report) in mantle cell lymphoma (MCL), follicular lymphoma (FL), diffuse large B-cell lymphoma (DLBCL) and marginal zone lymphoma (MZL) samples.

From Fitzgerald et al. 2023[1], originally produced by myself.

To understand whether mutations are associated with different stages of differentiation, I compared the expression of genes with detected variants across intratumor maturation states. I observed differential expression of mutated genes between maturation states in several tumors: an MCL tumor exhibiting multi-hit mutations in the *ATM* gene, typically inactivated in MCL, displayed regained expression in post-GC states; an FL tumor with multi-hit mutations in *HIST1H1E* and *DTX1*, linked to transformed FL, showed diminished expression exclusively in post-DZ states; a DLBCL tumor with multi-hit mutations in the *PIM1* proto-

oncogene, known for its association with ibrutinib resistance, exhibited decreased expression solely in the memory state; and an MZL tumor with a nonsense mutation in *TNFAIP3*, a frequently mutated NF-kB inhibitor in MZL, demonstrated reduced expression in post-GC states (Figure 19E). These instances suggest that the mutations may have occurred in only a subset of tumor cells, either halting or driving their maturation, thus promoting ITH in maturation states.

Discussion

Cell-type differentiation is central to normal cellular diversification, while cancer types and subtypes are thought to originate from different cell types and states[9,10]. On the other hand, intratumor heterogeneity (ITH), which is central to tumor evolution and treatment response, has typically been considered distinct from the differentiation of cell types. In this thesis, I investigated whether differentiation and ITH may be interconnected. Through a single-cell mult-omic and spatial atlas of nodal B-cell non-Hodgkin lymphomas (B-NHL), a set of cancers tied to distinct stages of the B-cell maturation differentiation lineage, I found that tumors instead consist of multiple co-existing maturation states. This reveals that differentiation is not fixed in cancer, but remains plastic and divergent as a major driver of ITH[1].

Maturation state composition varied significantly between tumors of the same diagnosis, while DLBCL and FL tumors contained mixtures of germinal center and activated B-cell cell-of-origin subtypes. Tumor maturation state composition shifted over time, pointing to a role of differentiation in tumor evolution. The transcription factors that drive different stages of maturation in healthy B cells often maintained differential activity between malignant maturation states, highlighting that the gene regulatory networks driving differentiation are partially preserved in cancer. Intratumor maturation states tended to occupy unique spatial niches, maintaining the cell-type interactions required for ongoing differentiation. I also observed an association between genetic variants and maturation states in tumors, indicating a link between genetic aberrations and cell-type differentiation in cancer evolution[1].

In this Discussion, I consider the conceptual and practical implications of this discovery and its related findings across fields, from cancer pathogenesis and evolution to its diagnosis and management.

Differentiation as a driver of intratumor heterogeneity

Viewing cell-type differentiation as a driver of ITH reshapes our understanding of cancer pathogenesis and evolution. While ITH has long been attributed to clonal evolution guided by genetic variation[5], a deluge of single-cell technologies and studies have propelled our appreciation of the role of non-genetic variation manifesting on the epigenetic, transcriptional, morphological, and metabolic levels[11,18–20]. This phenotypic variation has

been associated with a diverse range of mechanisms, including genetic mutations, epigenetic changes, cellular stress response, and microenvironmental interactions[29]. On the other hand, ITH has typically been viewed as distinct from differentiation. Blocking or evading differentiation is considered inherent to cancer pathogenesis, allowing tumor cells to proliferate at a particular stage indefinitely rather than reach terminal differentiation. For example, several forms of leukemia, including B-NHL, have been thought to arise from a disruption of the normal program of differentiation, such that a committed progenitor becomes able to divide indefinitely *rather* than continuing differentiation[29]. Our findings challenge this paradigm by showing that cancer cells can diversify through ongoing differentiation, revealing that malignancy and differentiation are not mutually exclusive.

The wide spectrum of states observed across the B-cell maturation lineage within tumors indicates that tumor differentiation resembles clonal evolution, a sequential process where at each stage of differentiation a portion of cancer cells continue to differentiate while the remainder stay at the same differentiation stage. This has similarities and differences to the cancer stem cell (CSC) theory, which postulates that within a tumor there exists a subset of cancer cells with stem cell-like properties, CSCs, that can differentiate into non-stem cell cancer cells[39,84]. Like in the CSC theory, we also observe cancer cells can differentiate into multiple cell types. Although the CSC theory holds that cancer cells must start as stem-like cells or acquire a stem cell phenotype before giving rise to differentiated cell types, with the CSCs being the only subset of the tumor capable of prolonged self-renewal. For example, myelogenous leukemia CSCs have been thought to arise from the malignant transformation of normal tissue stem cells, and some B lymphocyte leukemias revert from a committed precursor B cell to a CSC[29]. In B-NHL, however, as B-cell maturation is a sequential differentiation process beginning from mature naïve B cells, we would therefore expect to find a portion of cells within tumors that reflect more stem-like earlier stages in the hematopoietic process. In contrast, I observed the same mature B-cell states that exist in non-malignant lymph nodes, including an abundance of proliferating late-GC and post-GC states. This leads me to wonder to what extent the CSC theory accounts for ITH in cancers of more mature cell types, and how sequential and branched differentiation processes influence clonal evolution.

Differentiation-driven ITH opens several interesting questions. One is whether the directionality of the differentiation process is altered in cancer. Although B-cell maturation is predominantly unidirectional, from naïve to functionally specialized memory or plasma

states[44], we observed maturation states both upstream and downstream from the cell-of-origin associated with each entity, with post-GC states in GC tumors (eg., FL), and, conversely, GC states in post-GC tumors (eg., DLBCL non-GCB and MZL). This raises two potential roles of differentiation in cancer pathogenesis; it may either enable additional states to evolve from the cell of origin that determines the entity or drive the emergence of the entity-determining state from a cell-of-origin earlier in the differentiation process[1].

Revisiting the cell-of-origin model of cancer classification

This plasticity in differentiation and emergence of multiple cell types in tumors leads me to reconsider the role of cell-of-origin in cancer classification. Different cancer types and subtypes are thought to originate from different cell types and states[9], while cell of origin dominates the molecular classification of cancers[10]. It is well established that B-NHL entities relate to distinct stages of B-cell maturation, such as GCB versus ABC DLBCL, exhibiting varying clinical outcomes[49,85]. However, substantial variation in treatment response and survival is also observed between patients of a given cancer entity. We observed substantial variation in maturation state composition between tumors of the same entity, which may drive this *inter*-tumor heterogeneity in clinical outcomes beyond cell of origin.

ITH in cell types and states therefore poses challenges for cancer diagnosis. If the goal of cancer diagnosis is to classify tumors into biologically and clinically similar subgroups for patient stratification, this ITH blurs the boundaries between previously defined subgroups. Notably, using two widely-used DLBCL subtype classifiers, Lymph2Cx[64] and Tally[65], which separate GCB and ABC subtypes known for their distinct prognoses, I observed both GCB and ABC classes within single tumors. This observation is particularly intriguing in FL, where an ABC subtype has not yet been established. Consequently, ITH along the differentiation axis suggests a need to reevaluate the classification of cancer types in B-NHL and other cancers distinguished by cell-of-origin, considering the diverse and dynamic differentiation stages within tumors. Further exploring ITH in cancer cell types in larger clinical studies across cancer types may refine cancer diagnosis and management by accounting for tumor differentiation.

Tumor evolution through differentiation

The concept that tumor cell types and states can adapt over time through differentiation[86]

changes our understanding of tumor evolution. Differentiation status in tumors is not just a passive reflection of their origin, but an active participant in their evolutionary trajectory. Variation in a population confers a survival advantage in natural selection as it enables subpopulations resistant to the selective pressures to emerge. My findings suggest that tumors hijack the normal cellular diversification mechanisms of cell-type differentiation to achieve this variation.

In normal physiology, the diversification of cell types is essential for the function and structure of complex organisms, forming many diverse tissues and organs. Whereas in cancer, it is not clear whether the emergence of distinct cell types through differentiation confers an evolutionary advantage beyond driving variation. Tumors may also form a complex organization of functional cancer cell types, analogous to their normal counterpart tissues. In B-NHL, perhaps intratumor B-cell maturation states maintain survival by continuing the GC reaction, which promotes genetic diversity through somatic hypermutation and tumor growth through proliferation. PD-1 is upregulated following B-cell activation through the BCR, with maturation therefore potentially enabling immune escape. B cells in the tumor microenvironment play a critical role in controlling or promoting tumor growth[87], while tumor and immune cells are thought to co-evolve[88]. Ongoing differentiation in immune cancers provides a direct mechanism for tumors to shape their immune microenvironment.

It is interesting to contemplate how my theory of tumor evolution through differentiation relates to existing models of tumor evolution. The clonal evolution model posits that tumors evolve by a process of clonal expansion, genetic diversification, and clonal selection within the adaptive landscapes of tissue ecosystems. Mutations that confer a selective advantage are more likely to persist and come to dominate the tumor population. Another important aspect in tumor evolution is the tumor-immune microenvironment (TIME). The complex interplay between cancer cells and the immune system shapes the trajectory of tumor development and impacts the effectiveness of therapeutic interventions[89,90]. In the following sections, I will discuss the relationship between these features of clonal evolution in the TIME and differentiation.

Microenvironmental influence on tumor differentiation

In the pathogenesis, clonal evolution, and treatment response of B-NHL, the TIME is crucial, notably affecting survival, growth signaling, and immune response[91]. The organization of B-

cell follicles, consisting of germinal centers and mantle zones, is vital for regular B-cell development. Our finding that intratumor maturation states occupy distinct spatial areas, each characterized by a unique microenvironment, poses that distinct microenvironments within a tumor can promote ITH along differentiation lineages[86]. The presence of follicular dendritic cells or follicular helper T cells, which drive maturation from GC to post-GC states, among GC cancer cells despite the disruption of lymph node follicles shows that tumors maintain spatial niches that support differentiation. FL, which typically preserves B-cell follicles, exhibited a broader range of maturation states compared to GCB DLBCL, suggesting that TIMEs resembling the microenvironment of their normal counterpart may better enable ITH through differentiation.

Nature and nurture: genetic variation and differentiation in cancer

The association between genetic variants and intratumor maturation states in B-NHL implies a tight interplay between clonal evolution and tumor differentiation. The presence of distinct copy number variants as well as differential expression of mutated genes between maturation states within tumors support this hypothesis[86]. Certain genetic changes might predispose cells to malignancy at different maturation stages. This could help explain the varying prevalence of oncogenic mechanisms across different types of B-cell lymphomas[62].

By extending this concept to the topic of ITH, one can consider that differentiation may promote divergent survival and growth mechanisms within the same tumor, thereby driving clonal evolution. I expect exciting research to emerge on characterizing the association between genetic aberrations and intratumor maturation states across space and time.

