Dissertation

submitted to the

Combined Faculty of Natural Sciences and Mathematics of the
Ruperto Carola University Heidelberg, Germany
for the degree of
Doctor of Natural Sciences

Presented by
M.Sc. Abdelrahman Mahmoud Aly Mohamed
Born in Riyadh, KSA

Oral examination 28th of January 2021

Deciphering the Immune Evolution Landscape of Multiple Myeloma Long-Term Survivors Using Single Cell Genomics

Referees:

Prof. Dr. Benedikt Brors
Prof. Dr. Karsten Rippe

To my father …

Abstract

Multiple myeloma (MM) is a malignant bone marrow (BM) disease characterized by somatic hypermutation and DNA damage in plasma cells; leading to the overproduction of dysfunctional malignant myeloma cells. Accumulation of myeloma cells has direct and indirect effects on the BM and other organs. Despite the development of new therapeutic options; MM remains incurable and only a small fraction of patients experiences long-term survival (LTS). The past has shown that ultimately all patients still relapse; leading to the hypothesis that a state of active immune-surveillance is required to control the residual disease.

To understand the long-term survival phenomenon and its link to the immune-phenotypes in MM disease; we collected paired bone marrow samples from 24 patients who survived for about 7 to 17 years after Autologous Stem Cell Transplant (ASCT), with a high plasma cell infiltration in the BM (median 49.5%) at diagnosis time. Response assessment according to the International Myeloma Working Group (IMWG) revealed that 15 patients were in complete remission (CR), whereas 9 patients were in non-complete remission (non-CR) that had tumor cells which remained stable over recent years.

We performed single-cell RNA-seq sequencing on more than 290,000 bone marrow cells from 11 patients before treatment (BT) and in LTS, as well as three healthy controls using 10x Genomics technology. I developed a computational approach using the state-of-the-art single cell methods, statistical inference and machine learning models to decipher the bone marrow immune cell types and states across all clinical groups. I performed in-depth analyses of the bone marrow immune microenvironment across all captured cell types, and provided the global landscape of cellular states across all clinical groups.

In this work, I defined new cellular states, marker genes, and gene signatures associated with the patients' clinical and survival states. Additionally, I defined a new myeloid population termed Myeloma-associated Neutrophils (MAN) cells and a T cell exhaustion population termed Aberrant Memory Cytotoxic (AMC) CD8$^+$ T cells in newly diagnosed Multiple Myeloma patients.

Moreover, I propose new therapeutic targets CXCR3 and NR4A2 in AMC CD8$^+$ T cells, which could be further investigated to reverse the T cell exhaustion state in newly diagnosed MM patients. Furthermore, I defined new prognostic markers in the CD8$^+$ T cell compartment which could be predictive for the global disease state.

Finally, I propose that MM long-term survivors go through a complex and evolving immune landscape and acquire cellular states in a stepwise manner. Furthermore, I propose the Continuum Immune Landscape (CIL) Model which explains the immune landscape of MM patients before and after long-term survival. Additionally, I introduced the Disease-State Trajectories (DST) hypothesis regarding the disease-associated dysregulated cellular states in MM context, which could be generalized into other tumor entities and diseases.

Zusammenfassung

Das multiple Myelom (MM) ist eine maligne Erkrankung des Knochenmarks (KM), die durch somatische Hypermutation und DNA-Schäden in Plasmazellen gekennzeichnet ist. Dadurch kommt es zu einer massiven Proliferation maligner Myelomzellen, die direkte und indirekte Auswirkungen auf das Knochenmark und andere Organe hat. Obwohl die Entwicklung neuer Therapeutika neue Behandlungsmöglichkeiten geschaffen hat, bleibt das multiple Myelom eine unheilbare Erkrankung, die nur wenige Patienten langzeitig überleben lässt (long term survival, LTS). Fast alle Patienten erleiden letztendlich einen Rückfall. Das lässt vermuten, dass das Immunsystem beim Langzeitüberleben eine Resterkrankung kontrollieren muss.

Um zu verstehen, warum es zu LTS kommt und welche Rolle dabei das Immunsystem spielt, haben wir gepaarte KM-Proben von 24 Patienten untersucht, die 7-17 Jahre nach einer autologen Stammzelltransplantation (ASCT) noch lebten. Die mediane Plasmazellinfiltration im Knochenmark betrug 49,5% bei Diagnose. Die Beurteilung des Therapieansprechens nach den Kriterien der International Myeloma Working Group (IMWG) ergab, dass 15 Patienten zum Zeitpunkt der zweiten Probennahme in vollständiger Remission (CR) waren, während 9 Patienten nur eine unvollständige Remission (non-CR) hatten, d.h. bei ihnen konnte ein Anteil noch vorhandener Tumorzellen nachgewiesen werden, der jedoch über die Zeit konstant war.

Von 11 Patienten wurden die Proben vor Therapiebeginn und bei LTS mit Einzelzell-RNA-Sequenzierung (10x Genomics) von insgesamt mehr als 290.000 Knochenmarkszellen untersucht, außerdem KM-Proben von gesunden Spendern. Ich habe hier ein computergestütztes Verfahren auf der Grundlage neuester bioinformatischer Einzelzellanalysemethoden entwickelt,

wobei auch Methoden der statistischen Inferenz und des Maschinenlernens zum Einsatz kamen. Damit konnte ich Zelltypen und ihre Zustände in den jeweiligen klinischen Gruppen bestimmen und eine detaillierte Analyse der Zusammensetzung der Immunumgebung im KM durchführen. So konnte eine globale Landschaft der Zelltypen und Zellzustände beim multiplen Myelom beschrieben werden.

Ich habe in dieser Arbeit neue zelluläre Zustände, charakteristische Gene und Gensignaturen beschrieben, die mit den klinischen Zuständen der Patienten und ihrem Überleben assoziiert sind. Ich habe eine neue Population myeloider Zellen, Myeloma-assoziierte Neutrophile (MAN) identifiziert und eine T-Zell-Population mit Anzeichen von *Exhaustion* gefunden, die wir "aberrante cytotoxische CD8$^+$-T-Gedächtniszellen" (AMC) genannt haben.

Mit CXCR3 und NR4A2 schlage ich neue therapeutische Ziele in AMC CD8$^+$ T-Zellen vor, deren Rolle bei der Revertierung eines *exhausted* Zustands weiterer Erforschung bedarf. Außerdem habe ich neue prognostische Marker in CD8$^+$-T-Zellen identifiziert, die den Status der Erkrankung vorhersagen lassen.

Schließlich schlage ich ein Modell vor, bei dem MM-Langzeitüberlebende eine komplexe Abfolge von Veränderungen der Immunzelllandschaft durchmachen, bei dem zelluläre Zustände sich schrittweise ändern. Dieses Continuum-Immunzelllandschafts-Modell (CIL) erklärt die Immunantwort von MM-Patienten vor Therapie und in LTS. Ich entwickele auch eine Hypothese zu "Krankheitszustands-Trajektorien" (DST), die dysregulierte Zellzustände mit potentieller Relevanz auch für andere Kerbserkrankungen beschreibt.

Contributions

- **Sample collection, 10x Genomics library preparation, and wet lab experiments (FACS sorting and qPCR) were performed by other research groups:**
- Dr. med. Raphael Lutz, Mohamed H. S. Awwad, Prof. Dr. med. Michael Hundemer

  Laboratory for Translational Immunology (TRIM), Heidelberg University Hospital, Heidelberg - Germany
- Dr. Jan-Philipp Mallm

  Single-cell Open Lab (scOpenLab) - The German Cancer Research Center (DKFZ), Heidelberg - Germany

- **Next generation sequencing (NGS):**
- Genomics and Proteomics Core Facility - The German Cancer Research Center (DKFZ), Heidelberg - Germany

- **Data transfer and storage:**
- Omics IT and Data Management Core Facility (ODCF) - The German Cancer Research Center (DKFZ), Heidelberg - Germany

- **Bone marrow cell types marker genes:**
- Dr. Simon Haas

  Heidelberg Institute for Stem Cell Technology and Experimental Medicine (HI-STEM gGmbH) - The German Cancer Research Center (DKFZ), Heidelberg - Germany

- **All bioinformatics analyses were performed by:**
- Abdelrahman Mahmoud Aly Mohamed

  Division Applied Bioinformatics (ABI) - The German Cancer Research Center (DKFZ), Heidelberg - Germany

Contents

6

List of Figures

14

List of Tables

1 Introduction

1.1   The discovery of cells: the building block of the biological systems

From an early time point in history, humans have been trying to understand the vast diversity of biological systems in the surrounding nature; as a crucial step to understand health and disease. Back in 3000 B.C., ancient Egyptians started to describe a manifestation of tissue abnormalities and distinguished between benign and malignant "tumors" (Hajdu 2011, 1). Thousands of years later would follow without real progress in understanding the underlying cause of such abnormalities.

In 1665, Robert Hooke used a compound microscope **(Figure 1.1)** and started to observe tiny pores that looked like tiny rooms ("Cella" in Latin) which he named "cells" (Hooke 1667). He did not know what their real function is and did not think that they are alive (Inwood 2002, 1635–1703). By using more powerful microscopes (magnify objects ~ 300-fold), Anton van Leeuwenhoek started to observe that these cells are motile and he assumed that this motility is indicative of life (Mazzarello 1999). Over years of technological advancement, we started to gain more information about subcellular structures and cellular functions, and more evidence that the cell is the fundamental unit of life (Mazzarello 1999). Such a new understanding was the start of shaping one of the key ideas in the biological history "The cell theory".

The discovery of the cells changed our perspective on the underlying structure of tissues and organs. We started to understand that alterations in the cells of an organism would have a causal link to the diseases and started to catalog human diseases accordingly. It seems that the interaction between the human mind, tools, and technologies induced the next steps in understanding the

underlying chain of causes and effects of the physical world. That was a true case in shaping "The cell theory". These concepts helped Rudolf Virchow to develop the *Cellularpathologie* concept and to further propose "the theory of cancer origin" (R. P. Wagner 1999).
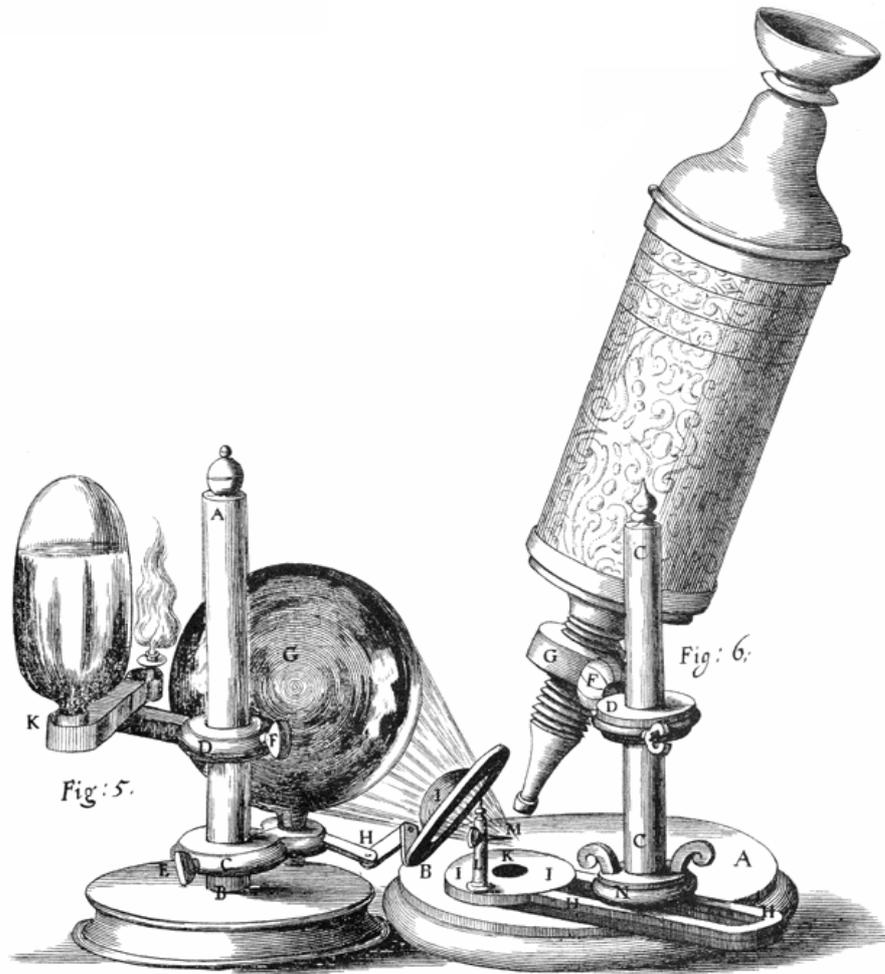
Figure 1.1: Early model of Robert Hooke's microscope.

----

*Adapted from (Micrographia by Robert Hooke, 1665)

1.2 The origin of the immune system:

Humans live in the natural world which is full of microbial communities and are in need to interact with such complex communities, either to eliminate the pathogenic ones or to keep a balanced relationship with the beneficial ones (D. Zheng, Liwinski, and Elinav 2020). Moreover, human cells tend to accumulate somatic mutations and cellular defects which need to be quickly eliminated (Goodnow 2007); to get rid of such potential pathogenic cells and to keep the whole organ-to-organ interaction systems in a healthy homeostatic state. To execute these crucial functions, humans and other mammals developed an intricate and complex immune system that keeps regulating and checking the cellular interactions in both homeostatic and diseased states; in addition to eliminating the pathogenic antigens.

Multicellular organisms (metazoans) started to evolve and exist in 500 million years ago. These organisms started to co-develop intricate systems of adaptive immune cells beside the innate immune cells to initiate protective responses against potential pathogens; including parasites, fungi, bacteria, and viruses (Cooper and Alder 2006). Over millions of years, these multicellular organisms started to adapt to different environmental challenges and stresses to survive such conditions in the natural world **(Figure 1.2).** These defensive mechanisms are mediated via a complex network of interacting cell types which evolved strategies to learn from the environment and the inner cellular world, and to elicit appropriate immune responses accordingly.

Over the last couple of decades, we started to gain deeper insights into the immune system and how it evolved to work in such an orchestrated manner. Moreover, we learned that agnathans (jawless vertebrates) developed an adaptive immune system based on a combinatorial assembly system of different genetic modules to produce lymphocytes with a unique receptor repertoire (Alder et al. 2005).
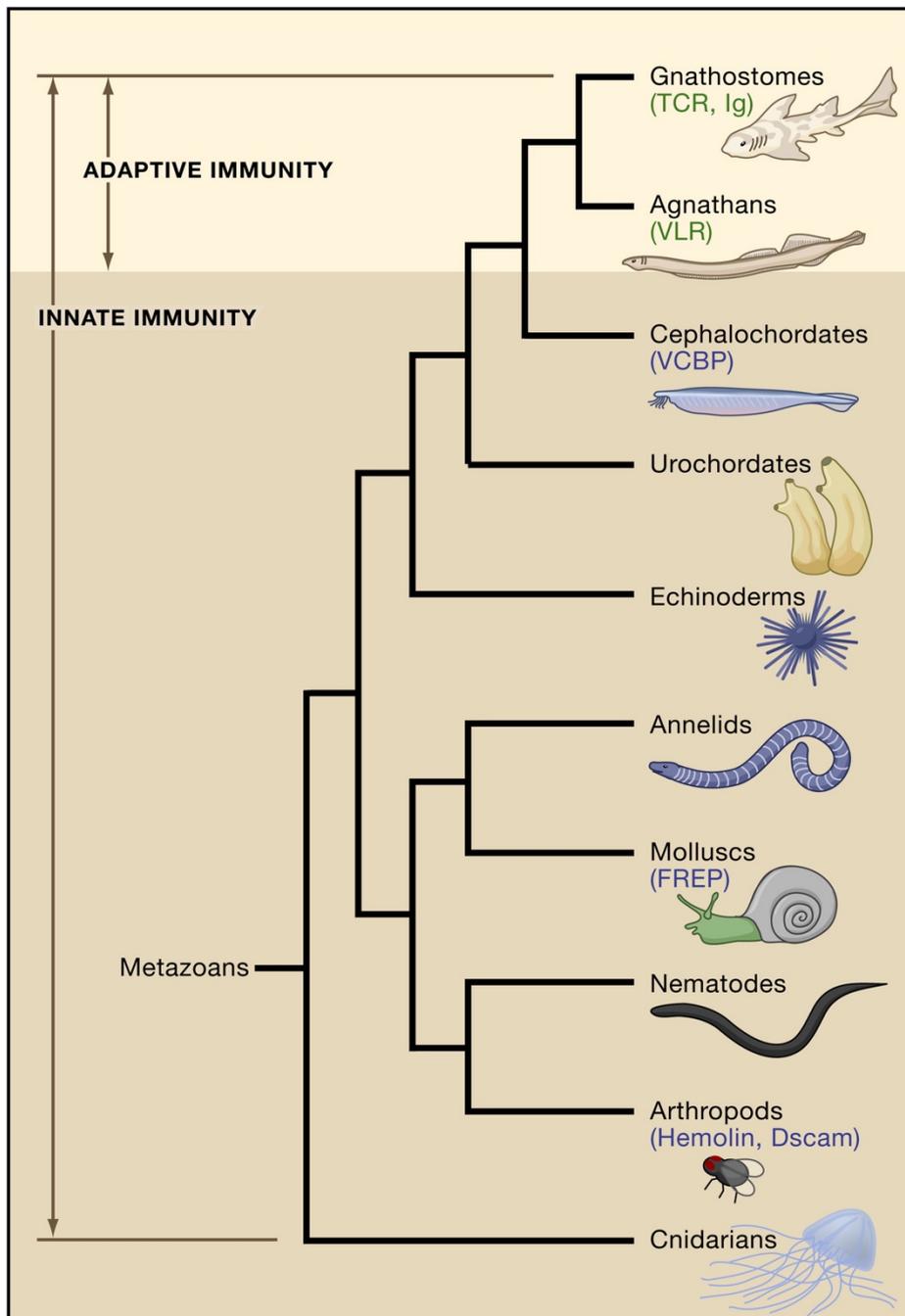
Figure 1.2: Phylogenetic Tree represents the evolutionary tree of the multicellular organisms' immune system.

----

*Adapted from (Cooper and Alder 2006) by permission from Copyright Clearance Center's RightsLink® service.

1.2.1 Hematopoietic system and hematopoietic stem cells (HSCs)

All immune cells originate from hematopoietic stem cells (HSCs) through a complex developmental process known as "hematopoiesis", which is localized mainly in the bone marrow. In 1868, Ernst Haeckel was the first to use the word stem cell (*Stammzelle)*. He thought that these stem cells are the root of a branching family of more mature cells in many biological systems which he described as a *tree-like* structure and he named it a *"family tree"* or '*Stammbaum'* in German (*Ernst Haeckel* 1868). It seems that the early description of such developmental processes as a *"Tree"* like model dominated our view of the hematopoietic system description for decades.

1.2.2 Hierarchical models of hematopoiesis and cell-fate decisions

Early definitions of the HSCs came from fluorescence-activated cell (FACs) sorting of the HSC populations (Spangrude, Heimfeld, and Weissman 1988) which are characterized by two essential properties after transplantation; self-renewal and multipotency. The non-HSCs populations (Progenitor cells) are the cells that can not retain such self-renewal properties and have a restricted lineage differentiation capacity (Doulatov et al. 2010).

Over the past twenty years, the scientific community derived many models to explain the differentiation and development process of HSCs from the progenitor states towards the final commitment to mature cell types. Early models started to separate the progenitors into two subbranches; the myeloid branch (common myeloid progenitors "CMP") and lymphoid branch (common lymphoid progenitor "CLM"). They further split into megakaryocyte–erythroid progenitor cell "MEP" and granulocyte–monocyte progenitors "GMP" (**Figure 1.3- a**).
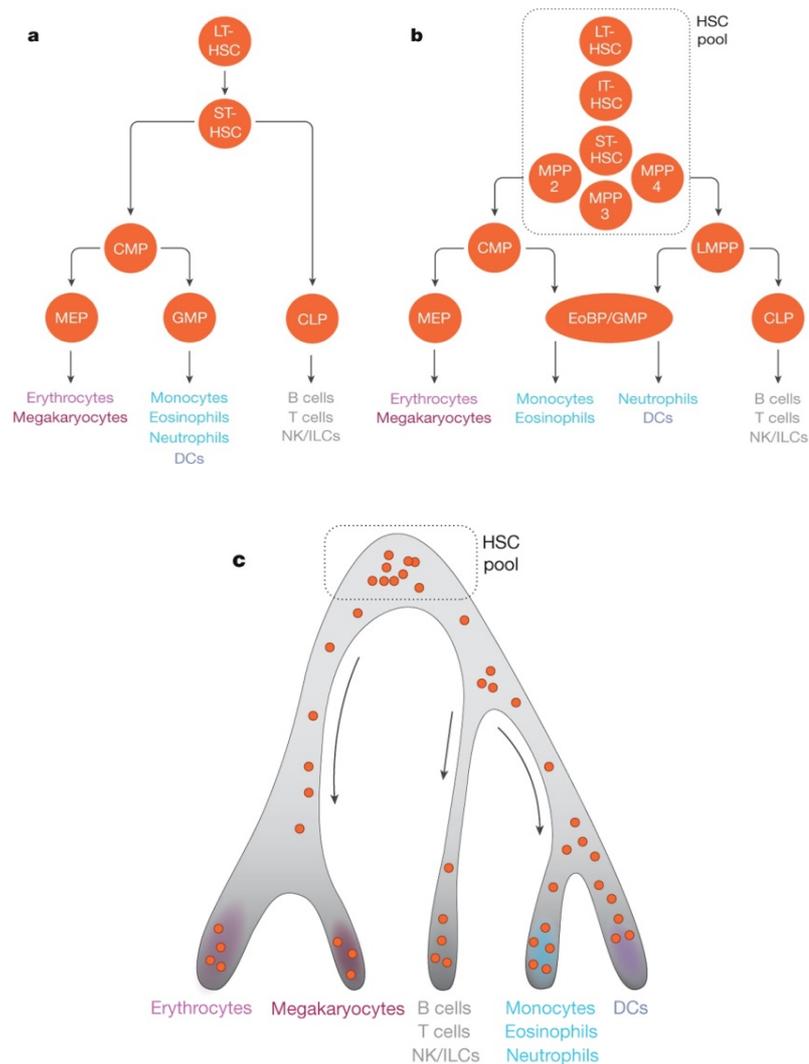
Figure 1.3: Proposed Haematopoiesis hierarchical models.

a) Early model shows that HSCs population is a homogeneous state of cells and through the differentiation process, it bifurcates into myeloid and lymphoid branches towards the final mature cell type. **b)** This model shows that the HSC pool is a more heterogeneous population. **c)** This model has been proposed after using single-cell transcriptomics technologies, which suggest that the haematopoiesis process is rather a *continuum* of cellular differentiation.

Around the time between 2005 and 2015, new data and pieces of evidence became available suggesting that HSCs are heterogeneous pools of cells that have different capacities of self-renewal and differentiation properties. The myeloid and lymphoid branches are still connected via the lymphoid-primed multipotential progenitor (LMPP) population (Görgens et al. 2013) (**Figure 1.3 - b**).

Due to the recent development in single-cell genomics technologies and especially in single-cell RNA-seq, it has been proposed that the hematopoiesis process is rather a *continuum* process of cellular differentiation where single cells can develop along a differentiation trajectory and that lineage-restrictions are regulated in an early time point of hematopoietic stem and progenitor cells (HSPCs) compartment (**Figure 1.3-c**) (Paul et al. 2015, Velten et al. 2017 and Giladi et al. 2018).

Cell-fate decisions are tightly controlled via gene expression programs, which are in turn regulated via transcriptomic and epigenetic regulators. Multipotent cells are proposed to harbor multilineage capacity where they express a low level of expression programs which can activate and shift the cellular function into alternative lineages under either *deterministic* or *stochastic* processes (Laurenti and Göttgens 2018). Such biological programs shape at the end a specific pattern of cell-type-specific related gene expression programs, which are capable of regulating the function, phenotypic state, and cellular identity of the immune cell types.

1.3 The immune system blueprint

Over the past several years, a substantial research effort has been invested to understand the mechanisms that regulate the immune systems' functions. The immune system can be subdivided into two large branches: innate and adaptive immunity. Innate immunity uses an embedded germline-encoded genetic memory of receptors that can directly recognize the molecular pattern of common antigens in a time scale of hours.

In contrast, adaptive immunity co-evolved to recognize unique patterns of antigens to elicit specific immune responses and eliminate these new antigens in a time scale of days **(Figure 1.4)**. The efficiency of the immune system depends mainly on the intricate interplay between the innate and adaptive immune branches to develop strategies; to eliminate potential pathogenic challenges.
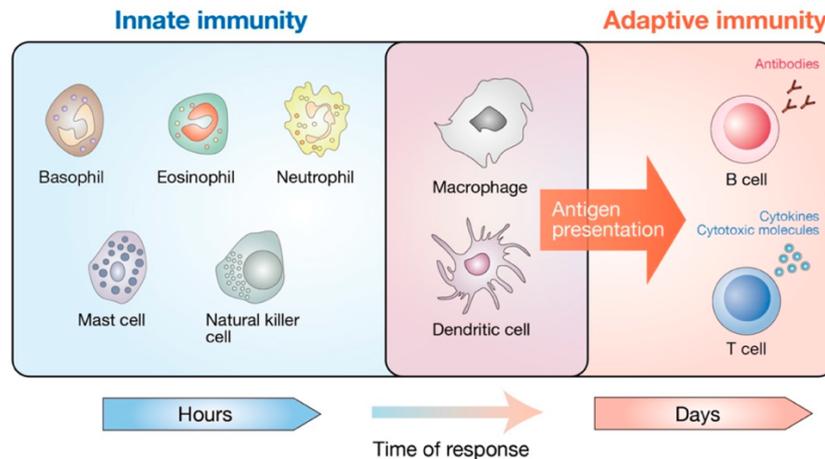
Figure 1.4: Immune system's two branches (innate and adaptive immunity) and the time scale of the immune response.

Through the next sections, we will dive into the key players of the immune system and how it is orchestrated and regulated.

1.3.1 The innate immune system: granulocytes, myeloid Antigen-Presenting (APC), and natural killer (NK) cells

The innate immune system (IIS) can induce immune responses mediated via specialized myeloid and lymphoid cells, and recognize foreign antigens through germline-encoded receptors (e.g: Toll-like receptors (TLRs) and pattern recognition receptors (PRRs) Galli, Borregaard, and Wynn 2011). Granulocytes are considered the frontline of the innate immune response and can be subdivided into many cell types according to the cellular morphology and characteristic staining (e.g., neutrophils, basophils, mast cells or eosinophils).

Neutrophils constitute the majority of leukocytes and are recruited by inflammatory molecules (e.g., chemokines) to the site of inflammation. They can phagocytose antigens, secrete quite a range of cytokines to recruit more neutrophil cells to the inflamed tissue, and potentially remodel the bone marrow to increase the production of myeloid cells. Myeloid progenitors can generate a group of phagocytic cells (dendric cells (DCs), monocytes, and macrophages) that acquire in addition the function of professional antigen-presenting cell (APC).

Myeloid APCs are considered the link between the innate and adaptive immune system as they can detect antigens and communicate that to the T cell and B cells **(Figure 1.4).** DCs have a primary and critical role in capturing and presenting antigens to naïve T cells and activating them (Collin and Bigley 2018). Macrophages also have the capacity of antigen-presenting and are more specialized in engulfing "phagocytose" antigens and pathogens (Mosser and Edwards 2008). More shreds of evidence suggest phenotypic plasticity of

innate immune cells in response to cytokines signals, and other microenvironment factors that can alter their cellular functions (Galli, Borregaard, and Wynn 2011).

Lymphoid progenitors can generate natural killer (NK) cells, which are considered as one of the IIS major players; they have germline-encoded receptors and can mediate cytolytic effector functions, which can directly kill infected cells or tumor cells (Moretta et al. 2008). However, this notion has been challenged. Other views consider NK cells as part of the adaptive immune system since they originate from the lymphoid lineage and can contribute to the "immunological memory" phenotype which is mainly associated with the T- and B- cells **(Figure 1.5)**. There are distinct receptor-ligand interactions which induce generating antigen-specific memory NK cells that have certain tissue restriction and adaptation patterns (Cerwenka and Lanier 2016). Therefore, NK cells can be considered as the interface between the innate and adaptive immune system (Moretta et al. 2008 and Vivier et al. 2011).
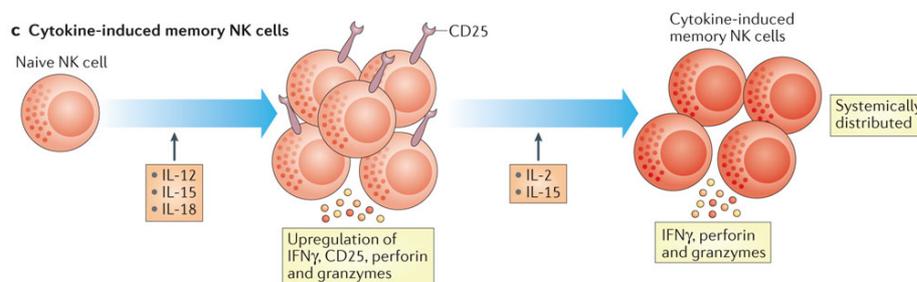
Figure 1.5: The generation of memory NK cells

In-vivo exposure of NK cells to IL-18, IL-12, and IL-15 induce a NK activation state and after the adoptive transfer of these cells; they can persist for the long term and produce abundant granzymes, perforin, and IFNγ.

1.3.2 The adaptive immune system: T- and B- cells

Throughout the evolutionary history of the immune system, vertebrates evolved a new branch of the immune system, which can generate new and diverse receptors against foreign antigens, in addition to the preexisting innate immune system. This new machinery is termed the "adaptive" immune system where each lymphocyte has a unique antigen receptor. This feature makes the adaptive immune system capable of generating a clonally diverse repertoire of lymphocytes; facilitating an evolutionary advantage for recognizing a wide range of new antigens and pathogens.

By the early 20th century, a new realization started to emerge in understanding that the adaptive immune system has two main functional branches, one branch is involved in cell-mediated immunity and the other branch is capable of generating soluble immune factors or humoral-mediated immunity (antitoxins or antibodies).

The key cellular players of these immune functions have not been discovered until the late 1950s. Two fundamental studies started to change the immunological thinking at this time and defined the key cell types mediating the adaptive immune system functions **(Figure 1.6)**.

In Miller 1961's study, he performed thymectomies in neonatal and older mice in the context of studying thymic lymphoma, and he observed that only mice that had undergone neonatal thymectomy developed a *wasting syndrome* in addition to infections and low proportions of lymphocytes. Miller proposed that in early life the thymus export lymphocytes which can support graft-rejection or antiviral responses.

A few years later, in Cooper, Peterson, and Good (1965)'s study, they showed that the chickens' bursa of Fabricius is the source of antibody-producing

lymphocytes, and they distinguished between the role of these cells and Miller's thymus-derived cells (cell-mediated lymphocytes).

Both studies established the key differences between the T (thymus-derived) and B (bone marrow-derived or bursa-derived) lymphocytes and their co-operative role in mediating adaptive immune functions **(Figure 1.6)**.
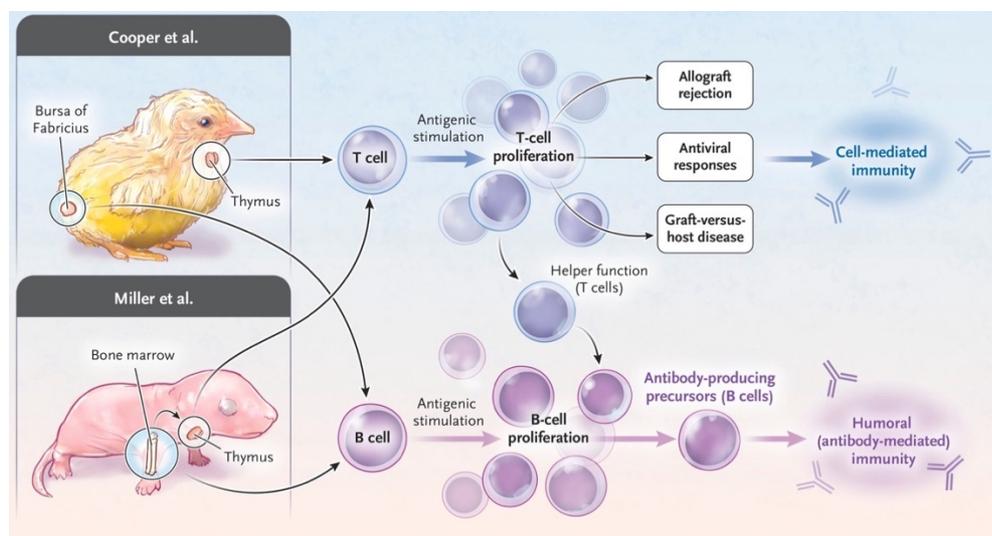
Figure 1.6: Early adaptive immune system models show the origin of cell-mediated immunity by the T cell (thymus-derived), and humoral-mediated immunity by the B cell (bursa-derived or bone marrow-derived).

----

*Adapted from (Germain 2019) by permission from Copyright Clearance Center's RightsLink® service.

1.3.3 T lymphopoiesis: T cell thymic development

T cells are the key players of cell-mediated immunity, they go through complex processes of development and differentiation to be fully functional cells. T cell development starts from HSC differentiation into lymphoid restricted progenitors which can give rise to thymic seeding progenitors (TSPs). The exact origin of the TSPs is debatable and further pieces of evidence suggest that TSPs originate from either fetal liver or bone marrow (Koch and Radtke 2011, Kernfeld et al. 2018, Zeng et al. 2019).

The thymus can be subdivided into four compartments (the subcapsular zone, cortex, medulla, and the corticomedullary junction), each of which has a unique microenvironment structure and has important roles in T-lymphopoiesis. TSPs start to enter the thymus and develop into early thymic progenitors (ETP); a subpopulation of double-negative (DN) 1 cell retains the capacity to develop into DCs, NK cells , and macrophages **(Figure 1.7)**.

In the DN1 stage, NOTCH signaling inhibits cell fate potential and allows it to develop into the DN2 stage where the TCR γδ and TCR β rearrangement occurs and can give rise to committed γδ T cells. Furthermore, the DN2 is developed into DN3 and DN4, which commit into double-positive T cells by further retaining TCR αβ lineage restriction. Finally, the DP T cells commit to a single positive (SP) state; either CD4 SP or CD8 SP T cells (Koch and Radtke 2011).

CD8$^+$ T cells have a crucial role in mediating cytotoxicity functions against antigens, pathogens, and tumors through releasing granzymes, perforin, and cytotoxic molecules. CD4$^+$ T cells have lower cytotoxicity functions and regulatory roles (Th1, Th2, Th17, and others), in addition to immunosuppressive roles (T$_{reg}$) mediated by binding to effector T cells and inhibiting the secretion of cytokines.
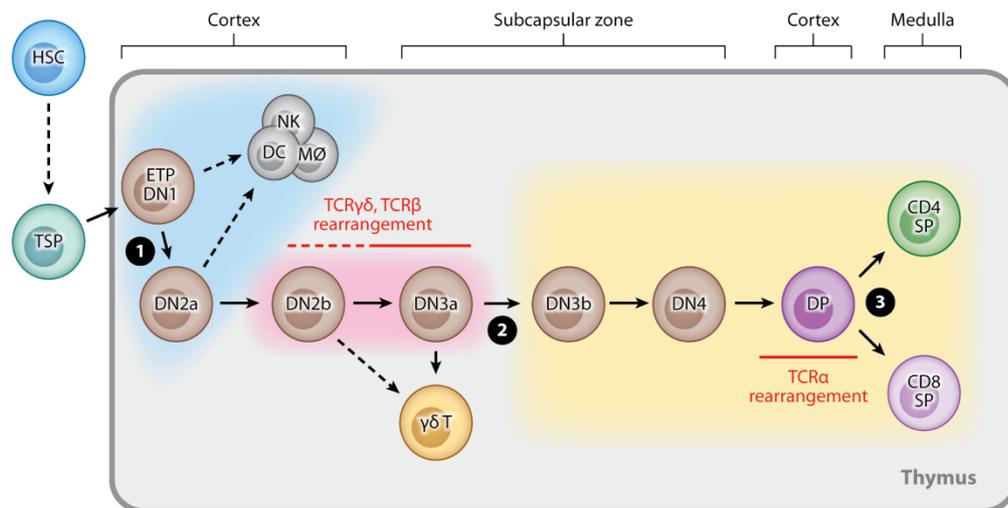
Figure 1.7: T cell thymic development.

HSC cells develop to TSPs which can enter the thymus cortex and go through multistep development and maturation into a single positive (SP) state either being SP CD4 or SP CD8 T cells.

----

*Adapted from (Koch and Radtke 2011) by permission from Copyright Clearance Center's RightsLink® service.

1.3.4 T cell-fate differentiation models

During acute and chronic infections, naïve CD8$^+$ T cells bind to specific antigens and undergo clonal expansion and differentiation into effector T cells, which could either be terminal effectors (effector cytotoxic) or memory cells (memory precursor effector) to further develop into long-lived memory T cells. In the chronic infection setting, the effector and memory CD8$^+$ T cells encounter persistent antigen exposure which leads to the emergence of a terminal nonfunctional state of differentiation that is defined as the "T cell exhaustion" state.

Many models have been proposed describing the T cell-fate differentiation processes and the generation of effector and memory T cells, which are tightly regulated via transcriptional and epigenetic regulators (Kaech and Cui 2012, Zebley, Gottschalk, and Youngblood 2020).

The first model is the "separate-precursor" model, which describes that naïve T cells are preprogrammed during thymic development to adopt a specific differentiation state either to effector T cells or long-lived memory T cells (Kaech and Cui 2012b). However, this model has not been supported anymore since several studies using cellular barcoding showed that single naïve T cells are multipotent and can differentiate to both memory and effector T cells (Stemberger et al. 2007, Gerlach et al. 2010) **(Figure 1.8 a)**.

The "decreasing-potential" model suggests that the differentiation states are regulated according to the repetitive stimulation and cumulative history of the signals encountered in the T cell history of antigen stimulation to the T cells **(Figure 1.8 b)**. The "signal-strength" model suggests that the heterogeneity in differentiated T cells is a result of the overall and combined signals of antigens, cytokines, and costimulatory molecules **(Figure 1.8 c)**.
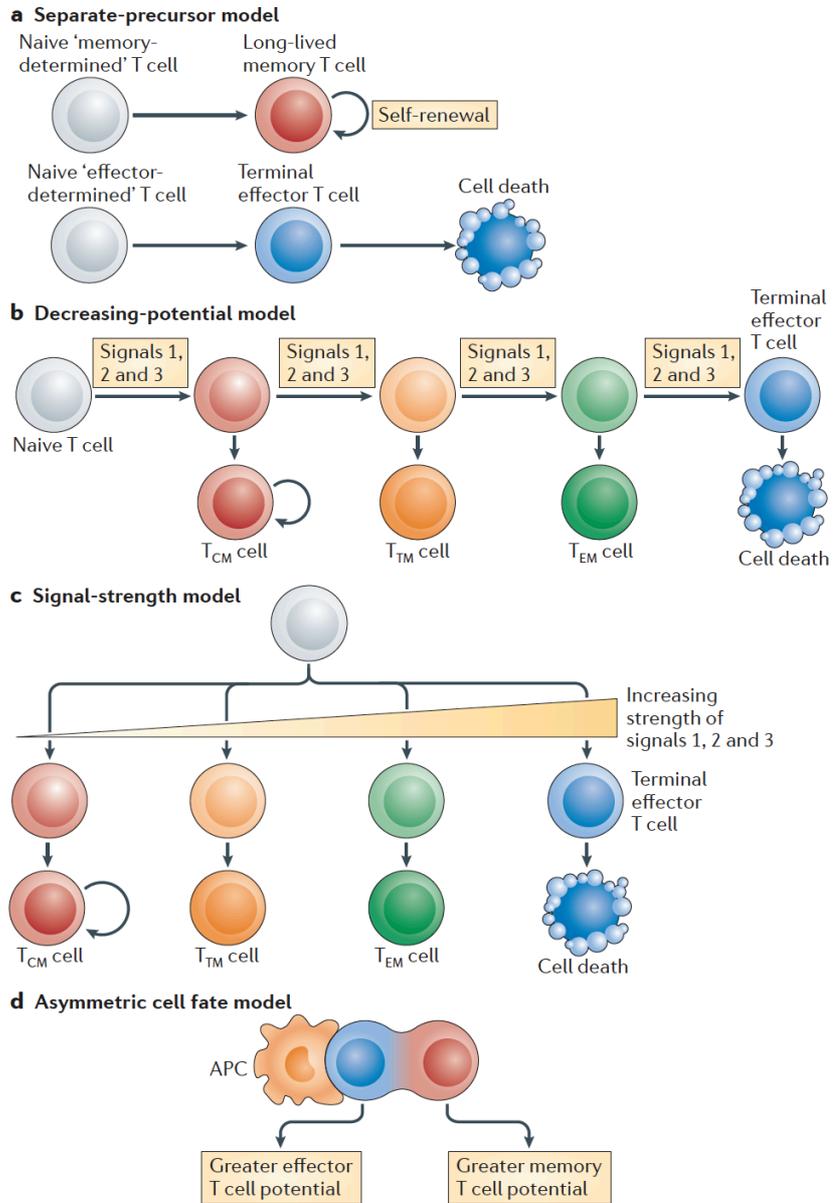
Figure 1.8: T cell-fate differentiation models describing potential mechanisms of generating effector and memory T cell states.