Therapeutic implications and opportunities

Investigating how tumor evolution through ITH relates to treatment resistance could reveal strategies for preventing relapse. These methods may include tailoring treatments to specific cancer-cell-type compositions or exploiting their differentiation trajectories to reign in tumor evolution. Therapeutic interventions might focus on exploiting weaknesses in certain cancer cell types or preventing the differentiation of tumor cells into more resistant forms. Our identification of distinct signaling pathways and gene regulatory networks associated with B-NHL maturation states offers potential targets for treatment, some of which are already linked

to existing drugs or known ligands, such as *XBP1* and *HDAC2*[92]. Nonetheless, the risk of immunosuppression posed by interfering with the differentiation of immune cells warrants careful consideration.

The growing prominence of T-cell-based immunotherapies, including CAR-T cells and Bi-specific T-cell Engagers (BiTEs), in treating B-cell malignancies underscores their potential in transforming cancer treatment[93,94]. We observed that the patterns of immune cell infiltration, especially cytotoxic T-cells around plasma tumor cells, vary among the tumor's maturation states[86]. This raises the possibility of tailoring immunotherapy to the cancer-cell-type composition, necessitating a combination approach to effectively address the ITH of immune microenvironments. Additionally, the diverse microenvironments surrounding different tumor states offers good reason for expanding immunotherapeutic strategies to include various other immune cell types, such as infiltrating macrophages, NK cells, and non-malignant B cells.

Further questions

The findings discussed in this thesis brings to light a conceptual framework of the role of cell-type differentiation in tumor evolution, but it also raises several further questions beyond the scope of this study which I hope will inspire exciting future research directions.

B-NHL proved a useful model to explore the relationship between differentiation and ITH due to the centrality of cell-of-origin in their pathogenesis and classification[34]. However, cell-of-origin is also considered important in several other cancer types, such as glioma[95], AML[37], and pancreatic adenocarcinoma[38]. Extending the study of ITH from differentiation to other cancer types would help understand its extent and nuances across cancers.

Through longitudinal analysis of maturation state composition in select B-NHL samples, this thesis exemplified how cancer-cell-type composition can change over time, elucidating a role of differentiation in tumor evolution. Larger, more comprehensive longitudinal follow-up studies would allow us to better understand cancer-cell-type evolution and the factors influencing these changes across time and space, including therapeutic and microenvironmental influences.

The association between genetic variants and intratumor maturation states inferred from pairing bulk DNA sequencing with single-cell transcriptomic data suggests a relationship

between genetic diversification and differentiation in clonal evolution. More direct profiling of genetic variation among intratumor maturation states, such as through single-cell DNA sequencing, is necessary to understand the extent of this relationship. Genetic perturbation studies may help understand the causal directionality of this relationship; does genetic diversification promote differentiation or vice versa? This would significantly advance our understanding of and ability to control tumor evolution.

Studying the relationship between intratumor maturation states and drug resistance mechanisms is key to understanding how differentiation influences treatment response. Large scale single-cell drug response screens, such as via flow cytometry or other microfluidic assays, could be employed to compare drug response profiles among cell types and states in a tumor. This may require advancing the feature coverage and cellular resolution of drug screening assays to accurately assign cell types and states (eg. via the transcriptome) and to comprehensively test the wide range of clinically actionable drug combinations and dosages. This would serve both to inform precision medicine strategies and predict tumor evolution according to a tumor's composition and treatment.

The diagnostic and therapeutic implications of differentiation driving ITH warrant extensive clinical investigation. The existence of multiple cell types within tumors, beyond the cell of origin, may require revision of existing cancer classifications, not least in B-NHL. The flux of tumor cell types brought about by differentiation gives cancers a fluid identity beyond what is typically accounted for in cancer diagnosis. Retrospective studies and/or clinical trials on the relationship between tumor cell type composition and patients' clinical outcomes, including survival and treatment response profile, could forge new paths for precision oncology.

Conclusion

In this thesis, I explored the intricate relationship between cell-type differentiation and intratumor heterogeneity (ITH) in nodal B-cell non-Hodgkin lymphomas (B-NHL), offering a novel perspective on cancer pathogenesis and progression. Our findings challenge the fixed cell-type paradigm of cancer by showing that cell types remain dynamic and divergent in tumor evolution through differentiation[86], with significant implications for cancer diagnosis, treatment, and research.

Key findings

1. **Differentiation drives ITH:** The discovery that B-NHL tumors comprise multiple co-existing B-cell maturation states shows that differentiation remains a dynamic and divergent process in cancer, driving ITH. This upends the view that cell-type differentiation is blocked in malignancy, showing that cancer cell types are mutable beyond their cell of origin.
2. **Tumor differentiation blurs cancer subtypes**: The observed variability in tumor maturation state composition suggests a need to revisit the cell-of-origin model in cancer classification, not least in B-NHL where the maturation state of the cell of origin has been associated with the tumor's diagnosis and clinical prognosis.
3. **Tumor evolution through differentiation**: My findings present differentiation as an active process in tumor evolution, driving cellular diversification for clonal evolution. This concept adds a new dimension to the understanding of tumor dynamics beyond the typical focus on genetic diversification and stress response.

Implications for diagnosis and treatment

1. **Diagnostic challenges**: The fluidity of cell types in tumors complicates the diagnosis of cancers, particularly in B-NHL. Diagnostic methods may need to incorporate the assessment of cell-type diversity for more accurate and personalized cancer diagnosis, which may evolve over time with ongoing differentiation.
2. **Therapeutic opportunities**: The identification of distinct signaling pathways and gene regulatory networks associated with B-NHL maturation states opens new avenues for targeted therapies. Understanding tumor evolution through differentiation can also aid in developing strategies to overcome treatment resistance and improve

patient outcomes.

Future research directions

1. **Broadening the study across cancers**: Extending this research to other cancer types will enhance our understanding of the role of differentiation in ITH and its implications across cancers.
2. **Longitudinal studies and genetic analysis**: Longitudinal and genetic profiling of cancer cell types is crucial to unravel the dynamics of tumor evolution through differentiation and the interplay between genetic changes and differentiation over time.
3. **Drug resistance and precision medicine**: Investigating the relationship between intratumor maturation states and drug resistance can inform precision medicine approaches and guide the development of more effective treatment strategies.
4. **Clinical investigations**: Large-scale clinical studies on the impact of tumor cell type composition on patient outcomes could revolutionize cancer diagnosis and treatment, moving towards a more nuanced approach of precision oncology.

In conclusion, this thesis highlights cell-type differentiation as a driver of ITH and tumor evolution, reshaping our understanding of cancer pathogenesis and opening promising pathways for advancements in precision oncology.

Figure 20: Tumor evolution through differentiation

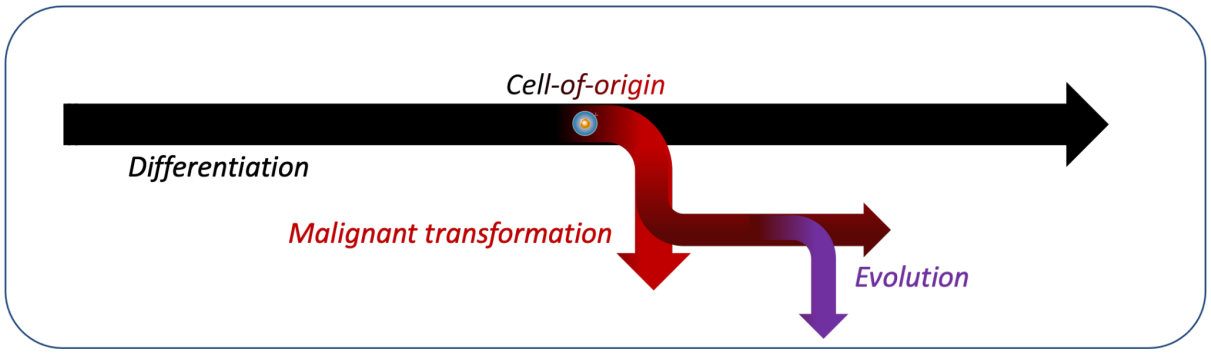


Diagram on the proposed role of differentiation in clonal evolution, whereby the cell of origin undergoes malignant transformation, and subclones diversify through further differentiation and aberrations.

Materials and methods

The text of the following section has been adapted from Fitzgerald et al, *bioRxiv*, 2023[1], and has been originally written by myself. The contributions of collaborators are outlined at the end of the Introduction section.

“

Lymph node sample processing

The University of Heidelberg's Ethics Committee approved our study (S-254/2016), and we secured informed consent from every patient beforehand. We processed and froze patient lymph node (LN) samples for later analysis, following previously described methods[19]. To mitigate the influence of treatment-associated effects on tumor cells and their surroundings, we excluded samples from patients who had undergone allogeneic stem cell transplantation, CAR T-cell, or bispecific antibody therapy from the CITE-Seq cohort. Furthermore, we ensured all samples were collected at least three months post the termination of the most recent treatment to maintain the same control. We provide an overview of the sample cohort in Table 1.

Single-cell 3’ RNA-seq and epitope expression profiling (CITE-Seq)

The cells were thawed, promptly washed to eliminate DMSO, and processed in groups of four to five that comprised a minimum of three distinct entities to avert entity-driven batch effects. A dead cell removal kit from Miltenyi Biotec was employed after thawing, aiming for cell viability of between 85% to 90%. Samples with less than 85% viability were not included. We then stained 5 x 10^5 live cells with a cocktail of oligonucleotide-linked antibodies (Table 2) and left them to incubate at 4°C for 30 minutes. The cells were washed thrice with chilled washing buffer and centrifuged for five minutes at 4°C each time. Following this, cell count and viability were re-evaluated; samples falling below 85% viability were discarded. Subsequently, we prepared the bead-cell suspensions and carried out the synthesis of complementary DNA, single-cell gene expression, and the production of antibody-derived tag (ADT) libraries. For these steps, we used a Chromium single-cell v3.1 3’ kit from 10x Genomics and followed the manufacturer's guidelines.

Single-cell 5' RNA-seq and B-cell receptor repertoire profiling

Apart from epitope staining, sample processing was identical to 3' scRNA-seq. The preparation of the bead-cell suspensions, synthesis of complementary DNA and single-cell gene expression, and BCR libraries were performed using a Chromium single-cell v2 5' and human BCR amplification kit (both 10x Genomics) according to the manufacturer's instructions.

Single-cell library sequencing and data processing

We pooled the 3' gene expression and ADT libraries in a 3:1 ratio, targeting 40,000 reads (gene expression) and 15,000 reads per cell (ADT) respectively, and sequenced them on a NextSeq 500 (Illumina). 5' gene expression libraries were sequenced on a NextSeq 2000 (Illumina), aiming for 50,000 reads per cell. BCR libraries, sequenced on a NextSeq 500 (Illumina), were aimed at achieving a minimum of 5,000 reads per cell.