----

*Adapted from (Kaech and Cui 2012a) by permission from Copyright Clearance Center's RightsLink® service.

Finally, the "asymmetric cell fate" model proposes that both memory and effector T cells can arise from one cell through an asymmetric process of cell division that can happen at very early steps of cell division after the T cell activation via an antigen (Kaech and Cui 2012b) **(Figure 1.8 d).**

Understanding T cell-fate differentiation states and the regulating mechanisms of this process are crucial for unleashing T cell cytotoxic functions against tumor cells and pathogens (Zebley, Gottschalk, and Youngblood 2020).

1.3.5 B-cell development and immunological memory formation

After a single infection, the immune system acquires two key aspects of an immunologic memory: **1)** the long-lived plasma cell, which is responsible for the antibody production. **2)** memory B-cells that can be activated and act after re-exposure to the same infectious agent. One key aspect to understand the immunologic memory of the B-cell is to understand the underlying developmental processes of the B-cells and cell-fate decisions after exposure to an antigen (Akkaya, Kwak, and Pierce 2020).

B- cell originates from the HSC compartment in the bone marrow where the HSCs differentiate into Pro-B cell and Pre-B cell states **(Figure 1.9 a)**. The next developmental step is the generation of the immature B-cell, which acquires the expression of the B-cell receptor (BCR) which "*marks*" the first specificity stage in B- cell development. Furthermore, these cells enter the periphery as a "transitional B-cell" which expresses IgM with or without IgD and would be ready to enter the second stage of B-cell development in the secondary lymphoid organs (Cambier et al. 2007). After exposure to an antigen, the naïve B-cell can further develop to a memory B- cell or a plasmablast and further develop to the final stage of the plasma cell state.
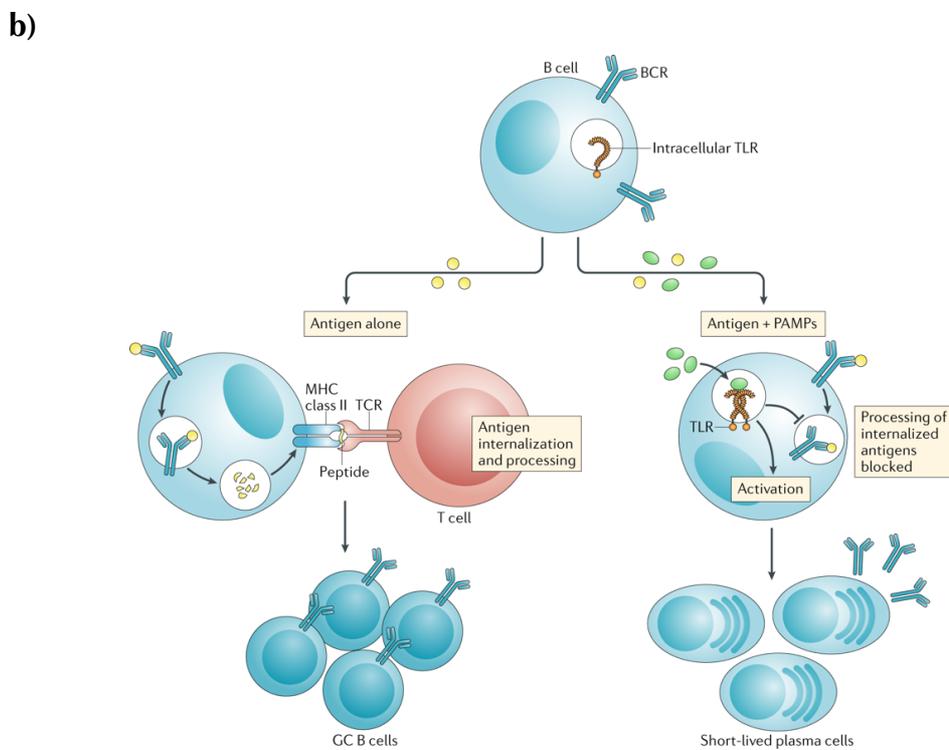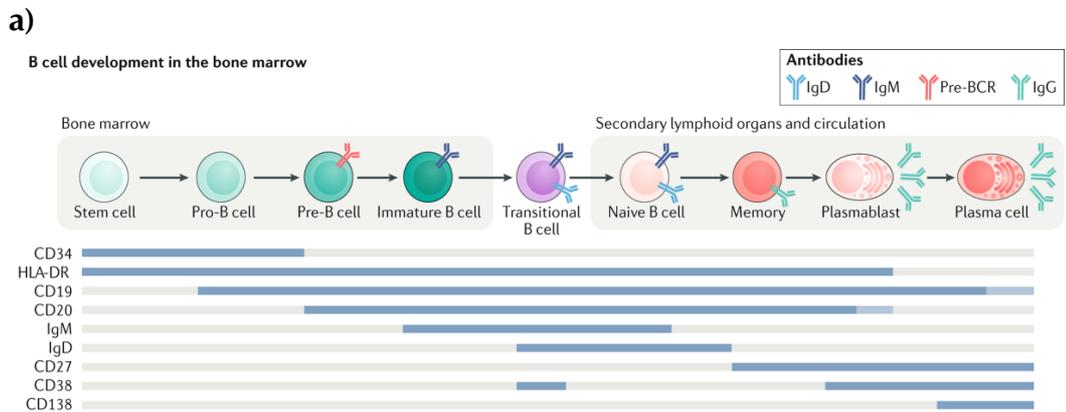
Figure 1.9: B cell developmental models.

a) B cell development in the BM and secondary lymphoid organs. **b)** Naïve B cell-fate and differentiation decisions according to the pathogenic exposure.

Naïve B- cells express both TLRs and BCRs; the TLRs role is to recognize pathogen-associated molecular patterns (PAMPS), which are conserved molecular motifs in microbes. BCRs can recognize new antigens. In the case of only recognizing antigens, the B-cells perform antigen presentation of the antigen to the T cells and this process induces the production of germinal center (GC) B cells **(Figure 1.9 b)**.

In contrast, B- cell's exposure to both antigens and PAMPs activates the TLRs to block the processing of the antigen via BCRs and induces a shift in the differentiation machinery toward producing short-lived plasma cells (Akkaya, Kwak, and Pierce 2020).

1.4   Multiple Myeloma (MM) disease

Multiple myeloma (MM) is a cancer of malignant plasma cells in the bone marrow (BM) microenvironment. MM represents about 10% of all hematological malignancies (Rajkumar 2020). Normal plasma cells have a unique role in generating diverse antibody structures which are accomplished by V(D)J recombination, somatic hypermutation, and class-switch recombination (Chi, Li, and Qiu 2020).

This mechanism of antibody generation is dysregulated, which leads to malignant transformation of plasma cells and the manifestation of many subtypes of hematological malignancies including MM, which can affect and dysregulate the microenvironment (Rajkumar et al. 2014). The overproduction of malignant plasma cells leads to disturbing normal bone marrow functions and the presence of clinical signs and manifestations like anemia, low platelet counts, and low white blood counts. One key characteristic of malignant myeloma cells is that they can produce abnormal immunoglobulins (M-proteins) which can be present in the bloodstream and or the urine.

a)
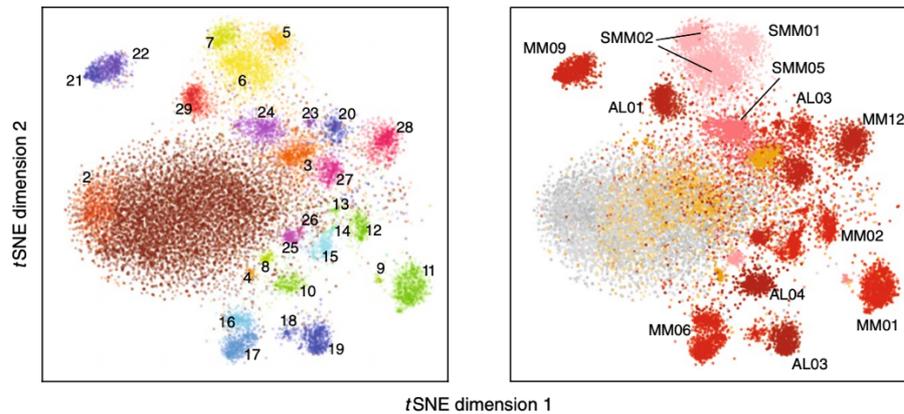
b)
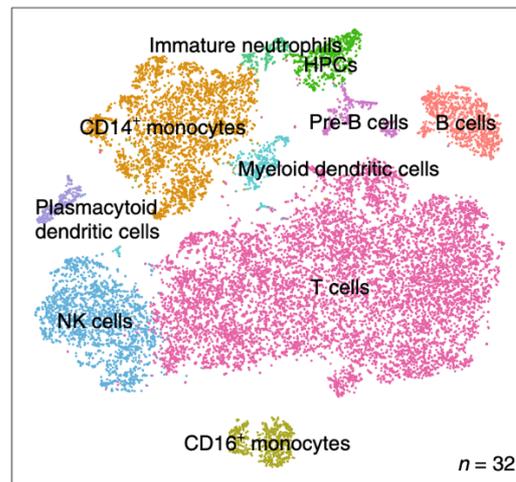
Figure 1.10: Single-cell genomics studies show the heterogeneity of the tumor cell compartment and bone marrow microenvironment cell types in the MM context.

a) t-distributed stochastic neighbor embedding (tSNE) representation shows the heterogeneity of the malignant plasma cell compartment between different patients' clusters. **b)** tSNE representation shows the diverse cell types and states in MM bone marrow microenvironment.

Furthermore, the M- proteins can adhere together and induce organ failure and cause hyperviscosity syndrome, as well as other clinical manifestations: hypercalcemia, **r**enal insufficiency, **a**nemia, and/or **b**one disease with lytic lesions. All these signs are known as CRAB clinical features (Kumar et al. 2017). Recent genomics and single-cell studies showed that malignant myeloma cells are composed of heterogeneous populations of malignant tumor cells where every single patient has a certain expression pattern of transcriptional programs (Ledergor et al. 2018). These *"transcriptional clones"* could instruct bone marrow changes and remodeling, and induce a cellular shift in the surrounding of the tumor microenvironment immune cells (Zavidij et al. 2020) **(Figure 1.10)**.

1.4.1 MM bone marrow niches and microenvironment

The bone marrow (BM) microenvironment is a very complex living system, which is composed of diverse cell types and states. They are interconnected through a complex chain of chemokine and cytokine signaling networks. The BM niches are constructed from cellular and non-cellular parts, where the cellular part is composed of mesenchymal stem cells (MSCs), osteoclasts, osteocytes, endothelial cells, macrophages, T cells, B cells, fibroblasts, and other cell types **(Figure 1.11 a)**. These cell types regulate and orchestrate the hematopoietic stem and progenitor cell (HSPC) self-renewal capacity in addition to their quiescence state, localization, and differentiation (Ghobrial et al. 2018).

Over the past decades, we started to understand the important role of the bone marrow microenvironment in supporting and mediating malignancy states in human and in vivo mouse models (Medyouf 2017, Tirado-Gonzalez et al. 2018). Early studies showed that stromal cells (e.g., fibroblasts) can establish premetastatic niches which can support malignant clones (Kaplan et al. 2005).
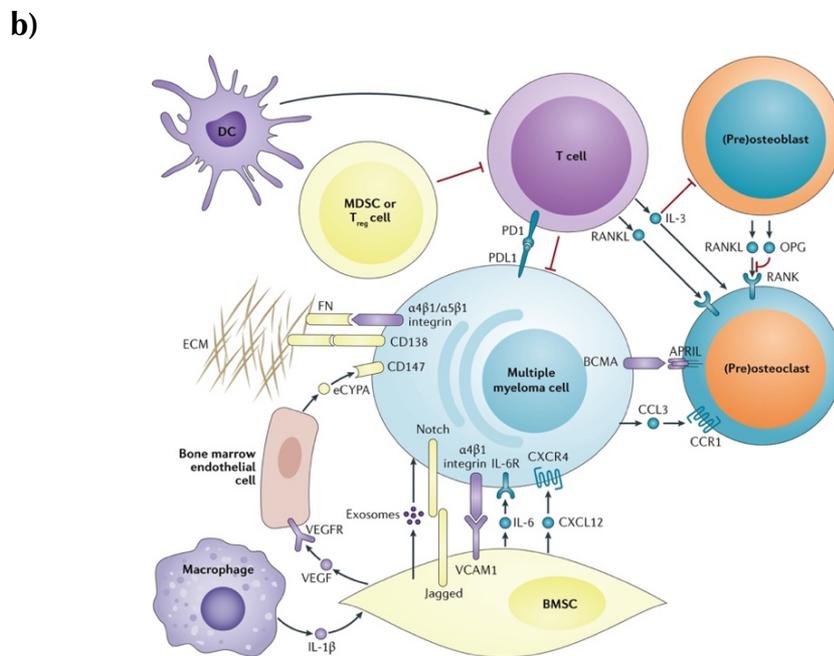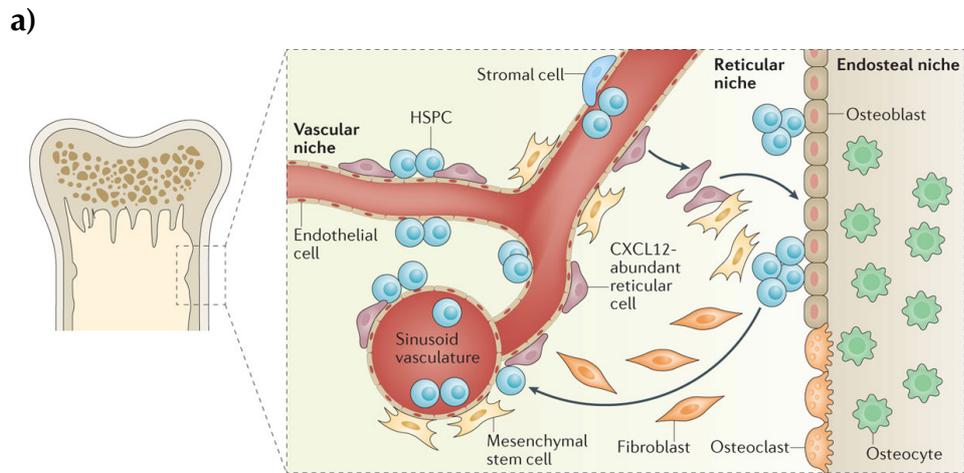
Figure 1.11: Multiple Myeloma BM microenvironment

a) Hematopoietic stem and progenitor cell (HSPCs) co-localization in the BM niche and cell types where cell-cell interactions are maintained via a network of chemokines and cytokines **b)** Cell-cell interaction network between myeloma cells and the bone marrow niche mediated via receptor-ligand interactions.

Studies in the MM context showed that there is a crucial interplay between the myeloma cells and the BM microenvironment (Kumar et al. 2017). Myeloma cells migrate to the BM by upregulating CXCR4 on their cells and bind to CXCL12 expressing cells in the BM niche **(Figure 1.11b)**. The interaction of myeloma cells with BM osteoblast causes high expression of RANKL, which binds to the RANK receptor (an activator of the NF-κB pathway and induces osteoclast differentiation). The imbalance between osteoblast and osteoclast differentiation leads to bone destruction, and the MM patient starts to experience bone disease with lytic lesions, which is one of the CRAB clinical signs of MM disease.

Additionally, the BM stromal cells express VEGFA, which is a strong angiogenic factor that mediates increased oxygen supply by forming new blood vessels. Myeloma cells can express VEGF, which induces angiogenic processes in the BM microenvironment (Vacca and Ribatti 2006). Clinical studies showed that elevated angiogenesis was associated with a worse MM disease outcome (Rajkumar et al. 2002). Still, we are in an early phase of understanding the relationships between the BM niche and malignant myeloma cells.

1.4.2 MM disease evolution trajectories and the emergence of high-risk states

Multiple myeloma is a genetically heterogeneous and complex disease mediated by several genetic and epigenetic changes that induce alterations driving the myelomagenesis processes (Morgan, Walker, and Davies 2012). Pre-clinical stages of MM have been defined; suggesting multistep evolutionary steps till the manifestation of full-blown MM state. The pre-clinical/pre-malignant states are monoclonal gammopathy of undetermined significance (MGUS) and smoldering myeloma (SMM) which are crucial steps in the natural history of the disease evolution.

It has been proposed that these multistep transitions require the acquisition of genetic abnormalities which lead to the development of myeloma malignant clones **(Figure 1.12 a)**. The myelomagenesis process starts with a normal plasma cell that acquires an initial event (e.g., chromosomal translocation t(4;14), t(14;16) and hyperdiploidy) which induces the MGUS state. Furthermore, the transformed clone starts to harbor copy number changes (e.g., gain (1q), deletion (1p), deletion (17p), and mutations) which transform the cells to be in the SMM state. The next stage of transformation is marked by MYC translocation, jumping translocations, and amplification (1q), which leads to subclonal diversification and expansion of certain clone structures and the development of the clinical symptoms of the MM disease.

Furthermore, there are major ecosystems and regional evolution resulting in (selective-sweeps) based on the competitive features of the malignant clones. These malignant clones access different niches and lead to creating focal lesions (FLs) in the BM, extramedullary disease (EMD) in other organs (e.g. liver) or plasma cell leukemia (PCL); which represent the key features of the high-risk MM state **(Figure 1.12 b)**.

a)
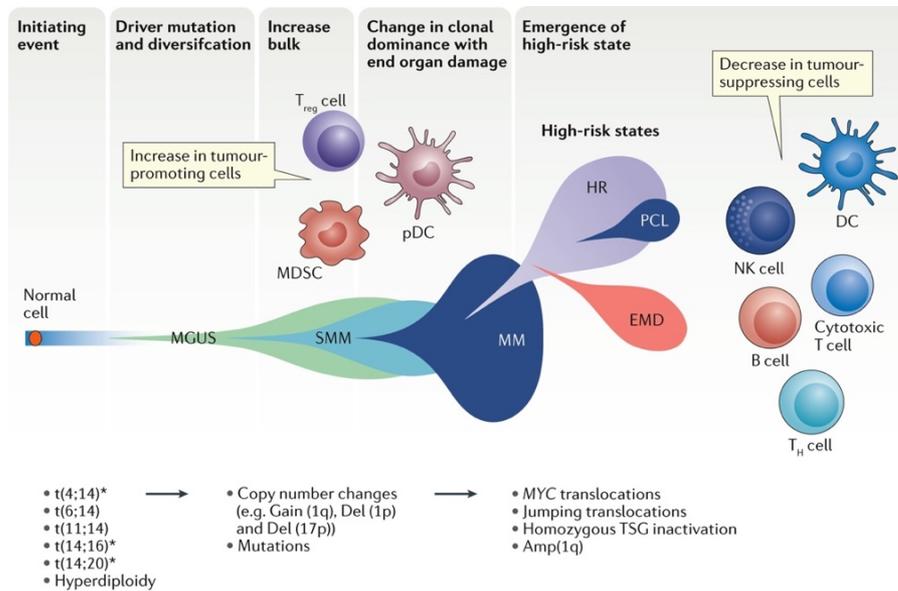
B)
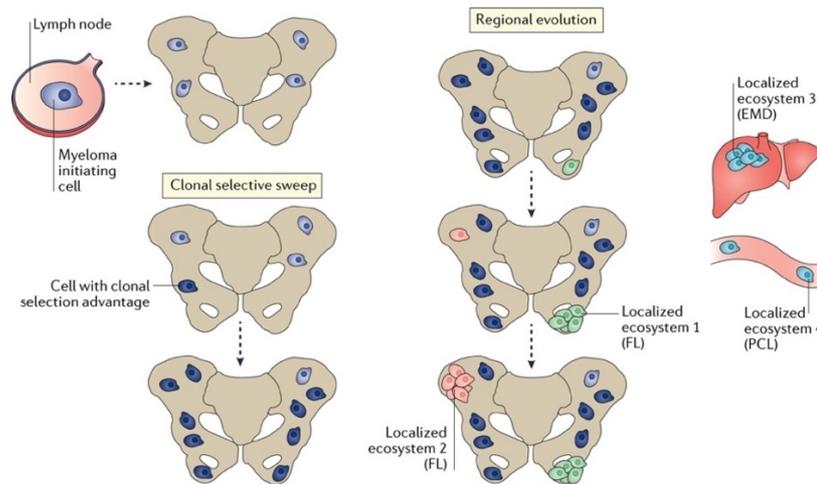
Figure 1.12: MM disease evolution trajectories.

a) Evolutionary model explaining the MM disease evolution from early initial transformation events in the plasma cell compartment to the manifestation of high-risk MM state. **b)** Proposed model for MM regional evolution in the bone marrow and extra-medullary.

----

*Adapted from (Pawlyn and Morgan 2017) by permission from Copyright Clearance Center's RightsLink® service.

1.4.3 The immune microenvironment, immunoevasion, immunosuppression, and T cell exhaustion

Cellular components of the immune system and its two branches (the innate and adaptive immune systems) have an important role and impact in tumor progression and elimination (Gajewski, Schreiber, and Fu 2013). In the normal BM microenvironment, there are cell types capable of deriving cytotoxicity functions and mediating potent immune responses (e.g., NK and T cells) against tumor cells, to orchestrate tumor elimination processes.

However, there are other populations such as myeloid-derived suppressor cells (MDSCs) and regulatory T cells (T_{regs}) that have an immunosuppressive function, which promote malignant clone proliferation and immunoevasion (Taube et al. 2018). Moreover, stromal cells have been shown to activate T_{reg} cells, inhibit B-cell proliferation, and regulate both innate and adaptive immune responses, which can be mediated through the secretion of IL-10, IFNγ, TNFα, and TGFβ (Ghannam et al. 2010).

Recent studies showed that there is a link between MM disease evolution from MGUS to SMM and the immunosuppressive state of the BM microenvironment, which mediates the escape of tumor cells and promotes their growth (Zhang and Zhang 2020). Recently, Zavidij et al. 2020 conducted a single-cell RNA-seq study of the BM aspirates from MM patients at different time points of the clinical and pre-clinical stages of the MM disease **(Figure 1.13).** Zavidij et al. 2020 showed that T_{regs}, NK, and CD16$^+$ monocytes cells were enriched in the MGUS stage, while memory CD8$^+$ T cells in the SMM low-risk stage were depleted. In the high-risk SMM stage, they showed that CD14$^+$ monocytes induce internalization of (MHC) class II, which establishes a T cell suppressive phenotype and increased expression of interferon-α signaling genes in many cell types.
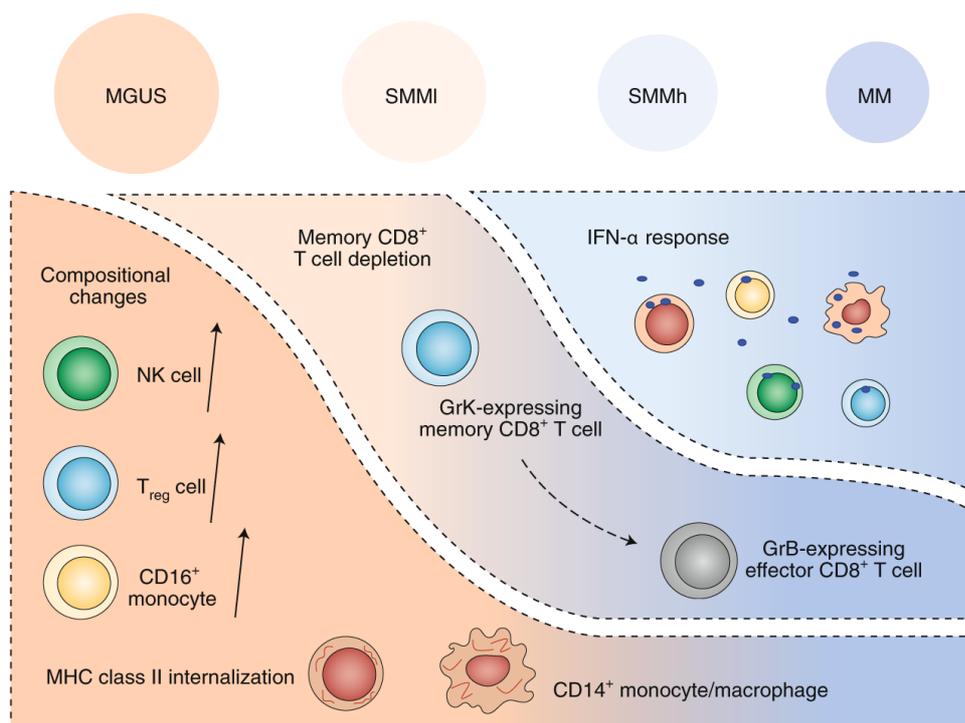
Figure 1.13: Stepwise immune microenvironment alterations through the evolution of MM.

T cell exhaustion is one of the key hallmarks of tumor-immune states. In the chronic inflammation state of tumors, the memory T cell differentiation is altered and these cells start to lose their effector function and upregulate many inhibitory receptors which mark the exhaustion state of the T cells (Wherry and Kurachi 2015). The T cell exhaustion pathways are multifaceted and have many factors that control and regulate the exhaustion state. The pathways can be grouped into <u>three</u> categories **(Figure 1.14):**

1. Cell-cell signals with prolonged T cell receptor (TCR) engagement (Wherry and Kurachi 2015).
2. High levels of inflammatory cytokines and soluble factors.
3. Microenvironmental and tissue remodeling as a result of the altered expression of adhesion molecules and chemokine receptors.

Despite recent advances in understanding the T cell exhaustion state in solid tumors and hematological malignancies; it seems that the T cell exhaustion in MM disease is difficult to define, since the typical exhaustion signatures and marker genes are not highly expressed, as have been recently shown by Zavidij et al. 2020. This means that we need more functional work to investigate and gain a deeper understanding of the T cell exhaustion state in the MM context.
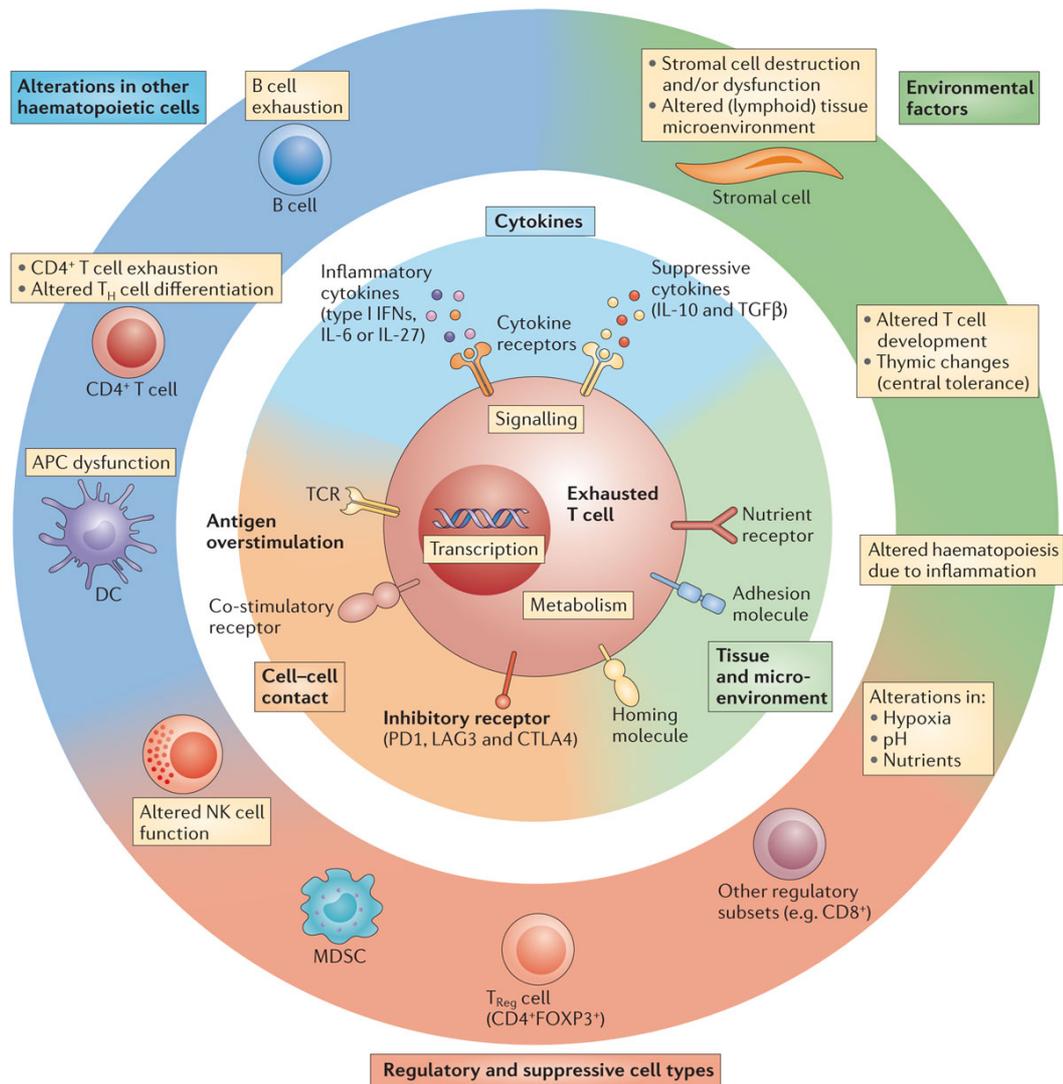
Figure 1.14: Cellular and molecular mechanisms regulating the T cell exhaustion state.

----

*Adapted from (Wherry and Kurachi 2015) by permission from Copyright Clearance Center's RightsLink® service.

1.4.4 Long-term survival and relapse: current treatment landscape and immunotherapeutic approaches

In the past 15 years, survival in MM has been improved due to the extensive efforts in understanding the biology and clinical aspects of the disease, which led to novel therapeutic approaches and drugs for treatment.

The treatment phases in MM can be grouped into four stages: initial therapy, autologous stem cell transplantation (ASCT), maintenance therapy, and treatment of relapse (Rajkumar and Kumar 2016, Agnarelli, Chevassut, and Mancini 2018). The therapeutic efficacy is reflected by the increased number of patients who achieved complete remission (CR) and extended periods of free progression.

Still, MM disease is incurable and most of the MM patients' relapse. The MM patients' survival is monitored via minimal residual disease (MRD) measures, which can be tested by real-time quantitative PCR, next-generation sequencing (NGS), and multicolor flow cytometric (MFC) (Kostopoulos et al. 2020).

The introduction of ASCT and therapeutic regimens using proteasome inhibitors and immunomodulatory drugs (IMiDs) showed the most effective ways of treating newly diagnosed multiple myeloma (NDMM) patients **(Figure 1.15)**. Lenalidomide is an immunomodulatory agent which can regulate T cell co-stimulation (via B7-CD28 costimulatory pathway), and increase the cytotoxicity functions of NK cells and alter cytokine production by downregulating IL-6 production which decreases myeloma cells proliferation and increase their apoptosis (Quach et al. 2010, Vo et al. 2018).

Immunotherapies with immune-checkpoint blockade seem to be a very effective strategy in treating MM patients; like Elotuzumab (anti-SLAMF7 antibody) and Daratumumab (anti-CD38 antibody). Elotuzumab activates NK

cells and macrophages. Daratumumab decreases T_{reg} proportion and increases helper and cytotoxic T cell proportions which leads to immune activation (Ghobrial et al. 2018). Recent studies showed that using Elotuzumab or Daratumumab alone or in combination with Lenalidomide or Bortezomib are most effective in comparison to the standard care (Palumbo et al. 2016).

Despite the fact that PD-L1 is highly expressed by MM cells, controversial data exist regarding the efficacy and toxicity of the use of Pembrolizumab (PD-1 antibody blockade) and Lenalidomide which has led the FDA to stop these studies (FDA Research 2019). These contradictions need to be resolved by better understanding the immune microenvironment and T cell cellular states in the MM context.

In addition, understanding the role of the immune-microenvironment in the MM context would be also helpful to increase the number of patients who experience the long-term survival state (> 10 years) and to potentially find a definitive cure for MM which is not an achieved target yet.

a)

FIGURE. Approach to the treatment of newly diagnosed multiple myeloma. ASCT = autologous stem cell transplant; CR = complete response; KRD = carfilzomib, lenalidomide, and dexamethasone; Rd = lenalidomide plus dexamethasone; VCD = bortezomib, cyclophosphamide, and dexamethasone; VGPR = very good partial response; VRD = bortezomib, lenalidomide, and dexamethasone. * = Rd if frail or age ≥75.

b)

Figure 1.15: Current treatment approaches for initially diagnosed MM patients.

a) MM treatment algorithm of newly diagnosed MM patients. **b)** Current drugs used for MM treatment include proteasome inhibitors and immunomodulatory drugs (IMiDs).

----

*Adapted from (Rajkumar and Kumar 2016) and (Kumar et al. 2017) by permission from Copyright Clearance Center's RightsLink® service.

1.5   The rise of single-cell genomics technologies

The scientific community started to realize the complexity of biological systems (e.g., immune system, tumor heterogeneity, and cellular differentiation and developmental processes). These systems can be dissected at the single-cell level, which will be helpful to further understand these systems and their underlying cellular interactions.

The first single-cell RNA-seq experiment has been published eleven years ago when Tang et al. (2009) started to perform mRNA sequencing for each single cell of a single 4-cell stage blastomere, which has been fully manually isolated. Since this first single-cell experiment, technological and methodological advances occurred over the years, and the scale of the sequenced cells grew exponentially **(Figure 1.16).**

Now, we can sequence thousands of cells per experiment, which can be easily scaled to sequence and profile millions of cells in one experimental run (Datlinger et al. 2019).
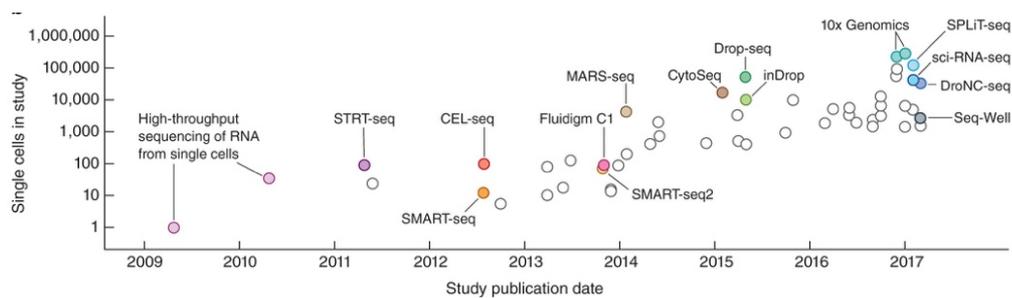
Figure 1.16: Single-cell experimental and technological advancements.

The reported number of cells per studies' publication date; representing the growth in the number of sequenced cells per study.

The droplet microfluidics technology enabled to reach the scale of sequencing to thousands of cells in the same experiment. Drop-seq was the first protocol that used the droplet microfluidics technology to sequence thousands of cells with a substantial decrease in cost per cell (6 cents/cell) (Macosko et al. 2015). The 10x genomics company started to commercialize the *inDrop* method and sell the required device and reagents, which made it available and accessible for more labs world-wide.

In the 10x Genomics setting, a single-cell RNA-seq experiment can be performed and scaled up to eight channels (each channel with different pools of cells and populations), which can be processed in parallel at the same time. In every single channel, the encapsulation process for thousands of cells occurs, where the cells are combined with the reagents of the microfluidic chip and gel beads at a fast formation rate (~100,000 Gel bead in EMulsion (GEM) / ~6-min) **(Figure 1.17)**.
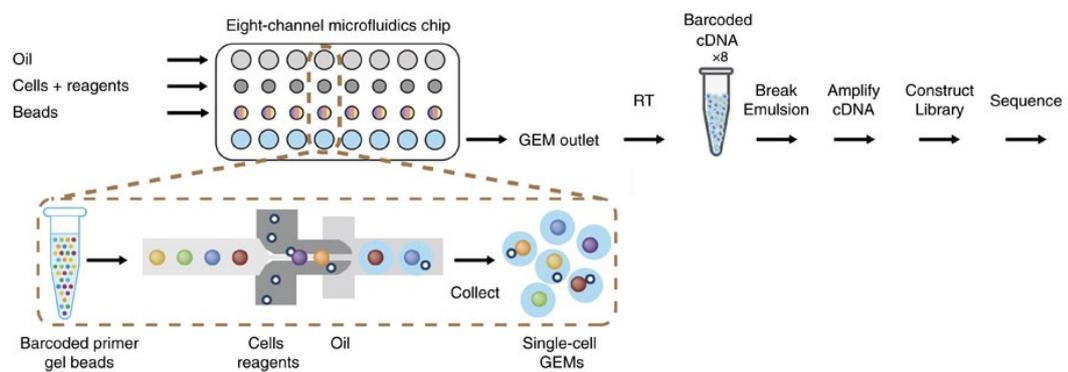
Figure 1.17: 10x Genomics' GemCode technology for single-cell RNA-seq library preparation.

----

*Adapted from (G. X. Y. Zheng et al. 2017) by permission from Copyright Clearance Center's RightsLink® service.

The gel beads are loaded with the barcoded oligonucleotides which are composed of four key parts: **1)** Sequencing adapters and primers **2)** 14 nucleotides (nt) barcode **3)** 10 nt unique molecular identifier (UMI) **4)** 30 nt oligo-dT to prime polyadenylated RNA transcripts (G. X. Y. Zheng et al. 2017).

The encapsulation process starts first when the gel beads are mixed with the cells and reagents. At the microfluidic junction, the cell/reagents/gel bead structure is mixed with the oil-surfactant solution to form the GEMs which are collected in the GEM outlet. After the success of the encapsulation process, the cell lysis process starts, and the GEM dissolves and releases its reagents for the start of reverse transcription (RT). After the GEM emulsion is broken, the barcoded cDNA is collected for PCR amplification. The final libraries are formed by sharing the amplified cDNAs and incorporating the adapter and sample indices into the cDNAs. By the end of this process, the libraries are ready for next-generation sequencing (NGS) **(Figure 1.17)**.

Further technological advancements that happened through the last couple of year; have led to capturing more layers and cellular information beyond the mRNA level (e.g., presence of cell surface proteins, intracellular proteins, chromatin accessibility, DNA methylation status, genome sequence, histone modifications, lineage tracing, and spatial co-localization) **(Figure 1.18)**.

These recent technological advances required developing new computational methods and algorithms to analyze the underlying patterns captured from each single data layer, and ultimately integrating these data types into one joint-learning process and forming holistic computational representations of cellular states in health and disease (Granja et al. 2019, Stuart and Satija 2019).
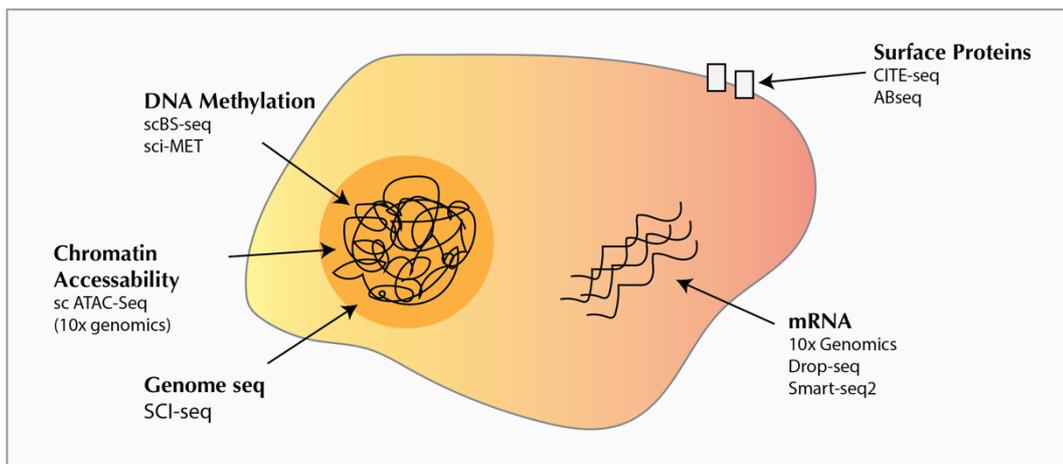
Figure 1.18: Multimodal single-cell technologies capture multiple layers of cellular information.

1.6   Challenges and advancements in single-cell computational methods and algorithms

Due to the rapid development in single-cell genomics technologies and readouts; innovative computational methods and algorithms have been developed to resolve the technical challenges inherited from different technologies, different data types, and integrating data from different batches, cell types, technologies, and species (Stuart and Satija 2019).

Single-cell RNA-seq analysis involves many different computational steps including alignment of the raw sequences to a reference genome, generating a count matrix, quality control (QC) check, normalization and dimensionality reduction, clustering, cell type annotations and further looking at cell-cell interaction, inferring developmental trajectories and gene regulatory networks (Hie et al. 2020). Even though there are similar and common steps in the analysis workflow of both bulk and single-cell RNA-seq **(Figure 1.19)**; still single-cell data analysis poses more challenges (Stegle, Teichmann, and Marioni 2015, Lähnemann et al. 2020).