Post sequencing, we utilized the Cell Ranger software's (10x Genomics, v6.1.1) *cellranger mkfastq* function for demultiplexing and aligning raw base-call files to the reference genome (hg38). For 3' gene and epitope expression libraries, we used the *cellranger count* command on the resulting FASTQ files, while we used *cellranger multi* for 5' gene expression and BCR libraries. For the BCR libraries, we used the VDJ Ensembl reference (hg38, v5.0.0) as a reference. Unless specifically stated otherwise, we adhered to default settings for all functions.

CITE-Seq data analysis

The *Seurat* R package (v4.1.0) was used to perform data quality control, filtering, and normalization (log-based normalization for RNA and centered log-ratio transformation for ADT data). Gene counts per cell, ADT counts per cell, and percentages of mitochondrial reads were computed using the built-in functions. Principal component analysis[96], Louvain clustering[97], and UMAP[98] were performed for the transcriptome (RNA) and epitope (ADT) data independently. After mapping the CD3 and CD19 epitope expression, non-B-cell transcriptomic clusters and doublets were removed. We used the IntegrateData function of the Seurat package for data integration across the different preparation batches. For multimodal clustering, multi-omic factor analysis was performed with the *MOFA2* R

package[60,61] (v1.8) based on the combined transcriptome and epitope data, and the resulting latent factors (n=30) were used as principal components.

5' single-cell RNA-seq data and B-cell receptor profile analysis

Transcriptomic analysis was performed using the R package Seurat (v4.1.0) as described in the *CITE-Seq data analysis* section above. B-cell receptor (BCR) clonotypes were added to the metadata from the *cellranger multi* output.

Sorting of B-cell maturation states from reactive lymph nodes

5 reactive lymph node and 2 tonsil samples were thawed as per a previously published protocol[19]. Using the BD FACSAria Fusion cell sorter equipped with the BD FACSDiva software v. 9.0, the antibody panel outlined in Table 3, and FlowJo analysis software version 10.9.0, B-cell maturation stated were identified and sorted with the gating strategy outlined in Figure 4A and as below. Maturation states and their marker panels were sourced from previous studies which defined naïve[99,100], germinal center[101], memory[99,100], and plasma[102] states with flow cytometry,

Naïve	Naïve B cells	CD19+, CD20+, CD38 low, CD27-, IgD high
DZ	Germinal center dark zone B cells	CD19+, CD20+, CD38+, CD184+, CD83-
LZ	Germinal center light zone B cells	CD19+, CD20+, CD38+, CD184-, CD83+
Mem IgM	IgD/IgM memory B cells	CD19+, CD20+, CD38 low, CD27+, IgM+
Mem IgG	Class-switched memory B cells	CD19+, CD20+, CD38 low, CD27+, IgG+
Plasma	Plasmablasts/plasma cells	CD19+, CD20 low, CD38 high, CD27 high, IgD low

RNA-seq of sorted maturation states

RNA was isolated by the RNeasy Micro Kit (Qiagen, Hilden, Germany) and quantified with Bioanalyzer RNA 6000 pico assay (Agilent, Santa Clara, US). The libraries were generated with NuGENs Trio RNA-Seq System (NuGEN, Redwood City, California) for whole RNA and sequenced on an Illumina NextSeq2000 (Illumina, San Diego, US). Reads were trimmed with TrimGalore v0.6[103] and aligned with hisat2 v2.2.1[104]. The *DESeq2* R package[105] (v1.38.3) was used for differential gene expression analysis between maturation states. Default parameters were used unless otherwise specified. Gene symbols, as per the scRNA-seq datasets, were obtained from *Ensembl*[106] HGNC symbols.

Characterization of B-cell maturation states in the reactive lymph node reference

Clustering and differential expression analysis of the scRNA-seq data from B cells in the integrated reactive lymph node samples (8 samples, 16625 cells) were performed as described in the *Seurat Guided Clustering Tutorial* [107], with the clustering resolution parameter set to 1. Clusters were assigned to known B-cell maturation states based on their differential expression (in RNA and ADT features) of established markers of B-cell maturation states from several sources in the literature[34,44,49,52–56] (Table 5).

Classification of rLN maturation states in single-cell RNA-seq data

The RNA-seq data from the sorted maturation states was filtered by the 2000 most variable RNA features in the rLN scRNA dataset (8 samples combined) and scaled. The resulting matrix was used as input for training a logistic regression model with nested cross-validation for the classification of maturation states using the *nestedcv* package[108–111]. The resulting best-fit model was used to predict maturation states in the log-normalized and scaled scRNA-seq data from reactive lymph nodes. Predicted states were used to validate marker-based maturation state annotations in the CITE-Seq rLN reference dataset.

Mapping of maturation states in all lymph node samples

B-cell maturation states defined in the reactive lymph node reference were mapped to each tumor sample in the CITE-Seq and 5' scRNA-seq datasets using an anchor-based single-cell integration approach outlined in the *Seurat multimodal reference mapping* tutorial[58,112]. Log-normalized counts (without batch-effect correction to prevent bias introduced by sample integration) were used to find transfer anchors and project samples on the reference reductions - PCA (50 dimensions) and UMAP (2 dimensions).

Isolation of malignant B cells

Malignant B cells in tumor samples were identified based on immunoglobulin light chain restriction, whereby malignant (monoclonal) populations of cells are restricted to either the kappa or lambda immunoglobulin light chain and non-malignant B-cell populations (polyclonal) show mixed kappa and lambda light chain positivity[1]. As a minority of ADT counts may be present from ambient unbound antibodies during CITE-Seq library preparation, the proportion of total light chain counts of the kappa subtype (Kappa

counts/(Kappa + Lambda counts)) per cell was used as a surrogate for binary positivity. Transcriptional B-cell clusters with an average kappa light chain proportion of >80% or <20% across all cells were considered malignant. Non-malignant B cells represented a median of 6% [0%, 94%] of all B cells in tumor samples.

Maturation state gene expression signature scoring

Maturation state gene expression signature scores were calculated by averaging the log-normalized counts for the 50 most differentially expressed genes (by fold-change) for each B-cell maturation state annotated in a published tonsil scRNA-seq dataset[49].

Inference of transcription factor activity from single-cell RNA-sequencing data

The pySCENIC[72] workflow was executed through a custom Snakemake pipeline. To infer the Gene Regulatory Network (GRN), we used the GRNBoost2 algorithm from the Arboreto[113] package with 10 perturbations. The analysis was performed on the raw scRNA-seq data. Transcription factor (TF) regulons were predicted using the human v9 motif collection from cisTarget (hg38__refseq-r80__10kb_up_and_down_tss.mc9nr.feather and hg38__refseq-r80__500bp_up_and_100bp_down_tss.mc9nr.feather databases). AUC scores per cell and GRNs were obtained for visualization and downstream analysis. For the final GRN reconstruction, only target genes occurring in more than 95% of the runs were considered. Differential expression (DE) analysis between B-cell maturation stages was conducted using Seurat's FindMarkers[114] function on the RNA assay, utilizing the MAST[74] method for DE analysis from single-cell data. DE genes between conditions in all cell populations were identified (p.adj < 10e-16 & log2FC > 0.4), and p-values were adjusted for multiple comparisons using the Benjamini-Hochberg[73] correction method. To determine differentially active TFs, we utilized the output of the SCENIC GRN, which provided TF activity at a single-cell level. Differentially active TFs were detected using Fisher's exact test to assess the enrichment of maturation stage-specific DE genes among all the TF target genes extracted from the SCENIC GRN (p.adj < 0.05).

CODEX sample preparation

Representative tumor or tumor-free lymph node areas were selected from archival FFPE tissue blocks belonging to18 patients. This selection was made by the certified pathologists at the National Center for Tumor Diseases' Tissue Bank and the University Hospital

Heidelberg's Institute of Pathology as previously described[76]. Two 4.5 mm cores per patient were incorporated into Tissue Microarrays (TMAs). TMA sections (4 μm) were affixed to Vectabond-precoated 25 x 25 mm coverslips, coated with paraffin, and stored for future staining.

Antibody conjugation, validation, and titration

We used the co-detection by indexing (CODEX) approach for multicolor immunofluorescence[51]. Antibodies utilized for CODEX experiments are summarized in Table 4. We reduced purified, carrier-free antibodies with Tris(2-carboxyethyl)phosphine (TCEP) and conjugated them with maleimide-modified CODEX DNA oligonucleotides, procured from TriLink Biotechnologies. A board-certified pathologist supervised the evaluation of the conjugated antibodies in singleplex stains on tonsil and/or lymphoma tissue, comparing with online databases, immunohistochemical reference stains, and published literature. We validated staining patterns in multiplex experiments with positive and negative control antibodies and titrated the appropriate dilution of each antibody starting from 1:100 to optimize the signal-to-noise ratio.

Multiplex tissue staining and fixation

We deparaffinized, and rehydrated coverslips, and subjected them to heat-induced epitope retrieval at pH9 and 97°C for 10 minutes in a Lab Vision PT module. After blocking non-specific binding with CODEX FFPE blocking solution, we stained the coverslips overnight with the full antibody panel at the dilutions shown in Table 4. Following staining, coverslips were fixed with 1.6% paraformaldehyde, methanol, and BS3 fixative, then stored in CODEX buffer S4 until imaging.

Multicycle imaging

We attached stained coverslips to custom acrylic plates and inserted them into a Keyence BZ-X710 inverted fluorescence microscope. We selected 7x7 fields of view and an appropriate number of z-planes (10-14) to capture the best focal plane across the imaging area. Multicycle imaging was performed using a CODEX microfluidics device. Post completion of multicycle imaging, coverslips were stained with hematoxylin/eosin, and the same areas were imaged in brightfield mode.

Image processing

We processed raw TIFF images using the RAPID pipeline[115] in Matlab with the default settings. Post-processing, images were concatenated to hyperstacks. Each tissue core was visually inspected for staining quality using ImageJ/Fiji.

Cell segmentation and cell type annotation

We segmented individual nuclei based on the Hoechst stain and quantified cellular marker expression levels using a modified version of the Mask R-CNN-based CellSeg software. A threshold based on the intensity of the nuclear markers Hoechst and DRAQ5 was used to exclude non-cellular events. Cells were then submitted to Leiden-based clustering using the scanpy Python package, and cluster annotations were assigned according to previously identified cell type marker profiles[76].

CITE-Seq to CODEX data integration for B-cell maturation state label transfer

B-cell maturation states in the CODEX data were classified sample-wise from annotations in the CITE-Seq data using shared features (n = 28) in the CITE-Seq and CODEX antibody panels (Tables 2 and 4). After selecting the shared features, CITE-Seq ADT counts and CODEX fluorescence intensities were subject to the same preprocessing steps of log-ratio normalization and scaling (z-scored) with *Seurat* v4. For each sample, a logistic regression classifier (*glmnet* package, 10-fold nested cross-validation)[110,111] was trained on the annotated CITE-seq data to classify B-cell maturation states. To prevent prediction bias toward majority classes, random sampling was performed to balance class distribution within the splits. For each sample, the resulting best-fit model (with the highest balanced accuracy on the outer folds) was used to predict B-cell maturation states in the sample's corresponding CODEX B-cell data. The median Pearson correlation coefficient between the samples' CITE-Seq and CODEX maturation state proportions was 0.91 ($p = 0.011$).