1.6.1 Single-cell data sparsity

Due to the shallow coverage of the transcriptomes, single-cell RNA-seq data suffers from a large number of observed zeros which have been termed as "dropouts" or "zero-inflation" (Kharchenko, Silberstein, and Scadden 2014). These zero values could originate from <u>two</u> sources:

1. **Sampling noise**: where the genes are not detected via sequencing machines.
2. **True biology:** where the genes are not expressed in certain cells in a given biological system.
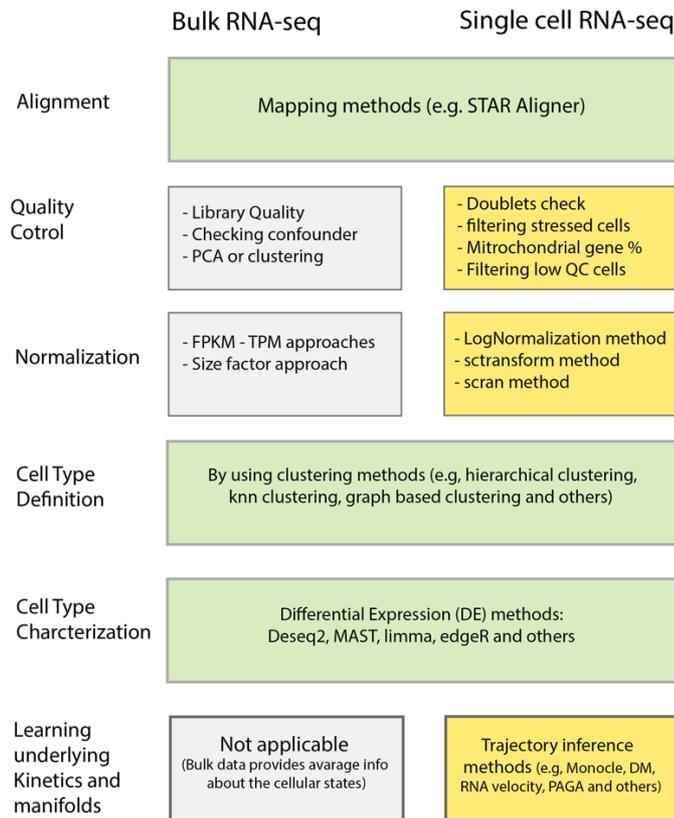
Figure 1.19: Common single-cell RNA-seq data analysis workflow and the key computational methods underlying these analysis workflows.

Many of these steps are common between the Bulk RNA-seq and single cell RNA-seq (green color). Some other steps are specific to each data type (gray: Bulk and yellow: single-cell).

These assumptions led the community to think and develop new methods, which could use these zero information, with consideration of using zero-inflation models and mixture models where every single cell has modeled as a mixture of dropouts (Poisson distribution) and amplification (NB distribution) (Kharchenko, Silberstein, and Scadden 2014, Pierson and Yau 2015). Furthermore, imputation methods have been developed to predict the expression of genes with zero values (Andrews and Hemberg 2018).

Despite these efforts to either model or correct for over-zeros, a recent study by Svensson 2020 and Choi et al. 2020 showed that the primary cause of zero-inflation is mainly due to the biological nature of the cells, and using a generalized linear model (GLM) with negative binomial count distribution would reflect the biological variation.

1.6.2 Sampling and biological variations

The sources of variations in the gene expression matrix could be due to sampling noise or biological variation. Normalization and selecting highly variable genes (HVGs) are two key methods that are currently used to minimize technical variation while preserving the biological variation.

Normalization methods try to generate consistent comparisons between the measured genes across many cells, and to correct for the variation in the number of captured reads/transcripts per gene and per cell. Many normalization methods have been introduced, e.g., SCnorm (Bacher et al. 2017), scran (Lun, McCarthy, and Marioni 2016), Deseq2 (Love, Huber, and Anders 2014, 2).

The size factor methods are the most frequently used; gene counts are normalized by the total counts per cell, then are scaled by an arbitrary factor (e.g., 10^4 or 10^6), then pseudocounts are added and finally log-transformation is performed. Such an approach ignores the real differences in the total mRNA counts and cell sizes between the different cell types and states (Hie et al. 2020). SCTransform models the counts per gene using a regularized negative binomial model, where the sequencing depth is used as a covariate in the GLM model (Hafemeister and Satija 2019).

The highly variable genes (HVGs) methods try to select the most variable genes across all cells based on their expression level. The HVGs methods assume that the true biological signal differences would manifest as an increased variation in the affected genes in comparison to the other genes which have been mainly affected by the technical factors. The simplest form of calculating the HVGs is by computing the variance of the log-normalized expression values of each gene for all cells. This per-gene variance is calculated by modeling the mean-variance relationship and selecting for the top variable genes (e.g. 2000-5000 genes) (Amezquita et al. 2020).

1.6.3 Dimensionality reduction

Since the single-cell RNA-seq technologies capture the full transcriptomic features of the samples, this does not imply that all genes are important for a certain biological system. Therefore, reducing this high data dimensionality to (low dimensional manifold) would be a reasonable task to preserve the most important structure in the data and to gain a more intuitive understanding of the data.

The dimensionality reduction methods can be grouped into <u>two</u> categories:

1. **Linear decomposition methods:** e.g., principal component analysis (PCA) (F.R.S 1901), singular value decomposition (SVD) (Kalman 1996) and GLM-PCA (Townes et al. 2019).

2. **Non-linear methods** (for mainly visualization purposes): e.g. t-distributed stochastic neighbor embedding (tSNE) (Maaten and Hinton 2008) and uniform manifold approximation and projection (UMAP) (McInnes, Healy, and Melville 2020).

The most commonly used linear method is the PCA, which tries to discover the principle components (PCs) or "axes" to capture the largest amount of variation in the data's high dimensional space which enables for more advantages in the downstream analysis steps (Amezquita et al. 2020).

The non-linear methods reduce the data to two or three dimensions as a representation for each cell in a more human intuitive space, in addition to preserving the local structures and patterns in that space. The non-linear methods preserve these local structures, but this process could introduce distortions.

Looking at the data in its high dimensional space is needed for a deeper interpretation and understanding. For example, UMAP tries to learn low-dimensional embeddings of high-dimensional data by introducing a set of heuristics, like force-directed embeddings (Kobourov 2012), which enables the visualization of k-nearest neighbor (KNN) graph topology (Altman 1992). This process is controlled by a set of parameters to enable higher density distortions (Hie et al. 2020).

1.6.4 Clustering cell types

Clustering methods are crucial tools to group cells based on the similarity of their gene expression patterns across many cells to define the underlying cell types and describe their heterogeneity. While many clustering algorithms have been used for single-cell RNA-seq data, graph-based clustering is largely adapted and used (Duò, Robinson, and Soneson 2018).

The graph-based clustering starts with constructing a network graph where each cell represents a node in the graph and is connected to its nearest neighbor (NN) via edges. The edge weights are calculated based on the similarity between the cells, and higher weights indicate that these cells are closely related e.g., by the Louvain and Leiden algorithms (Traag, Waltman, and van Eck 2019).

1.6.5 Differential expression (DE) analysis between clusters, cell types, and subpopulations

Differential expression (DE) analysis methods have been developed for the bulk RNA-seq and microarray data analyses, with the main interest to determine whether the mean expression of a given gene is showing a statistically significant difference for different sample groups.

Further methods have been developed and specifically tailored for single-cell RNA-seq data, e.g., MAST (Finak et al. 2015), and SCDE (Kharchenko, Silberstein, and Scadden 2014). A benchmark paper has been published (Soneson and Robinson 2018) and conducted a comparison between 36 DE tools. The authors showed that very simple tests like t-test or Wilcoxon rank-sum test ranked high scores, in addition to bulk-DE-methods (e.g. limma) and MAST.

Other approaches have been proposed to *average* single-cell data to pseudobulk and perform DE testing using bulk DE methods. This approach showed that it outperforms single-cell based methods and performs similarly well as complex mixed model-based approaches (Crowell et al. 2020). However, the pseudobulk approach could lose the important properties of the single cell data by averaging and *pseudobulking* the single cell signals.

Ntranos et al. 2019 revisited the use of the logistic regression (LR) model for microarray data analysis. They reasoned that the single-cell data size enables appropriate fitting of the data and showed that LR model outperforms both bulk and single-cell based methods. Moreover, Stuart et al. 2019 showed that the LR approach can be used for large-scale single-cell datasets and enables including donor information and other covariates in the model.

1.6.6 Trajectories and developmental process inference

Cellular states have been mainly defined by molecular surface markers (e.g., CD34$^+$). Due to current technological advances in capturing and mapping cellular states through measuring the whole transcriptome via single-cell RNA-seq; these cells have been observed to follow a *continuum* manifold of cellular states in many different tissue types and organs (D. E. Wagner and Klein 2020). To infer the *continuum* manifold, computational methods have been recently developed to construct data-driven models from single-cell data, predict the cellular dynamics of these cells, and visualize these manifolds (Saelens et al. 2019).

Most of the current methods begin with constructing a graph where each node represents a cell and the edges represent the gene expression similarity **(Figure 1.20 b)**. Then these algorithms start to extract the manifold structure topology or organize the cell into a trajectory axis to predict the future state of these cells **(Figure 1.20 c)**. To visualize these graphs in two or three dimensions, methods such as UMAP, SPRING (Weinreb, Wolock, and Klein 2018), and ForceAtlas2 (Jacomy et al. 2014) have been used. Such visualization can be intuitive for the human mind; however, it could be misleading as it represents a distorted representation of the high dimensional data (as discussed in the Dimensionality reduction section). More recent methods, e.g., RNA velocity (La Manno et al. 2018) or scVelo (Bergen et al. 2020) have been developed based on the idea of using nascent mRNA proportions to calculate the rate of change of the spliced and unspliced mRNA ratios across all genes to predict developmental directionality and future cellular states.

In general, these methods order the cells along a continuum and allow to study the average changes, variance, and gene expression correlation across a graph to infer tree-like structures (Qiu, Hill, et al. 2017) or other expected topologies and hierarchies (Saelens et al. 2019).
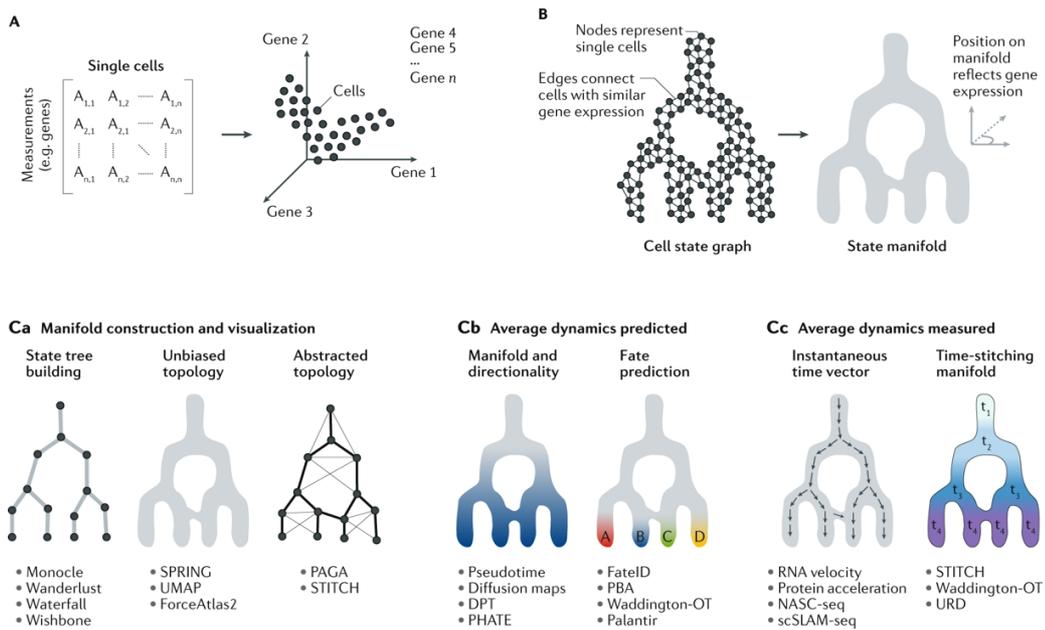
Figure 1.20: Trajectory inference from cellular state manifolds.

a) Gene expression matrix (genes x cells) which can be plotted in high dimensional space. **b)** Cells are connected according to the similarity in the gene expression space forming a graph. **c_{a,b,c})** Graph-based methods to construct and visualize the cellular state manifolds.

1.6.7 Data Integration across batches, technologies, and species

The current technology enables sequencing of thousands and millions of cells across different batches and cellular states and using different technologies, even from different species which pose many challenges to perform data integration.

These challenges echo with "batch-correction" techniques for bulk data. Newly developed data integration methods have been tailored to resolve single-cell data challenges (e.g., cell type heterogeneity, cellular state shifts, and others). These new integration methods can identify shared biological features and states e.g., matched cell type across batches and conditions (Stuart and Satija 2019) (**Figure 1.21**).
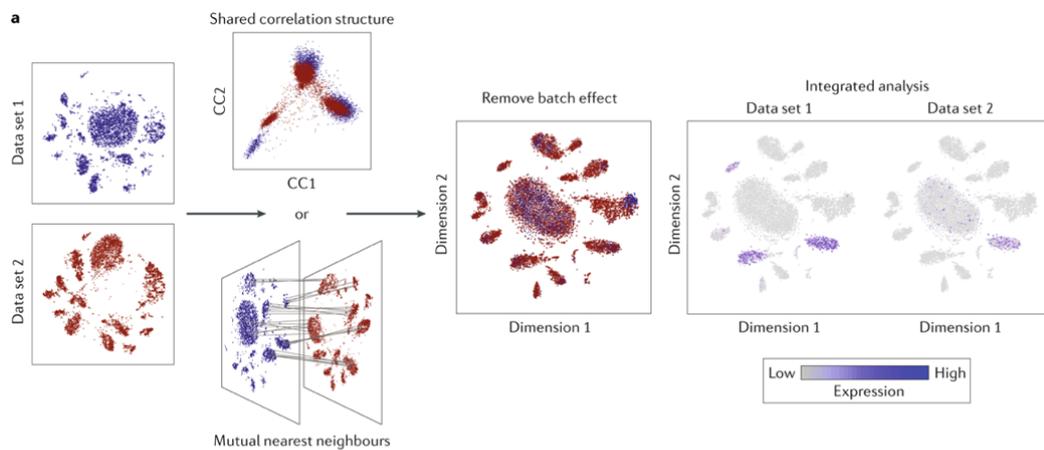
Figure 1.21: Workflow for data integration between different datasets and batches.

The purpose of these methods is to find shared correlation structures by using canonical correlation analysis (CCA) or mutual nearest neighbors (MNNs).

Two prominent methods have been developed recently and had an important impact on resolving the data integration challenges. The first method mainly uses canonical correlation analysis (CCA) to find shared sources of variation between the different datasets / batches which produces a vector of features, and by using a dynamic time warping algorithm, these vectors are aligned across datasets (Stuart and Satija 2019). These steps project the cells into a shared low-dimensional space and further locate the cells with the same biological state to close locations in that space, regardless of the batch design, experimental setup, or technical components (Butler et al. 2018). The second method uses the mutual nearest neighbors (MNNs) approach to define the cells which are mutually closest to the other cells in the datasets and which therefore could represent a shared cellular state (Haghverdi et al. 2018) **(Figure 1.21).**

Both methods had an important influence on developing other data integration methods (e.g., Scanorama, Conos, Harmony, BBKNN, and others). A recent benchmark paper (Luecken et al. 2020) showed that BBKNN, Scanorama, and Seurat v3 methods show overall high performance scores in many data integration tasks which are mainly based on the CCA and MNNs ideas.

1.6.8 Cell-cell interaction network

Computational methods have been developed to learn potential receptor-ligand interactions and communication between cell types from single-cell RNA-seq data, e.g., cellphoneDB (Efremova et al. 2020), and nichenet (Browaeys, Saelens, and Saeys 2020). These methods are based on a curated list of receptor-ligand pairs, which could bias the analysis since it is based on the curator's selection and expertise. Besides these methods are based on the expression of the mRNA of receptors and ligands, which is not always detectable by current single-cell technologies.

Despite that these methods have been used and showed benefits in many published studies (Vento-Tormo et al. 2018, Baccin et al. 2020, Browaeys, Saelens, and Saeys 2020). Further caution is needed while interpreting the results of these methods due to the lack of benchmark studies or ground truth data. Further development in the direction of cell-cell interaction methods is needed to leverage the full potential of the single-cell data and have detailed de novo and single-cell data-driven interactome maps.
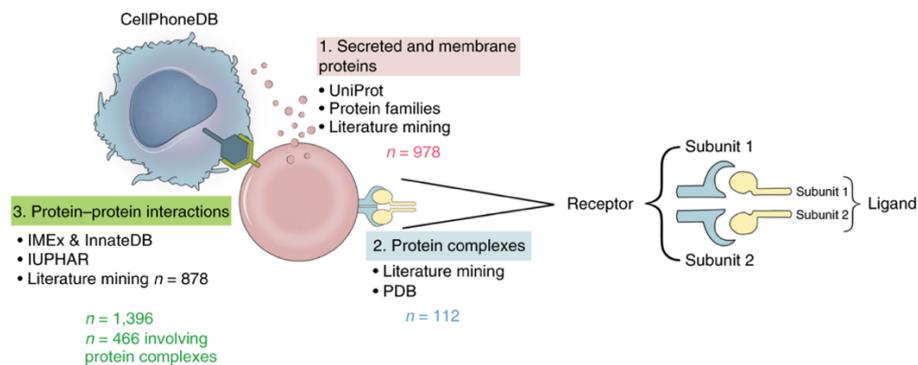
Figure 1.22: CellphoneDB method overview for cell-cell communication inference by using a curated list of receptor-ligand pairs.

----

*Adapted from (Efremova et al. 2020) by permission from Copyright Clearance Center's RightsLink® service.

2 Aims of the thesis

Through the past decades, we have started to gain deep insights into the MM disease and its clinical manifestation. We started to observe that few MM patients experience long-term remission (LTS) over several years (~ 7 to 17 years). Even though, they ultimately relapse!

The Long-Term Survival (LTS) phenomenon triggered our curiosity to establish this study in the context of MM. The study aims to investigate the molecular and cellular profiles of LTS patients' bone marrow immune microenvironment and its link to their LTS states. For this, we use computational models, single cell genomics technologies and wet lab validation.

The key objectives of this thesis:

1. Define the global landscape of MM patients' bone marrow immune microenvironment before treatment (BT) and in long-term survival (LTS).

2. Dissect the cell types and phenotypic states, and define the key biological programs and cellular states, which are linked to the LTS.

3. Define the potential role of BM immune microenvironment in controlling the malignant plasma cells' growth and progression.

4. Characterize the immuno-phenotypes of the MM patients via specific surface markers, transcriptional factors, gene signatures, and cell-cell interaction network patterns.

5. Define new therapeutic targets, prognostic markers, and signatures for MM disease states.

6. Build a new model that can explain the LTS phenomenon in the MM disease context.

The thesis objectives raised many key computational challenges which needed innovative ways to be tackled and resolved:

1- To broadly define cell types, I aimed to cluster the single-cell RNA-seq data and map the clusters to known cell types' marker genes and bulk data references.

2- To define characteristic marker genes, gene signatures, and biological pathways, I worked on finding the optimum differential expression (DE) model to include the complex experimental and clinical covariates, and run pathway enrichment analyses.

3- To predict and classify the cellular states across the clinical groups, I started to develop new approaches to resolve this challenge using a generalized linear mixed model (GLMM) and random forest (RF) model.

4- To study cell types development and differentiation processes, I started to use and finetune trajectory inference and RNA velocity methods to construct developmental trajectories.

5- To gain systems-level understanding and build a new immune model, I developed a new way to model the global pathways signature scores using generalized linear model (GLM).

6- To construct cell-cell interaction, I aimed to perform the receptor-ligand interaction networks which could mediate the cross-talk between cell types and states.

7- Furthermore, I worked extensively on the data interpretation, to link the analyses findings with current published work from the literature; to better understand the biological processes in the context of MM disease states.

Collectively, I provide a detailed molecular and cellular state description of the bone marrow immune microenvironment and the tumor compartment. I define new cellular states and populations associated with long-term survival states. I propose a new cellular-state therapeutic target and prognostic marker genes in the T cell compartment. Finally, I propose a new model that explains the MM long-term survival phenomenon.

3 Methods

I developed the computational approach and used the methods mentioned in this chapter except section **(3.1)**. The methods section **(3.1)** has been done in a collaboration with Prof. Dr. med. Michael Hundemer, Mohamed H. S. Awwad and Dr. med. Raphael Lutz from the Laboratory for Translational Immunology (TRIM), Heidelberg University Hospital, Dr. Jan-Philipp Mallm from Single-cell Open Lab (scOpenLab - DKFZ), and DKFZ Genomics and Proteomics Core Facility.