Cellular neighborhood analysis

We modified a previously described approach for neighborhood analysis[116]. For each cell, the 20 nearest neighbors were determined based on their Euclidean distance of the X and Y coordinates, thereby creating one 'window' of cells per individual cell. Next, we grouped these windows using k-means clustering according to the proportions of cell types within

each window. We selected K=11 for the number of neighborhoods as we observed that higher values of k did not result in an improved biologically interpretable number of neighborhoods. Neighborhoods were annotated based on their biological function in normal lymph nodes or their enriched cell type(s)/state(s).

Cellular interaction likelihood analysis

Spatial graph representations of immediately neighboring cells were constructed based on Delaunay triangulation between centroid coordinates using the *scipy.spatial* Python package[117]. To compute pairwise association strengths between clusters, relative frequencies were computed using the following metric:

$$\frac{Nij \quad \times Nt}{Ni \times Nj}$$

in which Nij is equal to the total number of edges between clusters i and j, Nt the total number of edges in the sample, and Ni and Nj the total degrees of clusters i and j respectively[116]. Computed association strengths were calculated separately for each disease entity, between B-cell states and other cell types.

DNA sequencing

DNA was fragmented (Covaris sonication) to 250 bp and further purified using Agentcourt AMPure XP beads (Beckman Coulter). Size-selected DNA was then ligated to adaptors during library preparation. Each library was quantified using qPCR and analyzed for quality after fragmentation and library preparation based on library yield and size on an Agilent Bioanalyzer. The sample MZL_2 failed at the library preparation stage. Finally, libraries were enriched for genes using the Sure Select XT Target Enrichment System for Illumina Paired-End Multiplexed Sequencing and each capture pool was sequenced at 300-400x.

Pooled samples were demultiplexed using a custom demultiplexing tool. Read pairs were aligned to the hg19 reference sequence using the Burrows-Wheeler Aligner[118], and data were sorted and duplicate-marked using Picard tools (version 2.23.3)[119]. All steps were performed within the bcbio-nextgen toolkit (version 1.2.9)[120].

The minimum quality criterion was 80% of target bases having > 30x sequencing coverage. Cases with 60-79% of target bases with > 30x sequencing coverage were also included if target bases not covered were < 1%. Cases with target bases covered 30x < 60% or cases with target bases covered 30x between 60-80% and target bases not covered > 1% were excluded. This was achieved for all sequenced samples (Supplemental DNA Sequencing Report). Metrics were collected using Picard tools (version 2.23.3)[119].

Variant Analysis

Mutation analysis for single nucleotide variants (SNV) and Insertions and Deletions (InDels) was performed using MuTect2[121] (GATK v4.1.9.0)[122] and annotated by Funcotator[123] (GATK v4.1.9.0). A panel-of-normals (PON) filter was generated using samples annotated as rLN and a panel of normal from the 1000 Genomes Project[124]. Variants were included in the PON if present in two or more normal samples.

Non-silent variants (Missense_Mutation, Nonsense_Mutation, Nonstop_Mutation, Splice_Site, Translation_Start_Site) resulting from BestEffect Funcotator annotation (dataSources.v1.6) at a variant allele frequency of > 10% are kept for further investigations. Germline polymorphisms and sequencing artifacts were excluded by comparison with the panel-of-normals and with the gnomAD database[125]. Known germline polymorphisms from the Exome Sequencing Project[126] and dbSNP[127] databases were excluded. An overview of the somatic variants identified is depicted in the Supplemental DNA Sequencing Report.

Genome-wide copy number aberrations (CNAs) were called using CNVkit (v0.9.9)[128]. Notably, this tool takes advantage of both on- and off-target sequencing reads and estimates the copy number using a pooled normal reference to compare binned read depths. Log2 change from a pool of normal control of ±0.2 was used as an indication of chromosomal gain or loss. Chromosomes X and Y are excluded from the analysis.

Inference of copy number variation from single-cell RNA-sequencing data

Copy number variants (CNVs) and ploidy were inferred from single-cell RNA-sequencing count data in each sample using the copykat R package as per the package vignette[82]. A cell filtering threshold of 5 genes per chromosome and a minimal segmentation window size of 25 genes was used. Copy number variation (Euclidean distance) was determined at a resolution of 5MB chromosomal segments, which was added as a new assay to CITE-Seq

Seurat objects for each sample for visualization of copy number variants across intratumor maturation states.

”

Appendix

Table 1: Sample overview

Sample ID	*Run*	*Entity*	*Age*	*Sex*	*Status*	*Ann-Arbor Stage*
ABC1	15	DLBCL, non-GCB	67	M	Initial diagnosis	IIIB
ABC2	6	DLBCL, non-GCB	63	M	Initial diagnosis	III
ABC3	8	DLBCL, non-GCB	80	M	Initial diagnosis	IIIB
ABC4	8	DLBCL, non-GCB	61	F	Initial diagnosis	IIIB
ABC5	3	DLBCL, non-GCB	38	F	Relapse	IVA
ABC6	8	DLBCL, non-GCB	69	M	Relapse	IAE
ABC7	6	DLBCL, non-GCB	54	M	Relapse	IIIA
FL1	10	FL	64	F	Initial diagnosis	IVB
FL10	2	FL	71	M	Relapse	IVA
FL11	11	FL	74	M	Relapse	IVA
FL12	9	FL	66	M	Relapse	IVA
FL2	16	FL	78	M	Initial diagnosis	IIIA
FL3	3	FL	67	F	Initial diagnosis	IVA
FL4	1	FL	33	M	Initial diagnosis	IVB
FL5	2	FL	54	M	Relapse	IIIA
FL6	1	FL	67	M	Relapse	IVA
FL7	10	FL	59	F	Relapse	IIIA
FL8	8	FL	42	M	Relapse	IIA
FL9	12	FL	72	F	Relapse	IIIA
GCB1	16	DLBCL, GCB	52	F	Relapse	IIIAE
GCB2	5	DLBCL, GCB	84	F	Initial diagnosis	IVA
GCB3	4	DLBCL, GCB	57	M	Initial diagnosis	IIIA
GCB4	13	DLBCL, GCB	45	M	Relapse	IVB
GCB5	15	DLBCL, GCB	77	M	Relapse	IIB
MCL1	4	MCL	63	M	Initial diagnosis	IVB
MCL2	7	MCL	77	M	Initial diagnosis	IVA
MCL3	3	MCL	68	M	Initial diagnosis	IVB
MCL4	9	MCL	50	M	Initial diagnosis	IVB
MCL5	10	MCL	69	M	Initial diagnosis	IVB
MCL6	15	MCL	61	M	Relapse	IVA
MCL7	12	MCL	62	M	Relapse	IVA
MCL8	11	MCL	72	M	Relapse	IVA
MZL1	4	MZL	49	M	Initial diagnosis	IVA
MZL10	11	MZL	78	M	Relapse	IAE

MZL11	9	MZL	63	M	Relapse	IIIA
MZL2	7	MZL	58	F	Initial diagnosis	IV
MZL3	3	MZL	49	F	Initial diagnosis	IVB
MZL4	14	MZL	34	F	Initial diagnosis	IVA
MZL5	16	MZL	71	F	Initial diagnosis	IA
MZL6	14	MZL	59	F	Relapse	IIA
MZL7	6	MZL	82	M	Relapse	IV
MZL8	5	MZL	52	F	Relapse	IVA
MZL9	16	MZL	54	F	Relapse	IVA
rLN1	7	rLN	73	F	n / A	n / A
rLN2	10	rLN	35	F	n / A	n / A
rLN3	6	rLN	33	F	n / A	n / A
rLN4	11	rLN	51	M	n / A	n / A
rLN5	14	rLN	20	M	n / A	n / A
rLN6	4	rLN	46	F	n / A	n / A
rLN7	7	rLN	23	F	n / A	n / A
rLN8	9	rLN	61	M	n / A	n / A

Table 2: CITE-Seq antibody panel

SPECIFICITY	*ALTERNATIVE*	*CLONE*	*ISOTYPE*	*CATALOGUE*	BARCODENO	CATALOGPE	TITRATION	CONCENTRATION	AMOUNTUSED
CD10		HI10a	Mouse IgG1, κ	312231	0062	312203	1.25 µl	100 µg/ml	125 ng
CD103		Ber-ACT8	Mouse IgG1, κ	350231	0145	350205	1.25 µl	12 µg/ml	15 ng
CD11B		ICRF44	Mouse IgG1, κ	301353	0161				
CD11C		S-HCL-3	Mouse IgG2b, κ	371519	0053	371504	0.63 µl	50 µg/ml	31 ng
CD127	IL7R	A019D5	Mouse IgG1, κ	351352	0390	351303	0.63 µl	100 µg/ml	63 ng
CD134		Ber-ACT35	Mouse IgG1, κ	350033	0158	350003	0.63 µl	200 µg/ml	125 ng
CD137		4B4-1	Mouse IgG1, κ	309835	0355	309803	1.25 µl	100 µg/ml	125 ng
CD150	SLAM	A12 (7D4)	Mouse IgG1, κ	306313	0870				
CD152	CTLA4	BNI3	Mouse IgG2a, κ	369619	0151	369603	0.63 µl	200 µg/ml	125 ng
CD16		3G8	Mouse IgG1, κ	302061	0083	302007	0.63 µl	100 µg/ml	63 ng
CD161		HP-3G10	Mouse IgG1, κ	339945	0149	339903	1.25 µl	120 µg/ml	150 ng
CD183	CXCR3	G025H7	Mouse IgG1, κ	353745	0140	353705	0.63 µl	100 µg/ml	63 ng
CD184	CXCR4	12G5	Mouse IgG2a, κ	306531	0366				
CD185	CXCR5	J252D4	Mouse IgG1, κ	356937	0144	356903	1.25 µl	100 µg/ml	125 ng
CD19		HIB19	Mouse IgG1, κ	302259	0050	302207	0.63 µl	50 µg/ml	31 ng
CD194	CCR4	L291H4	Mouse IgG1, κ	359423	0071	359411	0.63 µl	50 µg/ml	31 ng
CD195	CCR5	J418F1	Rat IgG2b, κ	359135	0141	359105	1.25 µl	100 µg/ml	125 ng
CD197	CCR7	G043H7	Mouse IgG2a, κ	353247	0148	353203	2.50 µl	160 µg/ml	400 ng
CD2		TS1/8	Mouse IgG1, κ	309229	0367				
CD20		2H7	Mouse IgG2b, κ	302359	0100	302305	2.50 µl	25 µg/ml	63 ng
CD200	-	OX-104		na	na	329206	0.63 µl	200 µg/ml	125 ng
CD21		Bu32	Mouse IgG1, κ	354915	0181	354903	0.63 µl	50 µg/ml	31 ng
CD22		S-HCL-1	Mouse IgG2b, κ	363514	0393	363503	0.63 µl	100 µg/ml	63 ng
CD223	LAG3	11C3C65	Mouse IgG1, κ	369333	0152	369305	1.25 µl	100 µg/ml	125 ng
CD23		EBVCS-5	Mouse IgG1, κ	338523	0897	338507	2.50 µl	50 µg/ml	125 ng
CD24		ML5	Mouse IgG2a, κ	311137	0180	311105	1.25 µl	200 µg/ml	250 ng
CD244		C1.7	Mouse IgG1, κ	329527	0189	329507	0.63 µl	50 µg/ml	31 ng

CD25	IL2RA	BC96	Mouse IgG1, κ	302643	0085	302605	1.25 µl	50 µg/ml	63 ng
CD27		O323	Mouse IgG1, κ	302847	0154	302807	0.63 µl	100 µg/ml	63 ng
CD273	PDL2	24F.10C12	Mouse IgG2a, κ	329619	0008	329605	0.63 µl	100 µg/ml	63 ng
CD274	PDL1	29E.2A3	Mouse IgG2b, κ	329743	0007	329705	0.63 µl	400 µg/ml	250 ng
CD278	ICOS	C398.4A	Armenian Hamster IgG	313555	0171	313507	0.63 µl	200 µg/ml	125 ng
CD279	PD1	EH12.2H7	Mouse IgG1, κ	329955	0088	329905	1.25 µl	50 µg/ml	63 ng
CD28		CD28.2	Mouse IgG1, κ	302955	0386	302907	1.25 µl	100 µg/ml	125 ng
CD29	part of VLA-4	TS2/16	Mouse IgG1, κ	303027	0369			50 µg/ml	
CD3		UCHT1	Mouse IgG1, κ	300475	0034	300407	0.63 µl	100 µg/ml	63 ng
CD31		WM59		303137	0124	303105	0.63 µl	100 µg/ml	63 ng
CD32		FUN-2	Mouse IgG2b, κ	303223	0142	303205	0.63 µl	50 µg/ml	31 ng
CD357	GITR	108-17	Mouse IgG2a, κ	371225	0360	371203	2.50 µl	100 µg/ml	250 ng
CD366	TIM-3	F38-2E2	Mouse IgG1, κ	345047	0169	345005	1.25 µl	100 µg/ml	125 ng
CD38		HIT2	Mouse IgG1, κ	303541	0389	303505	1.25 µl	100 µg/ml	125 ng
CD39		A1	Mouse IgG1, κ	328233	0176	328207	0.63 µl	50 µg/ml	31 ng
CD4		RPA-T4	Mouse IgG1, κ	300563	0072	300507	0.63 µl	100 µg/ml	63 ng
CD43		CD43-10G7	Mouse IgG1, κ	343209	0357	343203	0.63 µl	400 µg/ml	250 ng
CD44		IM7	Rat IgG2b, κ	103045	0073	103023	0.63 µl	50 µg/ml	31 ng
CD45		HI30	Mouse IgG1, κ	304064	0391	304007	1.25 µl	10 µg/ml	13 ng
CD45RA		HI100	Mouse IgG2b, κ	304157	0063	304107	0.63 µl	12 µg/ml	8 ng
CD45RO		UCHL1	Mouse IgG2a, κ	304255	0087	304205	0.63 µl	40 µg/ml	25 ng
CD47		CC2C6	Mouse IgG1, κ	323129	0026	323108	0.63 µl	80 µg/ml	50 ng
CD48		BJ40	Mouse IgG1, κ	336709	0029	336707	1.25 µl	200 µg/ml	250 ng
CD5		UCHT2	Mouse IgG1, κ	300635	0138	300607	0.63 µl	100 µg/ml	63 ng
CD56		QA17A16	Mouse IgG1, κ	392421	0084	392403	0.63 µl	80 µg/ml	50 ng
CD57		QA17A04	Mouse IgG1, κ	393319	0168	393307	0.63 µl	200 µg/ml	126 ng
CD62L		DREG-56	Mouse IgG1, κ	304847	0147	304805	1.25 µl	25 µg/ml	31 ng
CD69		FN50	Mouse IgG1, κ	310947	0146	310905	1.25 µl	50 µg/ml	63 ng
CD7		CD7-6B7	Mouse IgG2a, κ	343123	0066	343105	0.63 µl	100 µg/ml	63 ng

CD70		113-16	Mouse IgG1	355117	0027	355103	1.25 µl	200 µg/ml	250 ng
CD73		AD2	Mouse IgG1, κ	344029	0577	344003		400 µg/ml	
CD79B		CB3-1	Mouse IgG1, κ	341415	0187	341404	1.25 µl	50 µg/ml	63 ng
CD86	B7-2	IT2.2	Mouse IgG2b, κ	305443	0006				
CD8A		RPA-T8	Mouse IgG1, κ	301067	0080	301007	0.63 µl	100 µg/ml	63 ng
CD95	Fas	DX2	Mouse IgG1, κ	305649	0156	305607	0.63 µl	100 µg/ml	63 ng
ISOTYPE CTRL		MOPC-21	Mouse IgG1, κ	400199	0090				
ISOTYPE CTRL		MPC-11	Mouse IgG2b, κ	400373	0092				
ISOTYPE CTRL		MOPC-173	Mouse IgG2a, κ	400285	0091				
ISOTYPE CTRL		RTK4530	Rat IgG2b, κ	400673	0095				
ISOTYPE CTRL		HTK888	Armenian Hamster IgG	400973	0241				
KAPPA		MHK-49	Mouse IgG1, κ	316531	0894	316507	1.25 µl	20 µg/ml	25 ng
KLRG1/MAFA		SA231A2	Mouse IgG2a, κ	367721	0153	367711	0.63 µl	100 µg/ml	63 ng
LAMBDA		MHL-38	Mouse IgG2a, κ	316627	0898	316607	1.25 µl	20 µg/ml	25 ng
TIGIT		A15153G	Mouse IgG2a, κ	372725	0089	372703	1.25 µl	25 µg/ml	31 ng

Table 3: FACS antibody panel

Specificity	*Species*	*Fluorochrome*	*Clone*	*Supplier*	*Catalogue*
anti-IgM	human	BUV-395	G20-127	BD Biosciences	563903
anti-CD10	human	FITC	HI10a	BD Biosciences	332775
anti-IgD	human	PerCP-Cy5.5	IA6-2	BD Biosciences	561315
anti-CD23	human	APC	M-L233	BD Biosciences	558690
anti-CD38	human	APC-R700	HIT2	BD Biosciences	561979
anti-CD24	human	BV421	ML5	BD Biosciences	562789
anti-CD5	human	BV510	UCHT2	BD Biosciences	563381
anti-CD20	human	BV650	2H7	BD Biosciences	563780
anti-CD21	human	PE	B-ly4	BD Biosciences	555422
anti-IgM	human	PE-CF594	G20-127	BD Biosciences	562539
anti-CD27	human	PE-Cy7	M-T271	BD Biosciences	560609
anti-CD184	human	BV421	12G5	BD Biosciences	566282
anti-CD83	human	PE	HB15e	BD Biosciences	550634
anti-Ig κ Light Chain	human	PE-Cy7	G20-193	BD Biosciences	561328
anti-Ig λ Light Chain	human	PE	JDC-12	BD Biosciences	555797
anti-CD3	human	PE	HIT3a	BD Biosciences	555340
anti-CD19	human	APC-Cy7	SJ25C1	BD Biosciences	557791
anti-CD45R/B220	mouse	PE	RA3-6B2	BD Biosciences	553090
anti-CD45RB	human	FITC	MEM-55	Biolegend	310206

Table 4: CODEX antibody panel

Target	*Alternative*	*Clone*	*Supplier*	*Catalogue*	*CODEX oligo*	*Dilution*	*Exposure [ms]*	*Cycle*
Blank								1
BCL6		K112-91	BD Biosciences	561520	79	1:25	500	2
GATA3		L50-823	Cell Marque	custom	2	1:50	500	3
CD185	CXCR5	D6L3C	Cell Signaling Technology	custom	69	1:100	500	4
Tbet		D6N8B	Cell Signaling Technology	custom	68	1:100	500	5
CD62L		B-8	Santa Cruz Biotechnology	custom	38	1:400	250	6
FoxP3		236A/E7	Invitrogen	14-4777-82	61	1:100	500	7
CD163		EDHu-1	Novus Biologicals	NB110-40686	59	1:50	500	8
Ki67		B56	BD Biosciences	556003	6	1:200	333	9
CD366	TIM3	polyclonal	Novus Biologicals	AF2365	44	1:100	500	10
PAX5		D7H5X	Cell Signaling Technology	custom	66	1:200	200	11
CD134		Ber-ACT35	Biolegend	350002	75	1:100	500	12
IL10		polyclonal	R&D Systems	AF-217-NA	67	1:100	500	13
CD5		vC5/473 + CD5/54/F6	Novus Biologicals	NBP2-34583	25	1:50	500	14
CD206		MM0820-48L31	Abcam	n/a	55	1:200	500	15
CD25	IL2RA	4C9	Cell Marque	custom	24	1:200	500	16
CD16		D1N9L	Cell Signaling Technology	custom	60	1:50	500	17
CD152	CTLA4	BSB-88	BioSB	BSB 2885 (ASR)	30	1:25	500	18
CD79a		HM47	Biolegend	333502	46	1:200	250	19
CD57		HNK-1	Biolegend	359602	29	1:50	500	20
CD34		QBEnd/10	Novus Biologicals	NBP2-34713	11	1:50	500	21
CXCL13		polyclonal	Novus	AF801	41	1:200	500	22
CD21		SP186	Abcam	ab240987	15	1:100	500	23
CD7		MRQ56	Cell Marque	custom	63	1:100	500	24
Podoplanin		D2-40	Biolegend	916606	32	1:200	500	25
CD279	PD1	D4W2J	Cell Signaling Technology	custom	23	1:50	500	26
HLA-DR		EPR3692	Abcam	ab209968	65	1:100	333	27
CD223	LAG3	D2G4O	Cell Signaling Technology	custom	42	1:25	500	28