3.1   Sampling strategy, library preparation, and next generation sequencing (NGS)

We collected bone marrow samples from three healthy controls and eleven multiple myeloma patients at two different time points (a diagnosis time - Before treatment (BT), and in long-term survival (~ 7-15 years after initial diagnosis) in collaboration with Prof. Dr. med. Michael Hundemer, Mohamed H. S. Awwad and Dr. med. Raphael Lutz from the Laboratory for Translational Immunology (TRIM), Heidelberg University Hospital.

After thawing, the bone marrow cells were initially stained with anti-human CD45 and CD3 antibodies, followed by a Vybrant DyeCycle Violet stain according to the manufacturer's recommendations. For each time point, live total bone marrow and live $CD45^+CD3^+$ cells were sorted using a BD FACSAria Fusion (was performed by Dr. med. Raphael Lutz) **(Figure 3.1)**.

The 10x genomics platform (kit version 2) has been used for single-cell RNA sequencing, according to the manufacturer's recommendations (was performed by Dr. med. Raphael Lutz in collaboration with scOpenLab at DKFZ).

For each library, an individual sequencing run was performed on Illumina HiSeq 4000 machine (using paired-end sequencing protocol) and allocated one lane per sample for sequencing was performed by the DKFZ Genomics and Proteomics Core Facility.
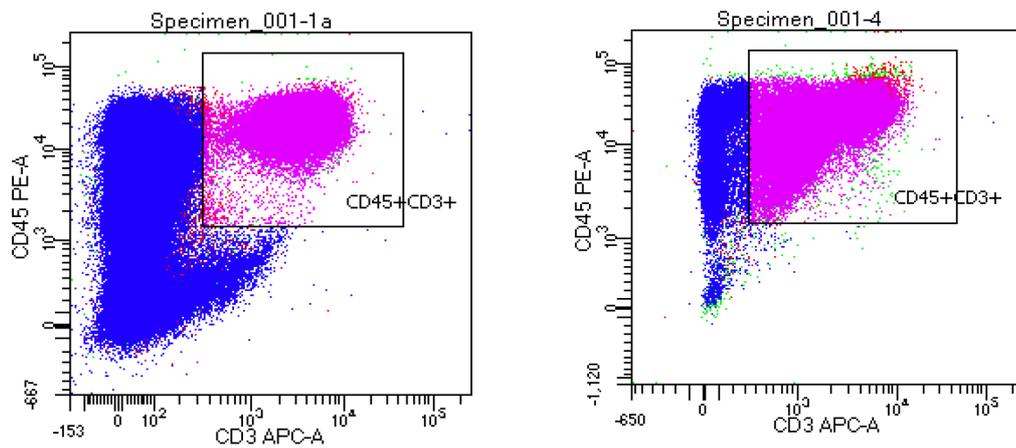
Figure 3.1 Sorting strategy for CD45$^+$/CD3$^+$ population from the bone marrow samples at different time points.

Source (collaborators: Dr. med. Raphael Lutz and Prof. Dr. med. Michael Hundemer)

3.2 Developing single-cell RNA-seq bioinformatics analysis workflow

I developed a computational approach for processing and analyzing the single-cell RNA-seq data, which can be divided into two main parts **(Figure 3.2)**:

- **The first part** starts with the upstream processing of the raw sequencing data using the Cell Ranger pipeline and constructing a gene-barcode matrix.

- **The second part** is based on statistical inference methods and performing downstream analyses, which are based on many standard methods in R and Python computational environments. Furthermore, I developed a generalized linear mixed model (GLMM) for the single-cell abundance analysis; to build conclusions about the association of cell types proportions and the clinical state of our patients. In addition, to quantify the cellular states; I developed a random forest model to predict and quantify the phenotypic expansion of the cellular states and phenotypes within our clinical groups.
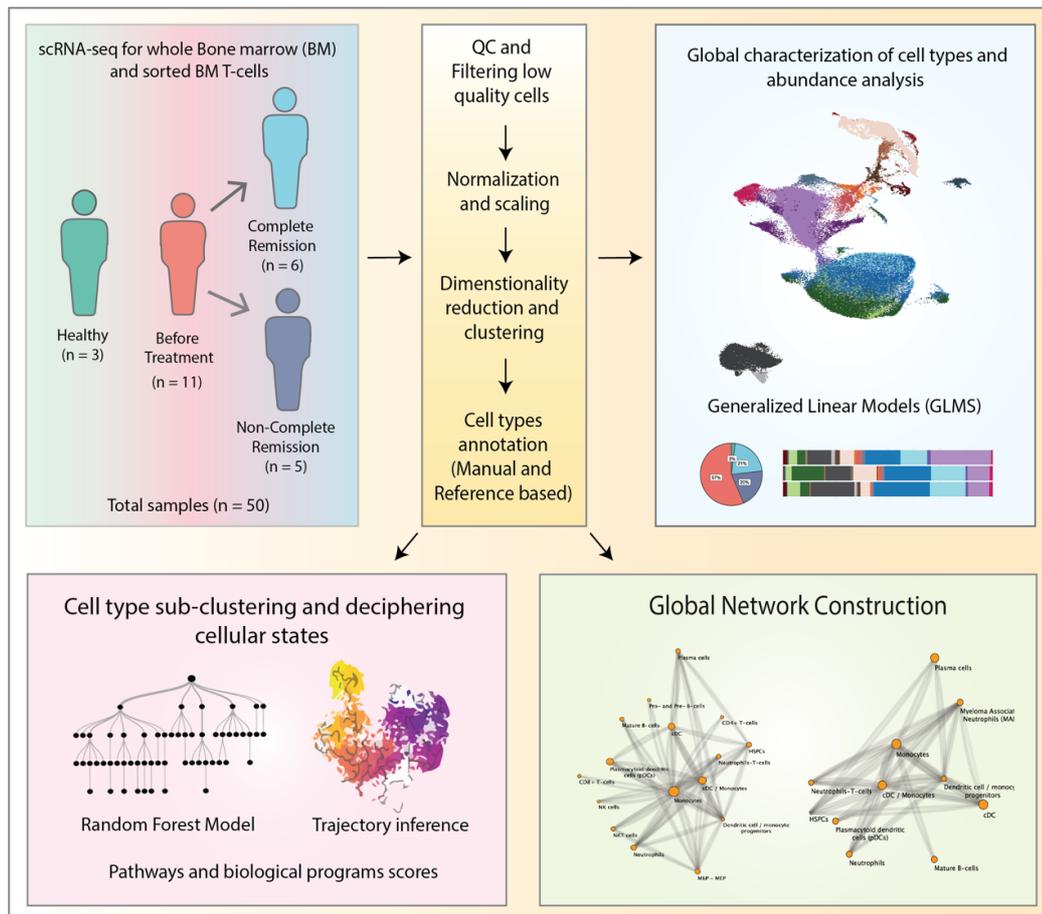
Figure 3.2: The overall bioinformatics workflow and developed approaches to analyze single-cell RNA-seq data.

3.2.1 Upstream analyses: Cell Ranger pipeline

To generate the gene-expression matrix; I used the Cell Ranger pipeline (version 3.0.1) (G. X. Y. Zheng et al. 2017). This is a pipeline which can process Chromium single-cell RNA-seq output to perform read alignment, generate gene-barcode matrix and perform other downstream analysis tasks **(Figure 3.3)**.

I started the upstream analysis with the raw data of the samples in FASTQ format (yellow box). I used (refdata-cellranger-GRCh38-1.2.0) as the human reference genome provided by 10x genomics to generate the gene-expression matrix (green box). I started using the Cell Ranger pipeline by running *cellranger count* function on each sample separately to extract the cell-barcode, UMI, and RNA reads, and to correct the cell-barcode sequencing errors. Furthermore, the read alignment step is mainly based on the STAR aligner (Dobin et al. 2013), which maps reads into exonic, intronic, and intergenic regions of the genome. The confidently mapped reads to the annotated transcripts model are considered for the UMI counting step.
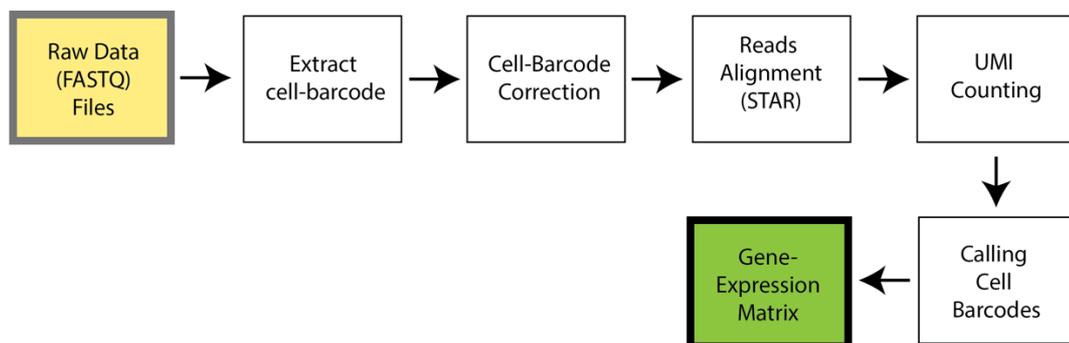
Figure 3.3: Overview of the Cell Ranger pipeline's main steps.

The yellow box represents the start point of the raw FASTQ files and the green box is the final output of the pipeline, which represents the gene-expression matrix for further downstream analyses.

After the correction step, Cell Ranger starts the UMI counting to generate gene-barcode matrix (green box). Cell Ranger (version 3.0.1) introduced a new algorithm in the pipeline that is more efficient in identifying populations with low RNA content per cell and defining real cells from empty droplets (calling cell barcodes). This is a critical step especially if the data represents an underlying heterogeneous population of cell types with different sizes and RNA content (Lun et al. 2019).

3.2.2 Downstream analyses: statistical inference and learning

After composing the gene-expression matrix for all samples, I imported this matrix into R for the downstream analysis. I used two main R packages: Seurat v3 (Stuart et al. 2019), and Monocle v3 (Qiu, Mao, et al. 2017) for QC check, statistical analysis, biomarker discovery, and trajectory analysis.

3.2.3 Quality control (QC) and selecting cells for the downstream analysis

I started the downstream analysis by creating a Seurat object and setting initial filtering criteria for the raw absolute count data (non-normalized data). I selected genes that are expressed in ≥ 3 cells and kept the cells with ≥ 200 detected genes. I exclude cells with an outlier number of detected genes (according to the cell types detected genes), which could be considered as potential doublets. Furthermore, I filtered cells based on the percentage of mitochondrial genes to the total UMI counts per cell (5 to 10%).

3.2.4 Normalizing the data and detecting highly variable genes (HVGs)

I normalized the data using *NormalizeData* function which uses *normalization.method ="LogNormalize";* to normalize the gene expression measurements for each cell by total expression, multiplies this value by a scale factor (10^4 by default), and log-transforms the result. I detected the highly variable genes by using *FindVariableGenes* function and specifying the following parameters *selection.method = "vst"* and *nfeatures =2000* for the downstream steps.

3.2.5 Data Scaling and regressing out undesirable sources of variation

Single-cell data often has several technical confounders (e.g., batch effect, number of detected molecules, etc.) and biological variability (like that introduced the cell cycle stages), which should be regressed out to gain a clear biological signal that represents the true biological variations between the cells. I used *ScaleData* function and *model.use = "negbinom"* to regress on the number of detected molecules per cell and the percentage of mitochondrial gene content.

3.2.6 Linear dimensionality reduction

I performed principal component analysis using the *RunPCA* function on the scaled data. Then, I used *PCElbowPlot* function to determine the statistically significant principal components (PCs) by ranking the PCs depending on the percentage of explained variance by every PC and selecting the top PCs (~ 20-50 PCs) for the next steps.

3.2.7 Clustering cells

I started clustering the cells using a graph-based algorithm (shared nearest neighbor (SNN)-Cliq). I constructed the shared nearest neighbor (SNN) graph which is implemented in Seurat (FindNeighbors function) (Xu and Su 2015). Then, I used a smart local moving algorithm (implemented in FindClusters function) to optimize the modularity function and define clusters (Waltman and van Eck 2013).

3.2.8 Non-linear dimensionality reduction

To learn the underlying data manifold, I used the uniform manifold approximation and projection (UMAP) algorithm to co-localize the cells in a low-dimensional representation (2d or 3d) (McInnes, Healy, and Melville 2020). I used the same PCs and ran the *RunUMAP* function to compute the UMAP algorithm based on the scaled gene expression data.

3.2.9 Finding differentially expressed genes and biomarkers

To define the biological markers and differentially expressed genes that are specific for clusters or clinical groups, I used the *FindMarkers* function and used the Logistic Regression (LR) model by specifying *test.use = "LR"* (Ntranos et al. 2019). Furthermore, I used patient IDs, gender, and cellular detection rate (CDR) as latent variables (Finak et al. 2015). In addition, further parameters have been specified (logfc.threshold = 0.25, min.pct = 0.1, and only.pos = TRUE).

3.2.10   Cell type annotation

We manually annotated the main cell types in the bone marrow according to the gene expression of known canonical markers derived from the literature (based on the marker genes provided by Dr. Simon Haas) **(Supplementary Figure 6.2)**. Regarding the T cell subtypes annotation, I faced major challenges to initially identify the CD4$^+$ and CD8$^+$ T-cells, since they are transcriptionally similar and CD4 gene expression is very sparse in 10x genomics data. To tackle this challenge, I used the reference-based method SingleR (Aran et al. 2019) by using the *SingleR* function to define the CD4$^+$ and CD8$^+$ T cell subtypes, which enabled us to expand the definitions of these cells beyond just the sparse expression of CD4$^+$ and CD8$^+$ surface marker genes.

3.2.11   Single-cell Abundance Analysis (GLMM approach)

To test the compositional shift of cell types across the clinical groups, I modeled the association of cell types and clinical states with cell types abundance using a generalized linear mixed model (GLMM) implementation in lme4 R-package and specified *family= poisson, link=log*. Specifically, I used the following model:

$$Y_{ij} \sim \text{celltypes + clinicalgroups + celltypes:clinicalgroups + (1 | paired sample IDs)}$$

Here, Y_{ij} denotes the raw counts of celltype i in the sample j where the celltypes, clinicalgroups, and the interaction term of both covariates are modeled as fixed effects, and the paired sample IDs are modeled as a random effect. I then used the statistically significant model estimates (p-value <0.001) of the interaction term between the cell types and clinical groups based on Wald-Test for the downstream visualization.

3.2.12 Single-cell RNA-seq copy number alterations (CNV) analysis

To define the chromosomal alterations (large segments of chromosomal gain or deletion), I used the inferCNV method which leverages the single-cell RNA-seq data to detect such genome-wide alterations (Tirosh et al. 2016). In general, inferCNV computes the average expression of many genes across several positions of the tumor cells' genome in comparison to normal cell types as a reference.

More specifically, I used the *infercnv::run* function which starts by filtering genes below a certain threshold (cutoff=0.1), performing normalization, and log transformation. Then the infercnv algorithm performs centering by using the normal reference mean value of each gene and subtracting these values from the corresponding genes in the tumor cells.

Furthermore, a smoothing step is performed at the chromosome level, and the relative adjustment to the normal reference is computed. The log-transformation is performed and the final values are inverted for efficient representation of the symmetry in the gains and losses per chromosome. In addition, I used a de-noising step *denoise=T* to further reduce the noise ratio. Finally, I used a hidden Markov model *HMM=T* to predict the final CNV states (deletion, neutral, or amplification) (Fan et al. 2018).

3.2.13   Classifying cellular states - Random Forest (RF) model

I developed Random Forest (RF) classifier **(Figure 3.4)** to quantify the phenotypic expansion and classify the cellular states using the randomForest R package (Breiman 2001). I started training the RF model with only healthy and Before Treatment (BT) cells after a down-sampling step *n=2000 cells*; to avoid imbalance sampling. Additionally, I performed feature selection and used the top *2000* highly variable genes (*HVGs*) for training the model *decision trees = 1000*. Then I used the trained model to predict the phenotypic composition and to classify the cellular states in all clinical groups including the CR and non-CR groups.

After performing the prediction step, I obtained continuous probabilistic scores (from 0 to 1) for every single cell, reflecting the extent of similarity of every single cell to the healthy-like or before treatment (BT-like) state respectively. Finally, I used the highest prediction score for each cell to assign the final label to a cell.
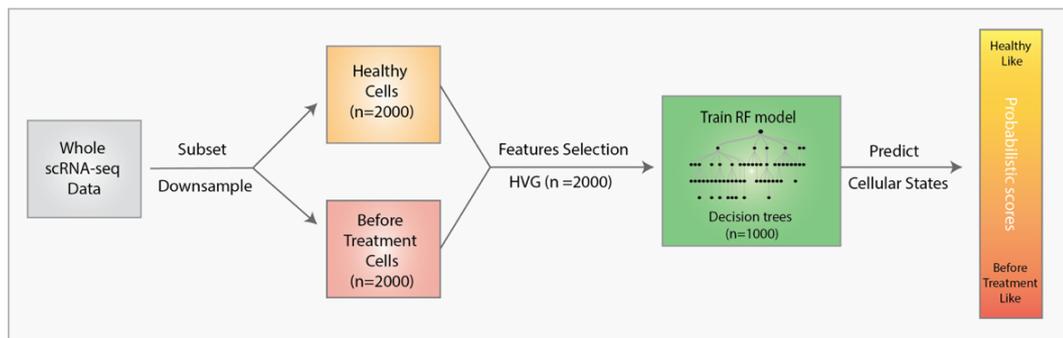
Figure 3.4: An overview of the Random Forest (RF) model to quantify cellular states in each cell type and defining Healthy (H) -Like and Before treatment (BT) -like cells.

3.2.14 Gene set enrichment analysis (GSEA) and biological program scoring

To define the underlying enriched biological processes, I used the *fgsea* function in the Fast Gene Set Enrichment Analysis R package (version 1.12.0) (Korotkevich, Sukhov, and Sergushichev 2019). I used a priori defined gene signatures and pathways from the MSigDB database (specifically, Hallmark:H, Curated gene set:C2, Gene Ontology:C5, and Immunological signature gene sets:C7) (Liberzon et al. 2015).

To compute the overall program score/per cell for a given gene set or biological program, I used *Seurat::AddModuleScore* function to calculate the average expression of the given gene set for every single cell and subtract the averaged values from the aggregated expression of a randomly selected control gene set (*n=100*).

3.2.15 Trajectory inference and mapping cellular states

To construct a developmental trajectory of CD8$^+$ T cell subtypes and states, I performed log-normalization and calculated lower-dimensional space PCA (*n=50*). Consequently, I performed clustering using Leiden community detection algorithms (Traag, Waltman, and van Eck 2019) and UMAP representation. In addition to the clustering, I performed graph partitioning and abstraction using the PAGA connectivity measure, by partitioning the graph obtained from the clustering representation into smaller homogeneous subclusters and subpopulations (Wolf et al. 2019).

To assign pseudotime, I specified the healthy CD8$^+$ cells as the initial starting point of the trajectory and ran the *learn_graph* function and the *order_cells* function with the default parameters to learn the entire trajectory which are implemented in Monocle v3 (Cao et al. 2019).

I performed spatial autocorrelation analysis using Moran's I (H. Li, Calder, and Cressie 2007) to define co-regulated genes through the developmental process of interest within the trajectory. Then I used these correlated and anticorrelated genes to define certain modules of co-expressed genes which could be linked back to our clinical groups and possibly explain the clinical states.

In addition, I applied RNA velocity (La Manno et al. 2018) based on the ratios of spliced / unspliced mRNA to predict the potential origins of aberrant memory-cytotoxic (AMC) CD8$^+$ T cells from different CD8$^+$ memory subtypes. I used scVelo (Bergen et al. 2020) to resolve the transcriptional dynamics of the splicing kinetics by running the dynamical model implementation *scv.tl.recover_dynamics* function in the python environment.

I computed the RNA velocity estimates by using *scv.tl.velocity* function. In-addition, I calculated a connectivity score based on PAGA connectivity measure and random walk based-distance by running *sc.tl.paga(groups='leiden', model='v1.0') function* (Wolf et al. 2019).

3.2.16   Cell-cell interactions and constructing global networks

To analyze the receptor-ligand interaction patterns between cell types and states in different clinical groups; I used a curated list of known receptor-ligand interactions from CellPhoneDB *v.2.0* (Efremova et al. 2020) and ran *cellphonedb method* function in a Python virtual environment to infer potential receptor-ligand interactions between the cell types for each clinical group separately.

Since each clinical group has different cell types and proportions, I performed subsampling of the data using a geometric sketching algorithm (Hie et al. 2019) by activating the subsampling function and specifying *--subsampling-num-cells 18000*. I selected the statistically significant interaction partners for the downstream visualization (p-value < 0.05) for each cell type per clinical group. Finally, I used Cytoscape (*v3.8.0*) to construct and visualize the global cell-cell interaction networks between cell types for each clinical group.