CD20		rIGEL/773	Novus Biologicals	NBP2-54591	48	1:200	250	29
CD56		MRQ-42	Cell Marque	custom	58	1:100	250	30
CD45RO		UCH-L1	Santa Cruz Biotechnology	custom	5	1:50	500	31
CD278	ICOS	D1K2T	Cell Signaling Technology	custom	74	1:200	500	32
CD90		EPR3132	Abcam	ab181885	57	1:150	500	33
CD4		EPR6855	Abcam	ab181724	20	1:100	500	34
CD11c		EP1347Y	Abcam	ab216655	49	1:200	333	35
CD3		MRQ-39	Cell Marque	custom	33	1:50	500	36
CD68		KP-1	BioLegend	916104	62	1:200	250	37
CD69		EPR21814	Abcam	ab234512	36	1:500	250	38
CD14		EPR3653	Abcam	ab226121	7	1:300	250	39
CD8		C8/144B	Cell Marque	custom	8	1:100	250	40
Kappa light chain		L1C1	Cell Marque	custom	70	1:100	200	41
CD45RA		HI100	Biolegend	304102	21	1:200	125	42
CD11b		EPR1344	Abcam	ab209970	28	1:200	167	43
Granzyme B		EPR20129-217	Abcam	ab219803	81	1:200	250	44
CD31		C31.3 + C31.7 + C31.10	Novus Biologicals	NBP2-47785	51	1:200	167	45
CD45		2B11+PD7/26 3	Novus Biologicals	NBP2-34528	56	1:200	167	46
CD38		EPR4106	Abcam	ab176886	3	1:200	333	47
CD44		IM7	Biolegend	103002	45	1:200	250	48
CD15		MMA	BD Biosciences	559045	14	1:200	25	49
Lambda light chain		Lamb14	Cell Marque	custom	26	1:200	118	50
Mast cell tryptase		AA1	Abcam	ab2378	71	1:200	250	51
DRAQ5		n/a	Cell Signaling Technology	4084L	n/a	1:100	167	52
Hoechst 33342		n/a	Thermo Fisher Scientific	62249	n/a	1:600	7	all cycles

Table 5: B-cell maturation state markers

Marker*	Maturation State(s)		Physiological Function(s)**	CITE-Seq Panel	Reference***
IGKC (Kappa)	All	Clonality indicator	Antigen and Ig receptor binding, immune activation, phagocytosis, Ig immune response to other organisms	✅	*Seifert M et al, Methods Mol Biol, 2019*
IGLCx (Lambda)	All	Clonality indicator	Antigen and Ig receptor binding, immune activation, phagocytosis, Ig immune response to other organisms	✅	*Seifert M et al, Methods Mol Biol, 2019*
CXCR3	Plasma Memory		GPCR for CXCL9, CXCL10 and CXCL11 for leukocyte traffic and chemotactic migration.	❌	*Morgan D and Tergaonkar V, Trends in Immunology, 2022.*
IGHE	Centrocyte (LZ) Plasma Memory	Late (class-switched)	Antigen and Ig receptor binding, immune activation, phagocytosis, Ig immune response to other organisms, circulating Ig complex	❌	*Talay O et al, Nature immunology, 2012*
IGHAx	Centrocyte (LZ) Plasma Memory	Late (class-switched)	Ig receptor binding, antibacterial humoral response, glomerular filtration, respiratory burst.	❌	*Morgan D and Tergaonkar V, Trends in Immunology, 2022.*
IGHGx	Centrocyte (LZ) Plasma Memory	Late (class-switched)	Antigen and Ig receptor binding, immune activation, phagocytosis, Ig immune response.	❌	*Morgan D and Tergaonkar V, Trends in Immunology, 2022.*
FAS (CD95)	(Pre)Memory		TNF-receptor, pro-apoptotic regulator, activation of NF-kappaB, MAPK3/ERK1 and MAPK8/JNK.	✅	*Laidlaw B et al, Nat. Immunol., 2020*
CD38	Centrocyte Plasma Memory	Late (activated)	Synthesis and hydrolysis of cADP for intracellular calcium flux.	✅	*Morgan D and Tergaonkar V, Trends in Immunology, 2022.*
PRDM1	Plasma(blasts)		Repressor of beta-interferon gene expression	❌	*Holmes A et al, J. Exp. Med., 2020*
IRF4	Plasma(blasts)		Regulation of interferons and interferon-inducible genes, negatively regulates Toll-like-receptor (TLR) signaling.	❌	*Holmes A et al, J. Exp. Med., 2020*
TNFRSF17	Plasmablasts		B-cell development, autoimmune response, NF-kappaB and MAPK8/JNK activation (via TNFSF13B/TALL-1/BAFF). cell survival and proliferation (via TRAF)	❌	*Holmes A et al, J. Exp. Med., 2020*
CCR6	(Pre)Memory		Beta chemokine receptor family, B-lineage maturation and antigen-	❌	*Morgan D and Tergaonkar V, Trends in*

			driven B-cell differentiation, migration and recruitment of dendritic and T cells during inflammatory and immunological responses.		*Immunology, 2022.*
BACH2	Centrocytes (LZ) Memory		Enables DNA binding, primary adaptive immune response	❌	*Sidwell et al, Nature, 2016*
GPR183 (EBI2)	(Pre)Memory	Early differentiation	Homing to outer B-cell follicle	❌	*Holmes A et al, J. Exp. Med., 2020*
RELB	Centrocytes (LZ)	Late	RNA polymerase and protein kinase binding. Lymphocyte differentiation and suppression of IF-beta production	❌	*Morgan D and Tergaonkar V, Trends in Immunology, 2022.*
REL	Centrocytes (LZ)	Late	Regulate genes involved in apoptosis, inflammation, the immune response, and oncogenic processes. B-cell survival and proliferation	❌	*Morgan D and Tergaonkar V, Trends in Immunology, 2022.*
ICAM1	Centrocytes (LZ)	Late	Binds CD11a and CD11b integrins, migration to endothelial tissue	❌	*Morgan D and Tergaonkar V, Trends in Immunology, 2022.*
TRAF1	Centrocytes (LZ)	Late	TNF receptor, NFKB activation, anti-apoptotic signalling	❌	*Morgan D and Tergaonkar V, Trends in Immunology, 2022.*
NFKB2	Centrocytes (LZ)	Late	Transcriptional regulation, inflammation, alteration of cell growth	❌	*Morgan D and Tergaonkar V, Trends in Immunology, 2022.*
NFKB1	Centrocytes (LZ)	Late	Transcriptional regulation, inflammation, alteration of cell growth	❌	*Morgan D and Tergaonkar V, Trends in Immunology, 2022.*
CD40	Centrocytes (LZ)	Late	TNF receptor, class-switching, memory B-cell development, germinal center formation	❌	*Morgan D and Tergaonkar V, Trends in Immunology, 2022.*
BLNK	Centrocytes (LZ)		B-cell development	❌	*Morgan D and Tergaonkar V, Trends in Immunology, 2022.*
CD74	Centrocytes (LZ)		Regulation of antigen presentation	❌	*Morgan D and Tergaonkar V, Trends in Immunology, 2022.*
BLK	Centrocytes (LZ)		Cell proliferation, differentiation, B-cell	❌	*Morgan D and Tergaonkar V,*

			receptor signalling, B-cell development		*Trends in Immunology, 2022.*
BTK	Centrocytes (LZ)		B-cell development	❌	*Morgan D and Tergaonkar V, Trends in Immunology, 2022.*
EBI3	Centrocytes (LZ)		Formation of IL-27 (regulation of T cells)	❌	*Morgan D and Tergaonkar V, Trends in Immunology, 2022.*
CD83	Centrocytes (LZ)		Regulation of antigen presentation	❌	*Morgan D and Tergaonkar V, Trends in Immunology, 2022.*
FOXP1	LZ -> DZ	Transition	Transcription regulation (unspecified)	❌	*Morgan D and Tergaonkar V, Trends in Immunology, 2022.*
CFLAR	LZ -> DZ	Transition	Regulation of apoptosis	❌	*Morgan D and Tergaonkar V, Trends in Immunology, 2022.*
FCRL2	LZ -> DZ	Transition	Ig receptor	❌	*Morgan D and Tergaonkar V, Trends in Immunology, 2022.*
SLA	LZ -> DZ	Transition	Cell differentiation; innate immune response; and transmembrane receptor protein tyrosine kinase signaling pathway	❌	*Morgan D and Tergaonkar V, Trends in Immunology, 2022.*
PTPN6	DZ -> LZ	Transition	Hematopoietic cell growth, differentiation, mitotic cycle, and oncogenic transformation	❌	*Morgan D and Tergaonkar V, Trends in Immunology, 2022.*
MS4A1 (CD20)	DZ -> LZ	Transition	B-cell differentiation	✅	*Morgan D and Tergaonkar V, Trends in Immunology, 2022.*
CD72	DZ -> LZ	Transition	Enable signaling receptor binding activity, cell adhesion	❌	*Morgan D and Tergaonkar V, Trends in Immunology, 2022.*
CAMK1	DZ -> LZ	Transition	Calmodulin-dependent protein kinase cascade	❌	*Morgan D and Tergaonkar V, Trends in Immunology, 2022.*
AICDA	Centroblasts (DZ)		Somatic hypermutation	❌	*Morgan D and Tergaonkar V,*

					Trends in Immunology, 2022.
CCNB1	Centroblasts (DZ)	Proliferating	Mitosis	❌	*Cattoretti G et al, Blood, 2006*
BCL6	Centroblasts (DZ)		Transcription repression, blocking IL-4 response	❌	*Morgan D and Tergaonkar V, Trends in Immunology, 2022.*
MME (CD10)	Germinal centre (DZ.LZ)	Mainly centroblasts	Endopeptidase, hormone inactivation	✅	*Goteri G et al, Diagn. Pathol., 2017.*
SELL (CD62L)	Naïve		Lymph node homing	✅	*Morgan D and Tergaonkar V, Trends in Immunology, 2022.*
TCL1A	Naïve		Cell survival (activation of AKT)	❌	*Morgan D and Tergaonkar V, Trends in Immunology, 2022.*
IGHM	Naïve Memory		Antigen and Ig receptor binding, Ig secretion	❌	*Morgan D and Tergaonkar V, Trends in Immunology, 2022.*
IGHD	Naïve		Antigen and Ig receptor binding, IL-1 stimulation	❌	*Morgan D and Tergaonkar V, Trends in Immunology, 2022.*

* Marker names are obtained from the NCBI Gene Database. Aliases are listed in brackets if used elsewhere in this study.

** Gene functions from the NCBI Gene database are presented in summary form.

*** Publication linking the marker with its respective maturation state.

Abbreviations and glossary

B-NHL: B-cell non-Hodgkin lymphomas

ITH: Intratumor heterogeneity

R-CHOP: Rituximab with cyclophosphamide, hydroxydaunorubicin, oncovin, and prednisolone/prednisone chemoimmunotherapy combination therapy

GC: Germinal center of lymph node B-cell follicles

DZ: Dark zone germinal center B cells (centroblasts)

LZ: Light zone germinal center B cells (centrocytes)

Mem IgM: IgM+/IgD+ memory B cells

Mem IgG: Class-switched memory B cells, including IgG, IgA or IgE isotypes

Plasma: Plasma cells or plasmablasts

rLN: reactive lymph node(s)

MCL: Mantle cell lymphoma

FL: Follicular lymphoma

DLBCL: Diffuse large B-cell lymphoma

GCB: Germinal center B cell

ABC: Activated B cell

MZL: Marginal zone lymphoma

CNV: Copy number variation/variants

List of figures and tables

References

1. Fitzgerald, D. *et al.* A single-cell multi-omic and spatial atlas of B cell lymphomas reveals differentiation drives intratumor heterogeneity. *bioRxiv* 2023.11.06.565756 (2023) doi:10.1101/2023.11.06.565756.
2. Medawar, P. B. & Medawar, J. S. *The Life Science: Current Ideas of Biology*. (Wildwood House, 1977).
3. Darwin, C. *On the origin of species by means of natural selection, or, the preservation of favoured races in the struggle for life*. (MVB E-Books, 2010). doi:10.4324/9780203509104/origin-species-1859-charles-darwin.
4. Greaves, M. & Maley, C. C. Clonal evolution in cancer. *Nature* **481**, 306–313 (2012).
5. Nowell, P. C. The clonal evolution of tumor cell populations. *Science* **194**, 23–28 (1976).
6. Klein, C. A. Parallel progression of primary tumours and metastases. *Nat. Rev. Cancer* **9**, 302–312 (2009).
7. Malaise, E. P., Chavaudra, N. & Tubiana, M. The relationship between growth rate, labelling index and histological type of human solid tumours. *Eur. J. Cancer* **9**, 305–312 (1973).
8. Merlo, L. M. F., Pepper, J. W., Reid, B. J. & Maley, C. C. Cancer as an evolutionary and ecological process. *Nat. Rev. Cancer* **6**, 924–935 (2006).
9. Visvader, J. E. Cells of origin in cancer. *Nature* **469**, 314–322 (2011).
10. Hoadley, K. A. *et al.* Cell-of-Origin Patterns Dominate the Molecular Classification of 10,000 Tumors from 33 Types of Cancer. *Cell* **173**, 291-304.e6 (2018).
11. Black, J. R. M. & McGranahan, N. Genetic and non-genetic clonal diversity in cancer evolution. *Nat. Rev. Cancer* **21**, 379–392 (2021).
12. O'Leary, B. *et al.* The Genetic Landscape and Clonal Evolution of Breast Cancer

Resistance to Palbociclib plus Fulvestrant in the PALOMA-3 Trial. *Cancer Discov.* **8**, 1390–1403 (2018).

13. Landau, D. A., Carter, S. L., Getz, G. & Wu, C. J. Clonal evolution in hematological malignancies and therapeutic implications. *Leukemia* **28**, 34–43 (2014).
14. Harvey, W. T. *et al.* SARS-CoV-2 variants, spike mutations and immune escape. *Nat. Rev. Microbiol.* **19**, 409–424 (2021).
15. Boutros, P. C. *et al.* Spatial genomic heterogeneity within localized, multifocal prostate cancer. *Nat. Genet.* **47**, 736–745 (2015).
16. McPherson, A. *et al.* Divergent modes of clonal spread and intraperitoneal mixing in high-grade serous ovarian cancer. *Nat. Genet.* **48**, 758–767 (2016).
17. Marusyk, A., Almendro, V. & Polyak, K. Intra-tumour heterogeneity: a looking glass for cancer? *Nat. Rev. Cancer* **12**, 323–334 (2012).
18. McGranahan, N. & Swanton, C. Clonal Heterogeneity and Tumor Evolution: Past, Present, and the Future. *Cell* **168**, 613–628 (2017).
19. Roider, T., Brinkmann, B. J. & Dietrich, S. Processing human lymph node samples for single-cell assays. *STAR Protoc* **2**, 100914 (2021).
20. Alizadeh, A. A. *et al.* Toward understanding and exploiting tumor heterogeneity. *Nat. Med.* **21**, 846–853 (2015).
21. Gavish, A. *et al.* Hallmarks of transcriptional intratumour heterogeneity across a thousand tumours. *Nature* **618**, 598–606 (2023).
22. Quail, D. F. & Joyce, J. A. Microenvironmental regulation of tumor progression and metastasis. *Nat. Med.* **19**, 1423–1437 (2013).
23. Roider, T. *et al.* Dissecting intratumour heterogeneity of nodal B-cell lymphomas at the transcriptional, genetic and drug-response levels. *Nat. Cell Biol.* **22**, 896–906 (2020).
24. Goyal, Y. *et al.* Diverse clonal fates emerge upon drug treatment of homogeneous

cancer cells. *Nature* **620**, 651–659 (2023).

25. Nguyen, D. X., Bos, P. D. & Massagué, J. Metastasis: from dissemination to organ-specific colonization. *Nat. Rev. Cancer* **9**, 274–284 (2009).
26. Andor, N. *et al.* Pan-cancer analysis of the extent and consequences of intratumor heterogeneity. *Nat. Med.* **22**, 105–113 (2016).
27. Gerlinger, M. *et al.* Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *N. Engl. J. Med.* **366**, 883–892 (2012).
28. Bayat Mokhtari, R. *et al.* Combination therapy in combating cancer. *Oncotarget* **8**, 38022–38043 (2017).
29. Alberts, B. *et al. Molecular Biology of the Cell*. (Garland Science, 2002).
30. Tabula Sapiens Consortium* *et al.* The Tabula Sapiens: A multiple-organ, single-cell transcriptomic atlas of humans. *Science* **376**, eabl4896 (2022).
31. Waddington, C. H. *The Strategy of the Genes*. (Routledge, 2014).
32. Xu, J., Du, Y. & Deng, H. Direct lineage reprogramming: strategies, mechanisms, and applications. *Cell Stem Cell* **16**, 119–134 (2015).
33. Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**, 663–676 (2006).
34. Seifert, M., Scholtysik, R. & Küppers, R. Origin and Pathogenesis of B Cell Lymphomas. in *Lymphoma: Methods and Protocols* (ed. Küppers, R.) 1–33 (Springer New York, 2019).
35. Küppers, R. Mechanisms of B-cell lymphoma pathogenesis. *Nat. Rev. Cancer* **5**, 251–262 (2005).
36. Alizadeh, A. A. *et al.* Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* **403**, 503–511 (2000).
37. Shlush, L. I. *et al.* Identification of pre-leukaemic haematopoietic stem cells in acute

leukaemia. *Nature* **506**, 328–333 (2014).

38. Flowers, B. M. *et al.* Cell of Origin Influences Pancreatic Cancer Subtype. *Cancer Discov.* **11**, 660–677 (2021).
39. Reya, T., Morrison, S. J., Clarke, M. F. & Weissman, I. L. Stem cells, cancer, and cancer stem cells. *Nature* **414**, 105–111 (2001).
40. Batlle, E. & Clevers, H. Cancer stem cells revisited. *Nat. Med.* **23**, 1124–1134 (2017).
41. Plaks, V., Kong, N. & Werb, Z. The cancer stem cell niche: how essential is the niche in regulating stemness of tumor cells? *Cell Stem Cell* **16**, 225–238 (2015).
42. Cai, S. F. *et al.* Leukemia Cell of Origin Influences Apoptotic Priming and Sensitivity to LSD1 Inhibition. *Cancer Discov.* **10**, 1500–1513 (2020).
43. Sung, H. *et al.* Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J. Clin.* **71**, 209–249 (2021).
44. Morgan, D. & Tergaonkar, V. Unraveling B cell trajectories at single cell resolution. *Trends Immunol.* **43**, 210–229 (2022).
45. Koues, O. I., Oltz, E. M. & Payton, J. E. Short-Circuiting Gene Regulatory Networks: Origins of B Cell Lymphoma. *Trends Genet.* **31**, 720–731 (2015).
46. Bahler, D. W., Pindzola, J. A. & Swerdlow, S. H. Splenic marginal zone lymphomas appear to originate from different B cell types. *Am. J. Pathol.* **161**, 81–88 (2002).
47. BioRender. Created with BioRender. com. Preprint at (2021).
48. Shah, B. D., Martin, P. & Sotomayor, E. M. Mantle cell lymphoma: a clinically heterogeneous disease in need of tailored approaches. *Cancer Control* **19**, 227–235 (2012).
49. Holmes, A. B. *et al.* Single-cell analysis of germinal-center B cells informs on lymphoma cell of origin and outcome. *J. Exp. Med.* **217**, (2020).

50. Stoeckius, M. *et al.* Simultaneous epitope and transcriptome measurement in single cells. *Nat. Methods* **14**, 865–868 (2017).

51. Black, S. *et al.* CODEX multiplexed tissue imaging with DNA-conjugated antibodies. *Nat. Protoc.* **16**, 3802–3835 (2021).

52. Laidlaw, B. J., Duan, L., Xu, Y., Vazquez, S. E. & Cyster, J. G. The transcription factor Hhex cooperates with the corepressor Tle3 to promote memory B cell development. *Nat. Immunol.* **21**, 1082–1093 (2020).

53. Sidwell, T. & Kallies, A. Bach2 is required for B cell and T cell memory differentiation. *Nature immunology* vol. 17 744–745 (2016).

54. Cattoretti, G. *et al.* Nuclear and cytoplasmic AID in extrafollicular and germinal center B cells. *Blood* **107**, 3967–3975 (2006).

55. Goteri, G. *et al.* Comparison of germinal center markers CD10, BCL6 and human germinal center-associated lymphoma (HGAL) in follicular lymphomas. *Diagn. Pathol.* **6**, 97 (2011).

56. Delia, D. *et al.* CD1c but neither CD1a nor CD1b molecules are expressed on normal, activated, and malignant human B cells: identification of a new B-cell subset. *Blood* **72**, 241–247 (1988).

57. Cox, D. R. The Regression Analysis of Binary Sequences. *J. R. Stat. Soc. Series B Stat. Methodol.* **20**, 215–242 (1958).

58. Stuart, T. *et al.* Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888-1902.e21 (2019).

59. Woolson, R. F. Wilcoxon Signed-Rank Test. *Wiley Encyclopedia of Clinical Trials* (2008) doi:10.1002/9780471462422.eoct979.

60. Argelaguet, R. *et al.* Multi-Omics Factor Analysis-a framework for unsupervised integration of multi-omics data sets. *Mol. Syst. Biol.* **14**, e8124 (2018).

61. Argelaguet, R. *et al.* MOFA+: a statistical framework for comprehensive integration of multi-modal single-cell data. *Genome Biol.* **21**, 111 (2020).

62. Shaffer, A. L., 3rd, Young, R. M. & Staudt, L. M. Pathogenesis of human B cell lymphomas. *Annu. Rev. Immunol.* **30**, 565–610 (2012).

63. Chen, Y. *et al.* CTLA-4 promotes lymphoma progression through tumor stem cell enrichment and immunosuppression. *Open Life Sci* **16**, 909–919 (2021).

64. Scott, D. W. *et al.* Determining cell-of-origin subtypes of diffuse large B-cell lymphoma using gene expression in formalin-fixed paraffin-embedded tissue. *Blood* **123**, 1214–1217 (2014).