4 Results

4.1 The global landscape of the bone marrow immune microenvironment

To generate a deep transcriptional immune landscape of the multiple myeloma (MM) bone marrow (BM) microenvironment, we performed single-cell RNA-seq for the whole bone marrow and FACS-sorted CD3$^+$ T cells (n=50 paired-sample) of newly diagnosed multiple myeloma patients (before treatment), n=11, in long term survival (LTS), n=11, and healthy controls, n=3. Bone marrow cells were isolated and library preparation was prepared using the 10x genomics technology, and afterwards sequenced (see Sampling strategy, library preparation, and next generation sequencing in the Methods section).

After quality control checking (QC) and filtering **(Supplementary Figure 6.1)**, I performed merging for all data across all clinical groups since I did not observe a batch effect. The cells were clustered according to cell types from different batches and clinical groups. I used the graph-based clustering and UMAP representation to classify and catalogue cell types and subtypes (**Figure 1.1**). We performed manual cell type annotation using known canonical markers of each cell type (provided by Dr. Simon Haas) **(Supplementary Figure 6.2),** and I used a reference-based method to resolve the challenges in defining and annotating T cell subtypes, NK and NKT cells.
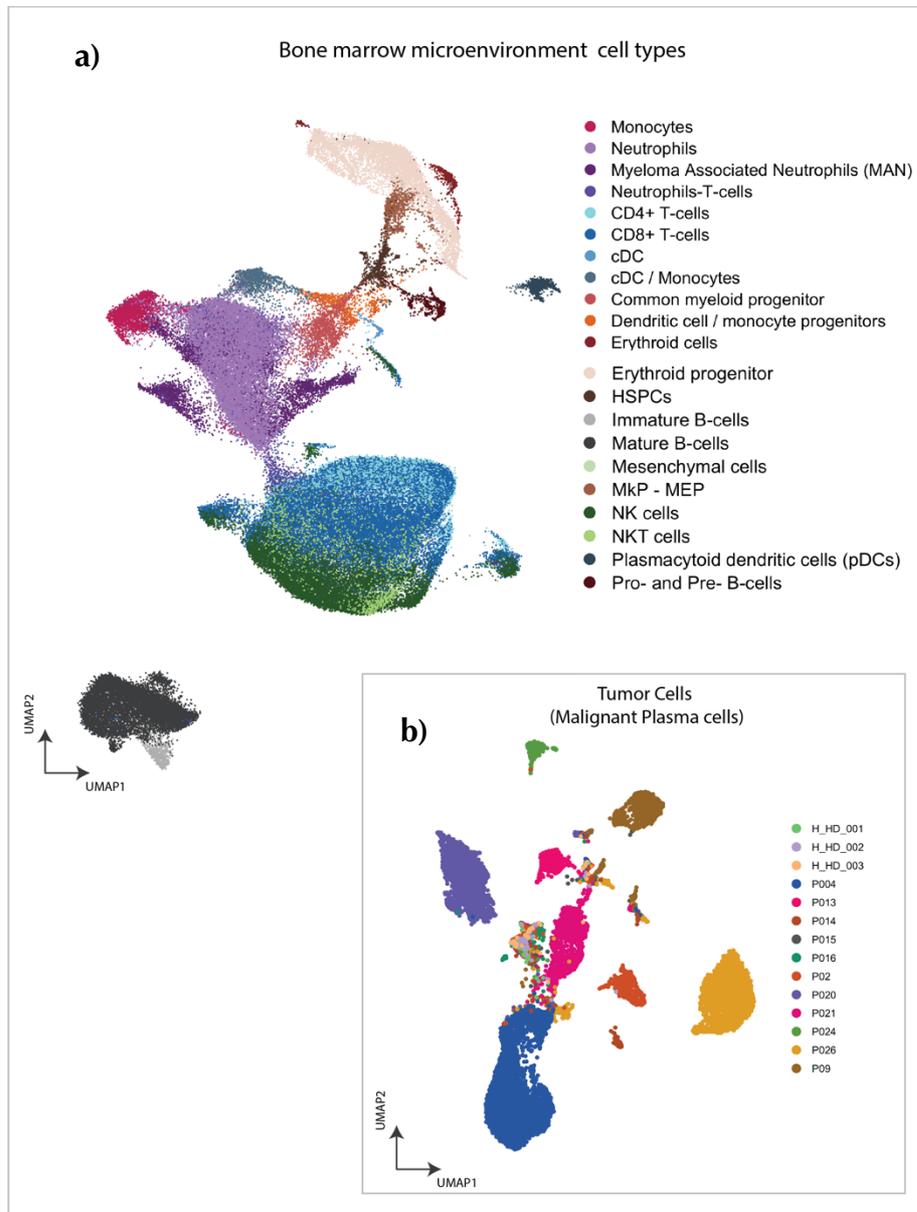
Figure 4.1: The global landscape of the bone marrow microenvironment.

a) UMAP representation of bone marrow cells of all clinical groups showing the main captured cell types and states in the bone marrow immune microenvironment. **b)** UMAP representation of the plasma cell compartment across patients' samples as defined by markers gene expression.

4.2 Cellular abundance and compositional shifts in the immune microenvironment before and after Long Term Survival (LTS)

We were able to define and annotate 23 subtypes of cells in the bone marrow microenvironment, including the expected immune cell types; T-cells, B-cells, NK, NKT, neutrophils, dendritic cells (cDC), monocytes, hematopoietic stem cells and progenitor, and plasma cells reflecting the underlying complexity of bone-marrow niches across all clinical groups (**Figure 4.1 a**). I observed a high degree of immune cell composition variation and shifts per patient's samples as well as across clinical groups in comparison to the healthy donors' samples (**Figure 4.2 a**)

I performed cell type's abundance analyses by constructing generalized linear mixed model (GLMM) to obtain statistical and predictive power to understand the association between cell type abundance and the disease clinical states. I observed significant enrichment as well as the depletion of different cell types across the clinical groups in comparison to the healthy controls (**Figure 4.2 b and Table 2).** In the BT group, I observed an enrichment of plasma cells and NK cells in line with previous studies (Ledergor et al. 2018) (Zavidij et al. 2020). I defined new population in the neutrophil compartment and termed myeloma associated neutrophils (MAN) cells. The MAN cells showed statistically significant positive estimates (p-value < 0.001) from the GLMM, indicating the high predictive power of these cells' abundance for the MM initial diagnosis state. In the non-CR group, I observed partial enrichment of plasma cells, MAN, and NK cells, which also showed statistically significant positive GLMM estimates. In contrast, the CR group showed high enrichment of CD8+ T-cells, NKT cells, neutrophils, neutrophils-t-cells, mature B-cells, common myeloid progenitors (CMP) and monocytes which showed positive GLMM estimates reflecting the predictive power of these cell types for the CR clinical state. Both CR and non-CR groups showed enrichment of dendritic cell subtypes and dendritic cell/monocyte progenitors.
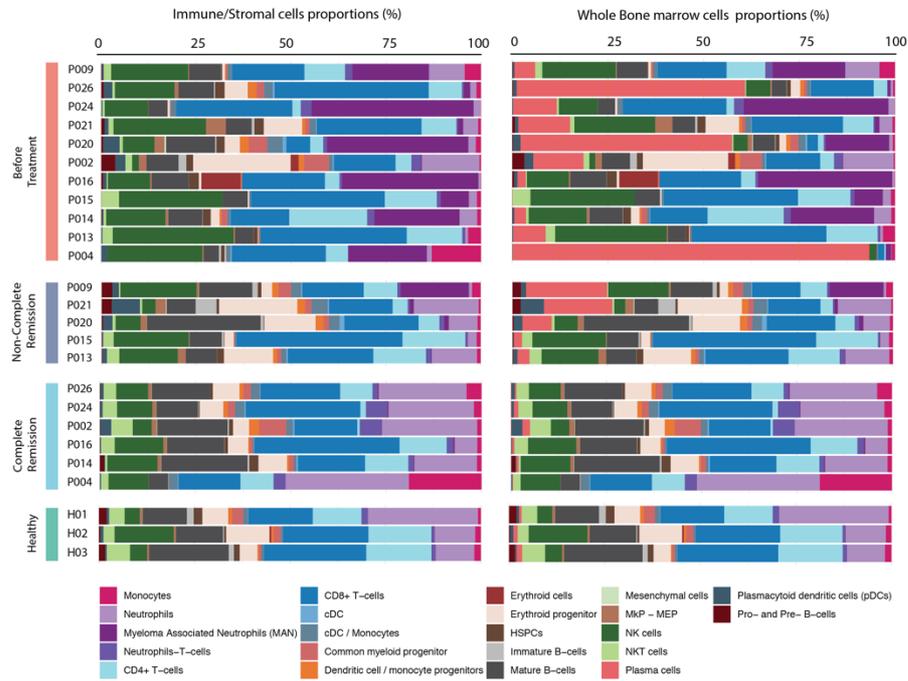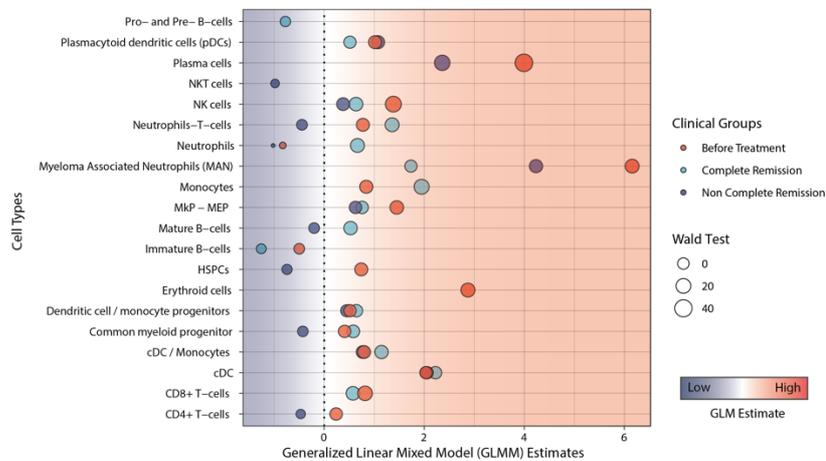
Figure 4.2: Cellular abundance shift in the bone marrow immune microenvironment before and after Long Term Survival (LTS).

a) Proportions of cell types in both healthy and MM patients before and after long-term survival for each sample. **b)** The GLMM of the cell types abundance count data. The x-axis shows the GLMM model estimates (p-value <0.001), the y-axis shows the cell types, the circles' color represents the clinical groups, and the size of the circles reflects the Wald statistical test values.

4.2.1 Tumor compartment: B cells and malignant plasma cells

Mature B cells can differentiate to plasma cells. Somatic hypermutations and chromosomal translocations can occur that lead to the manifestation of abnormal clonal plasma cells (Pawlyn and Morgan 2017). I performed clustering and dimensionality reduction on the plasma cell compartment and defined 17 heterogeneous clusters **(Supplementary Figure 6.3 a)**. The healthy donors (clusters 7 and 12) showed the co-clustering pattern of plasma cells with the CR group cells reflecting the healthy plasma cell states. In contrast, the other major clusters from BT and non-CR groups represent one individual patient reflecting the interpatient heterogeneity of the malignant plasma cell compartment (**Figure 4.1 b**).

To further define the healthy and malignant plasma cells, I performed differential expression analysis between these clusters to define the biological programs that could explain the underlying heterogeneity. I observed that each patient has a certain usage preference to the immunoglobulins heavy chains and light chains (Ig kappa and lambda) (**Figure 4.3 a**).

Moreover, I observed that the majority of the plasma cells express malignancy markers (SCD1, TNFRSF17). Besides, I observed that each patient has a certain preference to co-express other malignancy markers (CCND1, CCND2, ITGB7, FRZB, LAMP5, MAFB) as shown in a previous study (Ledergor et al. 2018), indicating that abnormal plasma cells adapt different biological programs according to the surrounding BM microenvironment. In-addition, I performed CNV inference analysis using our single cell RNA-seq data to infer the CNV status and determine the global chromosomal losses and duplications (**Figure 4.3 b**).

I ran the CNV analysis for all patients before and in LTS, and observed that the same malignant plasma cell clone at the diagnosis time of each patient is persistently present in LTS. As an example, P20 showed specific loss of chr13, partial duplication of chr14, and duplication of chr22 **(Supplementary Figure 6.3 b)** which is visible at both timepoints (BT and LTS, respectively). The majority of the plasma cells harbor both chr13 loss and chr22 duplication, which continue to be present at LTS, too. Furthermore, I performed clustering for P20 plasma cells and defined seven subpopulations which showed heterogeneous transcriptional expression states, indicating that we can detect intra patient heterogeneity in their plasma cell compartment **(Supplementary Figure 6.4 d).**

The Inter-patient and Intra-patient heterogeneity can not only be explained by the CNV status, suggesting that there are possible sources which could shape and influence the observed transcriptional heterogeneity; such as the BM microenvironment, epigenetic regulation and/or long noncoding RNA (ex: MALAT1)" expression **(Supplementary Figure 6.4 d)**.
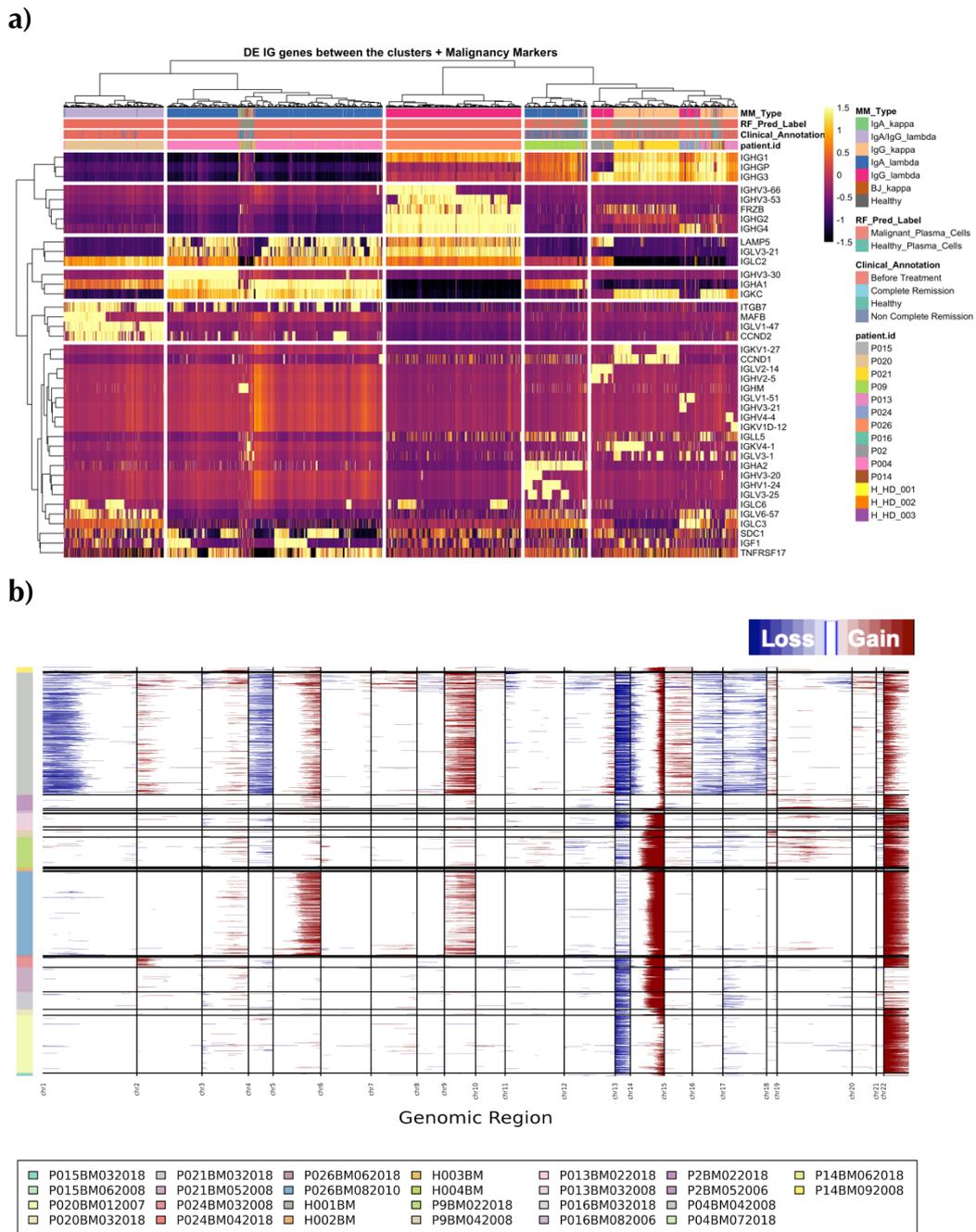
Figure 4.3: Malignant plasma cells in MM patients before and after LTS.

a) The heatmap shows the immunoglobulin usage and malignancy marker expression in the plasma cell compartment per patient sample. **b)** The heatmap shows the CNV single-cell RNA status inferred from the single-cell RNA-seq data of the MM patients' plasma compartment.

4.3 Dissecting the bone marrow microenvironment complex immune cellular states

In the following sections, I will dissect the cell types and states to define the key regulatory programs that control and co-evolve through the LTS groups (non-CR and CR).

4.3.1 The NK phenotypic expansion from healthy NK states to more diverse states in the BT group.

NK cells are a subset of innate lymphocytes that have an important role in mediating an effector cytotoxic function in the BM microenvironment. I grouped the NK cells (n= 14454 cells) into 11 major clusters (resolution parameter = 0.5) which are distributed over the clinical groups **(Figure 4.4)**. I found a statistically significant enrichment of NK cells relative abundance in both BT and CR groups to the healthy control **(Figure 4.2 b)**.

a)   **b)**
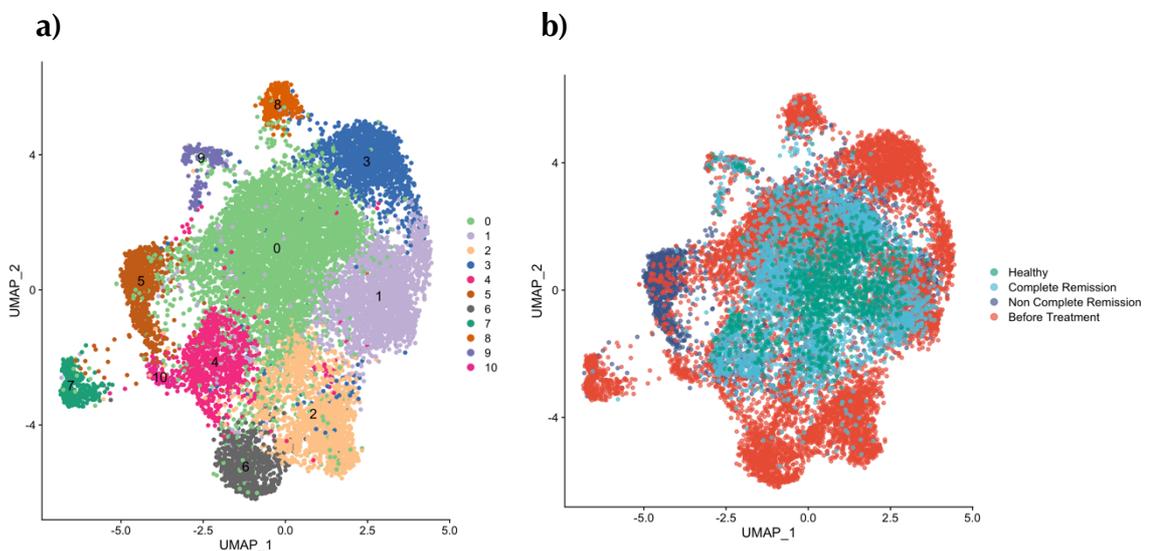
Figure 4.4: UMAP representation of the NK compartment clusters before and after long-term survival

a) UMAP representation for the 11 major clusters **b)** UMAP representation color coded by the clinical groups.

I observed NK phenotypic expansion from NK highly dense clusters in the healthy group toward more diverse states in CR and non-CR groups, with the highest diversity in the BT group **(Figure 4.5)**. To quantify this phenotypic expansion, I trained a random forest (RF) model on healthy and BT NK cells (check **Figure 3.4** in the methods section for more details). I used the trained model to predict the NK cell phenotypic composition in all clinical groups (including CR and non-CR groups).

Interestingly, I observed that non-CR and CR groups retain NK cells which have BT-like phenotype, indicating that the NK cells in the long-term survival groups are a highly dynamic population which is positioned in between the phenotypic states of the healthy and BT groups. These phenotypic states in the long-term survival groups reflect an active disease state **(Figure 4.6)**.