65. Meyer, P. N. *et al.* Immunohistochemical methods for predicting cell of origin and survival in patients with diffuse large B-cell lymphoma treated with rituximab. *J. Clin. Oncol.* **29**, 200–207 (2011).

66. Bustin, S. Molecular Biology of the Cell, Sixth Edition; ISBN: 9780815344643; and Molecular Biology of the Cell, Sixth Edition, The Problems Book; ISBN 9780815344537. *Int. J. Mol. Sci.* **16**, 28123 (2015).

67. Hagman, J. & Lukin, K. Transcription factors drive B cell development. *Curr. Opin. Immunol.* **18**, 127–134 (2006).

68. Nutt, S. L., Heavey, B., Rolink, A. G. & Busslinger, M. Commitment to the B-lymphoid lineage depends on the transcription factor Pax5. *Nature* **401**, 556–562 (1999).

69. Turner, C. A., Jr, Mack, D. H. & Davis, M. M. Blimp-1, a novel zinc finger-containing protein that can drive the maturation of B lymphocytes into immunoglobulin-secreting cells. *Cell* **77**, 297–306 (1994).

70. Dent, A. L., Shaffer, A. L., Yu, X., Allman, D. & Staudt, L. M. Control of inflammation, cytokine expression, and germinal center formation by BCL-6. *Science* **276**, 589–592 (1997).

71. Basso, K. & Dalla-Favera, R. Germinal centres and B cell lymphomagenesis. *Nat. Rev. Immunol.* **15**, 172–184 (2015).

72. Aibar, S. *et al.* SCENIC: single-cell regulatory network inference and clustering. *Nat. Methods* **14**, 1083–1086 (2017).

73. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Series B Stat. Methodol.* **57**, 289–300 (1995).

74. Finak, G. *et al.* MAST: a flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data. *Genome Biol.* **16**, 278 (2015).

75. Colombo, A. R. *et al.* Single-cell spatial analysis of tumor immune architecture in diffuse large B-cell lymphoma. *Blood Adv* **6**, 4675–4690 (2022).

76. Roider, T. *et al.* Multimodal and spatially resolved profiling identifies distinct patterns of T-cell infiltration in nodal B-cell lymphoma entities. *bioRxiv* 2022.11.04.514366 (2022) doi:10.1101/2022.11.04.514366.

77. Pasqualucci, L. *et al.* Inactivating mutations of acetyltransferase genes in B-cell lymphoma. *Nature* **471**, 189–195 (2011).

78. Cascione, L., Aresu, L., Baudis, M. & Bertoni, F. DNA Copy Number Changes in Diffuse Large B Cell Lymphomas. *Front. Oncol.* **10**, 584095 (2020).

79. Chen, W. *et al.* Array comparative genomic hybridization reveals genomic copy number changes associated with outcome in diffuse large B-cell lymphomas. *Blood* **107**, 2477–2485 (2006).

80. Sebastián, E. *et al.* High-resolution copy number analysis of paired normal-tumor samples from diffuse large B cell lymphoma. *Ann. Hematol.* **95**, 253–262 (2016).

81. Bouska, A. *et al.* Genome-wide copy-number analyses reveal genomic abnormalities

involved in transformation of follicular lymphoma. *Blood* **123**, 1681–1690 (2014).

82. Gao, R. *et al.* Delineating copy number and clonal substructure in human tumors from single-cell transcriptomes. *Nat. Biotechnol.* **39**, 599–608 (2021).
83. Gao, F. *et al.* Integrative genomic and transcriptomic analysis reveals genetic alterations associated with the early progression of follicular lymphoma. *Br. J. Haematol.* (2023) doi:10.1111/bjh.18974.
84. Hope, K. J., Jin, L. & Dick, J. E. Acute myeloid leukemia originates from a hierarchy of leukemic stem cell classes that differ in self-renewal capacity. *Nat. Immunol.* **5**, 738–743 (2004).
85. Shankland, K. R., Armitage, J. O. & Hancock, B. W. Non-Hodgkin lymphoma. *Lancet* **380**, 848–857 (2012).
86. Fitzgerald, D. *et al.* A single-cell multi-omic and spatial atlas of nodal B-cell lymphomas reveals B-cell maturation drives intratumor heterogeneity. *bioRxiv* 2023.11.06.565756 (2023) doi:10.1101/2023.11.06.565756.
87. Downs-Canner, S. M., Meier, J., Vincent, B. G. & Serody, J. S. B Cell Function in the Tumor Microenvironment. *Annu. Rev. Immunol.* **40**, 169–193 (2022).
88. Liu, R. *et al.* Co-evolution of tumor and immune cells during progression of multiple myeloma. *Nat. Commun.* **12**, 2559 (2021).
89. Bozic, I. & Wu, C. J. Delineating the evolutionary dynamics of cancer from theory to reality. *Nat Cancer* **1**, 580–588 (2020).
90. Zahir, N., Sun, R., Gallahan, D., Gatenby, R. A. & Curtis, C. Characterizing the ecological and evolutionary dynamics of cancer. *Nat. Genet.* **52**, 759–767 (2020).
91. Scott, D. W. & Gascoyne, R. D. The tumour microenvironment in B cell lymphomas. *Nat. Rev. Cancer* **14**, 517–534 (2014).
92. Pharos: Illuminating the Druggable Genome. https://pharos.nih.gov/.

93. Katz, D. A. *et al.* Open-Label, Phase 2 Study of Blinatumomab after First-Line Rituximab-Chemotherapy in Adults with Newly Diagnosed, High-Risk Diffuse Large B-Cell Lymphoma. *Blood* **134**, 4077–4077 (2019).

94. Sermer, D. *et al.* Outcomes in patients with DLBCL treated with commercial CAR T cells compared with alternate therapies. *Blood Adv* **4**, 4669–4678 (2020).

95. Alcantara Llaguno, S. R. & Parada, L. F. Cell of origin of glioma: biological and clinical implications. *Br. J. Cancer* **115**, 1445–1450 (2016).

96. Pearson, K. LIII. On lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* **2**, 559–572 (1901).

97. Blondel, V. D., Guillaume, J.-L., Lambiotte, R. & Lefebvre, E. Fast unfolding of communities in large networks. *J. Stat. Mech.* **2008**, P10008 (2008).

98. McInnes, L., Healy, J., Saul, N. & Großberger, L. UMAP: Uniform Manifold Approximation and Projection. *J. Open Source Softw.* **3**, 861 (2018).

99. Kibler, A. *et al.* Systematic memory B cell archiving and random display shape the human splenic marginal zone throughout life. *J. Exp. Med.* **218**, (2021).

100. Seifert, M. *et al.* Functional capacities of human IgM memory B cells in early inflammatory responses and secondary germinal center reactions. *Proc. Natl. Acad. Sci. U. S. A.* **112**, E546-55 (2015).

101. Victora, G. D. *et al.* Identification of human germinal center light and dark zone cells and their relationship to human B-cell lymphomas. *Blood* **120**, 2240–2248 (2012).

102. Avery, D. T. *et al.* Increased expression of CD27 on activated human memory B cells correlates with their commitment to the plasma cell lineage. *J. Immunol.* **174**, 4034–4042 (2005).

103. Krueger, F. Trim Galore!: A wrapper around Cutadapt and FastQC to consistently apply

adapter and quality trimming to FastQ files, with extra functionality for RRBS data. *Babraham Institute* (2015).

104. Kim, D., Paggi, J. M., Park, C., Bennett, C. & Salzberg, S. L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* **37**, 907–915 (2019).
105. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
106. Cunningham, F. *et al.* Ensembl 2022. *Nucleic Acids Res.* **50**, D988–D995 (2022).
107. Seurat - Guided Clustering Tutorial. https://satijalab.org/seurat/articles/pbmc3k_tutorial.html.
108. Stone, M. An Asymptotic Equivalence of Choice of Model by Cross-Validation and Akaike's Criterion. *J. R. Stat. Soc. Series B Stat. Methodol.* **39**, 44–47 (1977).
109. Piironen, J. & Vehtari, A. Sparsity information and regularization in the horseshoe and other shrinkage priors. *EJSS* **11**, 5018–5051 (2017).
110. Friedman, J. *et al.* Package 'glmnet.' *CRAN R Repositary* (2021).
111. Lewis, M. J. *et al.* nestedcv: an R package for fast implementation of nested cross-validation with embedded feature selection designed for transcriptomics and high dimensional data. *Bioinformatics Advances* vbad048 (2023).
112. Multimodal reference mapping. https://satijalab.org/seurat/articles/multimodal_reference_mapping.html.
113. Moerman, T. *et al.* GRNBoost2 and Arboreto: efficient and scalable inference of gene regulatory networks. *Bioinformatics* **35**, 2159–2161 (2019).
114. Hao, Y. *et al.* Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573-3587.e29 (2021).
115. Lu, G. *et al.* A real-time GPU-accelerated parallelized image processor for large-scale

multiplexed fluorescence microscopy data. *Front. Immunol.* **13**, 981825 (2022).

116. Schürch, C. M. *et al.* Coordinated Cellular Neighborhoods Orchestrate Antitumoral Immunity at the Colorectal Cancer Invasive Front. *Cell* **182**, 1341-1359.e19 (2020).
117. Virtanen, P. *et al.* SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat. Methods* **17**, 261–272 (2020).
118. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
119. Wysoker, A., Tibbetts, K. & Fennell, T. Picard tools version 1.90. *Jhpsn*.
120. Guimera, R. V. bcbio-nextgen: Automated, distributed next-gen sequencing pipeline. *EMBnet J.* **17**, 30 (2012).
121. Benjamin, D. *et al.* Calling Somatic SNVs and Indels with Mutect2. *bioRxiv* 861054 (2019) doi:10.1101/861054.
122. *gatk: Official code repository for GATK versions 4 and up*. (Github).
123. Funcotator Information and Tutorial. *GATK* https://gatk.broadinstitute.org/hc/en-us/articles/360035889931-Funcotator-Information-and-Tutorial.
124. 1000 Genomes Project Consortium *et al.* A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
125. Karczewski, K. The Genome Aggregation Database (gnomAD). https://ibg.colorado.edu/cdrom2019/nealeB/Gnomad/boulder_190307.pdf.
126. Tennessen, J. A. *et al.* Evolution and functional impact of rare coding variation from deep sequencing of human exomes. *Science* **337**, 64–69 (2012).
127. Sherry, S. T. *et al.* dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.* **29**, 308–311 (2001).
128. Talevich, E., Shain, A. H., Botton, T. & Bastian, B. C. CNVkit: Genome-Wide Copy Number Detection and Visualization from Targeted DNA Sequencing. *PLoS Comput.*

Biol. **12**, e1004873 (2016).

Publications

The work presented in this thesis is accompanied by the following preprint:

Fitzgerald, D. *et al.* A single-cell multi-omic and spatial atlas of B-cell lymphomas reveals differentiation drives intratumor heterogeneity. *bioRxiv* 2023.11.06.565756 (2023) doi:10.1101/2023.11.06.565756