Figure 4.5: The density plot shows the phenotypic expansion of the NK compartment and the enrichment and depletion of the NK compartment across the clinical groups.

* UMAP representation: the axes have been omitted for simplicity. The color scale represents number of neighbors
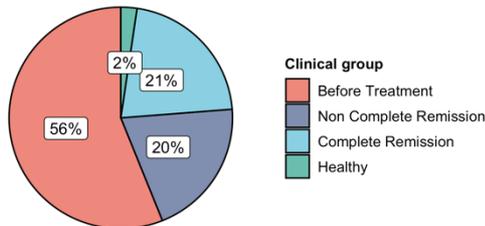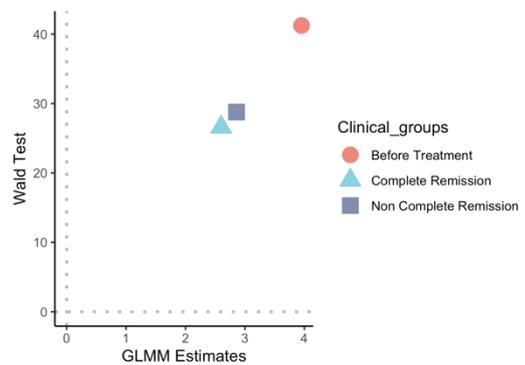
Figure 4.6: Random Forest Model prediction of BT-like and healthy-like states across all clinical groups in the NK compartment.

a) The UMAP representation shows the RF model prediction of the cellular states in the NK compartment*. **b)** The pie chart shows the proportions of BT-like cells in the different clinical groups. **c)** GLMM estimates of BT-like cells for each clinical group (p-value <0.001). *The axes have been omitted for simplicity.

4.3.2 NK cells control the tumor residual disease state via mediating high cytotoxic functions in the CR group

I performed differential expression analysis between our clinical groups and detected 41 differentially expressed genes (adj. p-value ≤ 0.05). I observed the upregulation of CXCR4 in the BT group, in line with previous study observations in MM context (Zavidij et al. 2020) **(Figure 4.7)**. In both BT and non-CR groups, I observed the upregulation of TNFA and NFKB1 inflammatory pathway genes (TNFAIP3, CD69, RGS1, KLF6, and NFKBIA) and downregulation of the main cytolytic effector molecules (GNLY, GZMB, NKG7, PRF1, and KLRD1) in NK cells (Duhan et al. 2019) **(Figure 4.8 a)**. Furthermore, I used these genes to calculate the overall biological program score for every single cell and constructed a predictive GLM model for these scores.
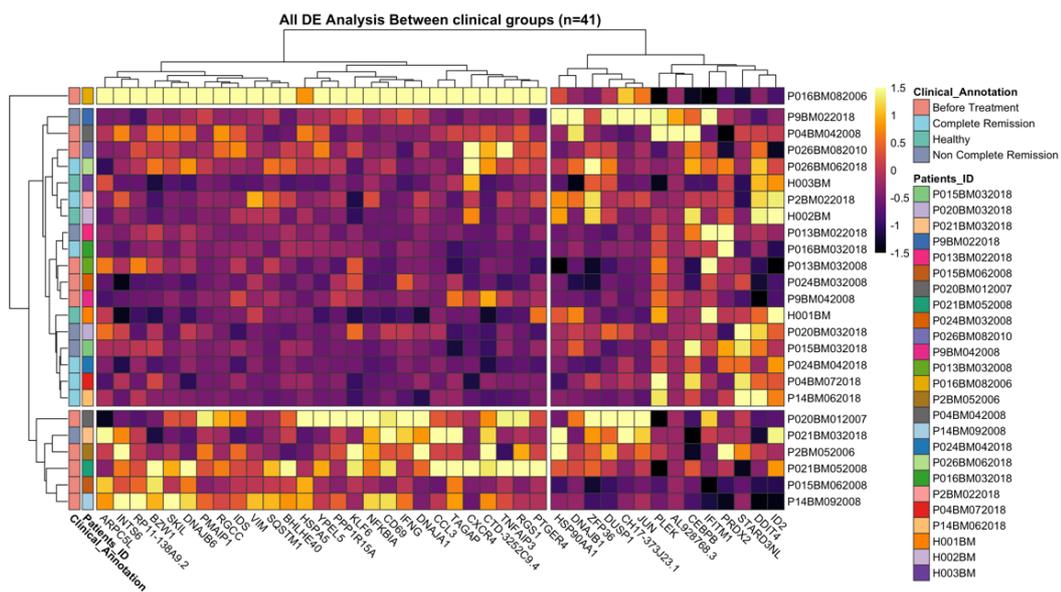
Figure 4.7: The heatmap shows the DE genes between the clinical group in the NK compartment cells across the patients' samples.

* The color scale reflects scaled values of the genes expression (z-score). The metadata is indicated by the color codes on the right-hand side

a)
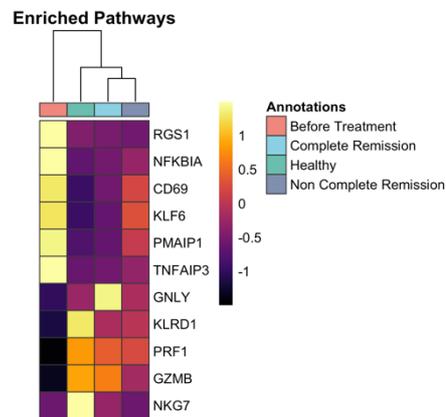
b)
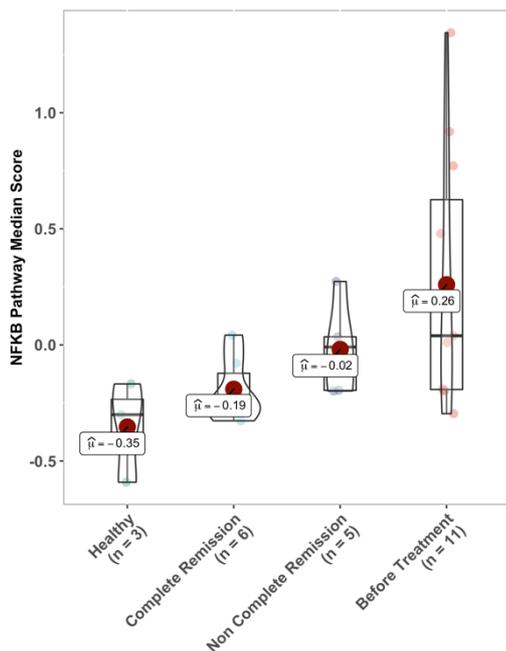
c)
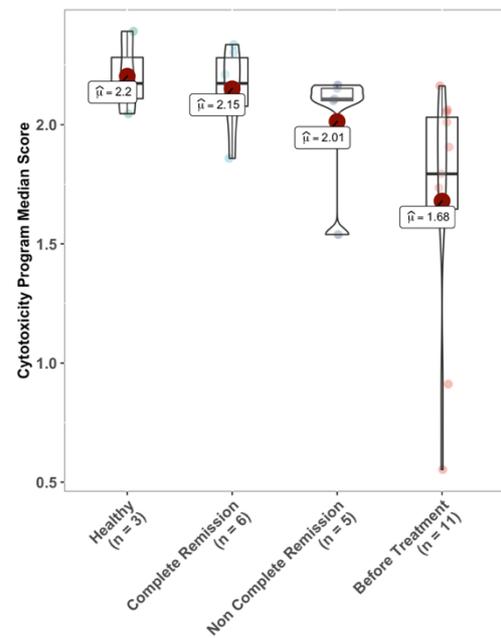
Figure 4.8: NFKB and inflammatory pathway scores and NK cytotoxicity program scores in the NK compartment

a) The heatmap shows the expression of NFKB and inflammatory pathway and NK cytotoxicity genes across the clinical groups. **b) c)** Median scores of NFKB and inflammatory pathway, and NK cytotoxicity program scores across clinical groups, respectively.

I observed a strong negative correlation between both scores in the BT group indicating that BT-NK cells induce a proinflammatory signal, and consequently, a lower NK cytolytic activity in the bone marrow microenvironment (**Figure 4.8 b and c, and Supplementary Figure 4.8).**

In contrast, both CR and healthy groups showed upregulation of cytolytic effector markers, indicating that both groups harbor higher NK cytotoxicity function and lower activity of NFKB1 inflammatory pathway. The non-CR group showed lower expression of GZMB, FCER1G, and GNLY, showing that the non-CR group retains a lower cytolytic activity than the CR group (**Figure 4.8 b, c and Figure 4.9).**
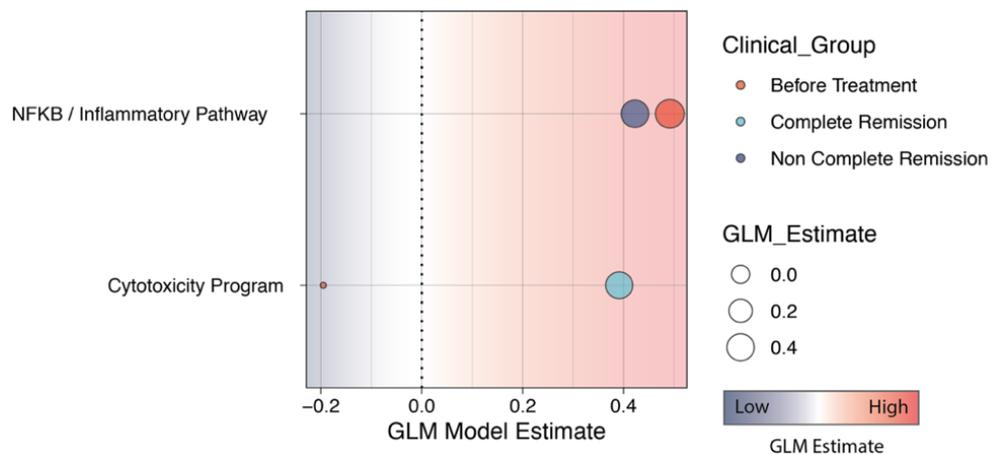
Figure 4.9: GLM estimates of NFKB and inflammatory pathway scores and NK cytotoxicity program single-cell scores per clinical group.

The circle colors represent the clinical groups and the size of the circle reflects the GLM model estimate values (p-value < 0.001).

I performed receptor-ligand interaction analysis to decipher cell-cell crosstalk patterns and observed common NK interaction partners like NK co-activating (CD2-CD58) (Rölle et al. 2018) partner, which is expressed in all clinical groups between NK cells and other cell types in the BM microenvironment.

More specifically, I observed that in the CR and healthy groups, the stimulatory interaction partner (CD94:NKG2C heterodimer - HLA-E) has a high interaction score between NK cells and many other cell types including CD8+T-cells, CD4+ T cells and mature B cells; reflecting a high NK activation state in the CR group (Pittari et al. 2017) **(Figure 4.10)**.

The non-CR group loses the stimulatory interaction partner (CD94:NKG2C heterodimer - HLA-E) and keeps upregulating the inhibitory interaction partner (CD94:NKG2A heterodimer - HLA-E) (Pump et al. 2019) **(Supplementary Figure 6.6)**. The BT group loses both (CD94:NKG2C heterodimer - HLA-E) and (CD94:NKG2A heterodimer - HLA-E) interaction partners and upregulate (TNFRSF1B - GRN) and (IFNG Type II - IFNR) interaction partners between NK cells and other cell types in the BM microenvironment. Such cell-cell crosstalk in the BT group would induce a negative impact on the NK's cytotoxicity functions and potentially mediate an immunosuppression state allowing myeloma cell growth and proliferation (Almishri et al. 2016, 2).
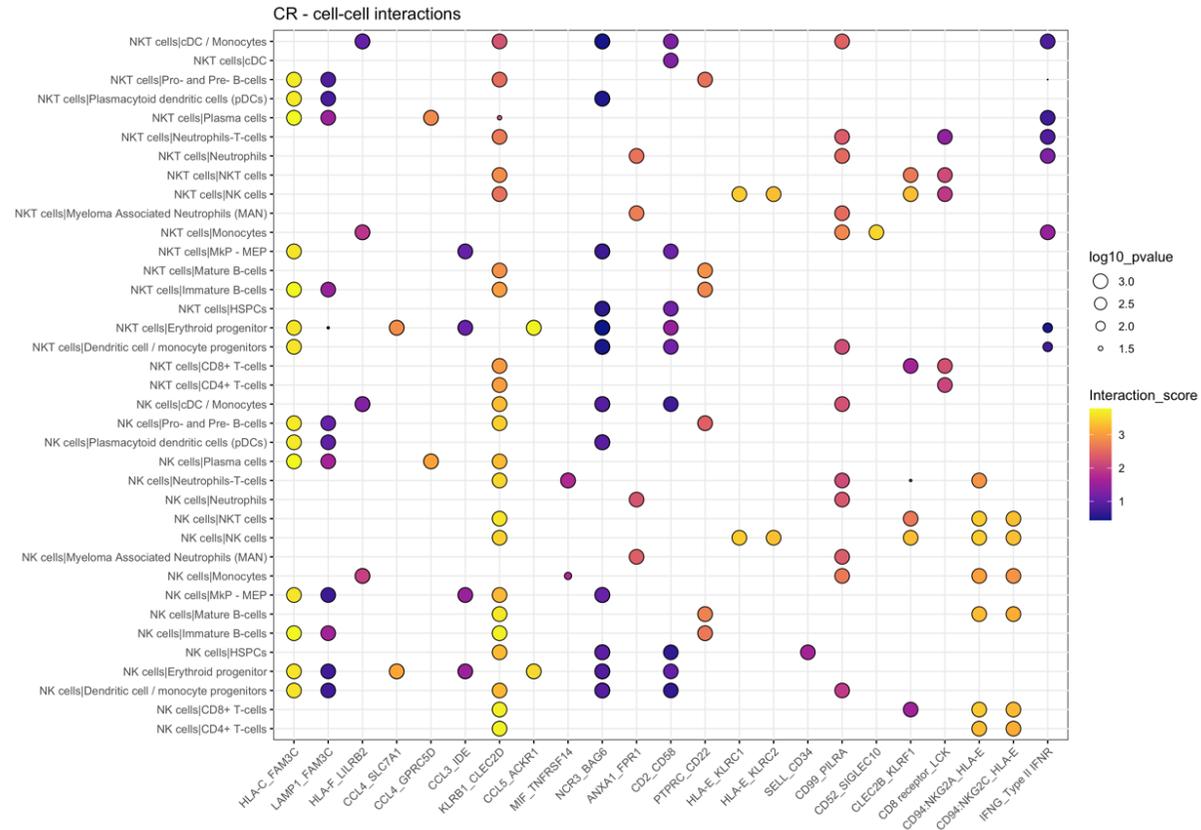
Figure 4.10: Balloon plot shows the Receptor-Ligand (R-L) interaction patterns between the NK cells and other cell types in the CR group bone marrow microenvironment.

The color scale represents the mean receptor-ligand interaction scores. The size of the circle represents the (-Log10) of the p values

4.3.3 T cell cellular states and phenotypes in the bone marrow microenvironment of MM patients.

T cell abundance and presence in spatial proximity to the malignant cell have been shown in many tumor entities (Binnewies et al. 2018). This spatial proximity has been correlated to better clinical outcomes when T cells harbor and retain certain phenotypic states and functions to elicit an immune response against malignant cells (van der Leun, Thommen, and Schumacher 2020). To gain an in-depth understanding of the T cells phenotypic state before and after MM LTS; we sorted CD3$^+$ T cells from the same BM samples of our cohort (was performed by Dr. med. Raphael Lutz) and performed single-cell RNA-seq (10x genomics). After QC and filtering out low-quality cells, I merged both the BM T cell and the sorted CD3$^+$ T cells single-cell RNA-seq data into one data object, and proceeded with the downstream analysis and T cell subtypes annotation.

4.3.4 CD8$^+$ and CD4$^+$ T cell heterogeneous cellular states and subtypes

Based on the clustering of the single-cell RNA-seq data, I defined 8 subtypes of CD8$^+$ T cells (n= 101975 cells) and 3 subtypes of CD4$^+$ T cells (n= 40821 cells) **(Figure 4.11)**. Interestingly, I defined a new population and termed aberrant memory-cytotoxic (AMC) CD8$^+$ T-cells, which is enriched in MM initial diagnosis state **(Supplementary Figure 6.7)**. I observe a heterogeneous distribution of the T cell subtypes between the clinical groups **(Supplementary Figure 6.7 and Figure 6.8)** and across samples before and after LTS **(Figure 4.12 a)**.

a)
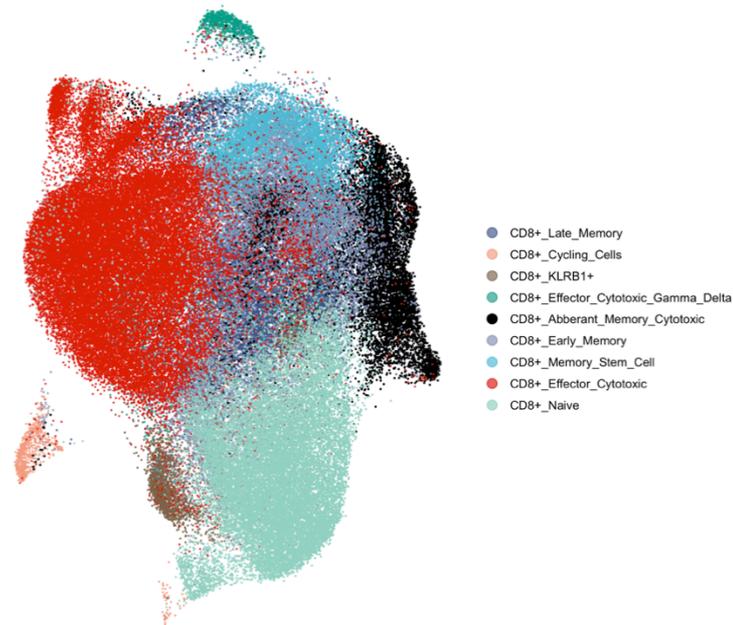
b)
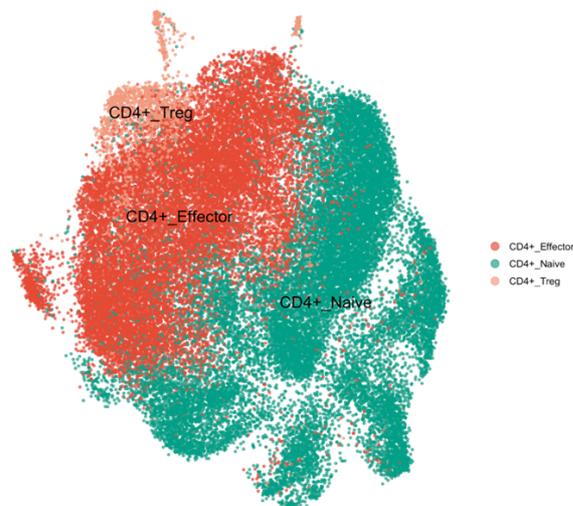
Figure 4.11: UMAP representation of the T cell subtypes in all clinical groups.

a) UMAP visualization of the integrated bone marrow and sorted CD8[+] T cells showing the subtypes of CD8[+] T-cells. **b)** UMAP visualization of integrated bone marrow and sorted CD4[+] T cells showing subtypes of CD4[+] T-cells. * The axes have been omitted for simplicity.

4.3.5 Compositional shifts in CD8$^+$ and CD4$^+$ subtypes across the MM patients

I performed cell type abundance analyses by generalized linear mixed model (GLMM) and observed significant relative enrichment as well as depletion of cell types across the clinical groups in comparison to the healthy controls **(Figure 4.12, Figure 4.13 and Table 3).** In the BT group, I observed an enrichment of aberrant memory-cytotoxic (AMC) and effector cytotoxic γδ CD8$^+$ T cells which obtained statistically significant positive GLMM estimates (p-value < 0.001) **(Figure 4.12 b)**.

In the non-CR group, I observed an enrichment of early and late memory CD8$^+$ T cells, T_{regs} and effector CD4$^+$ T cells, and partial enrichment of AMC CD8$^+$ T cells which also showed statistically significant positive GLMM estimates (p-value < 0.001). In contrast, the CR group showed high enrichment of naïve and effector cytotoxic CD8$^+$ T- cells and showed statistically significant positive GLMM estimates, indicating that the CR group has a high abundance of effector cytotoxic CD8$^+$ T cells.

4.3.6 T cell hallmark pathway analysis shows that the CR group harbor high cytotoxic functions while BT and non-CR harbor more inflammatory and exhaustion signatures

To characterize the phenotypic states of the T cell subtypes. I calculated global signature scores for each clinical group using T cell hallmark signatures and constructed GLMM to estimate the predictive power of these biological programs to the clinical state **(Figure 4.13 and Supplementary Figure 6.9)**.
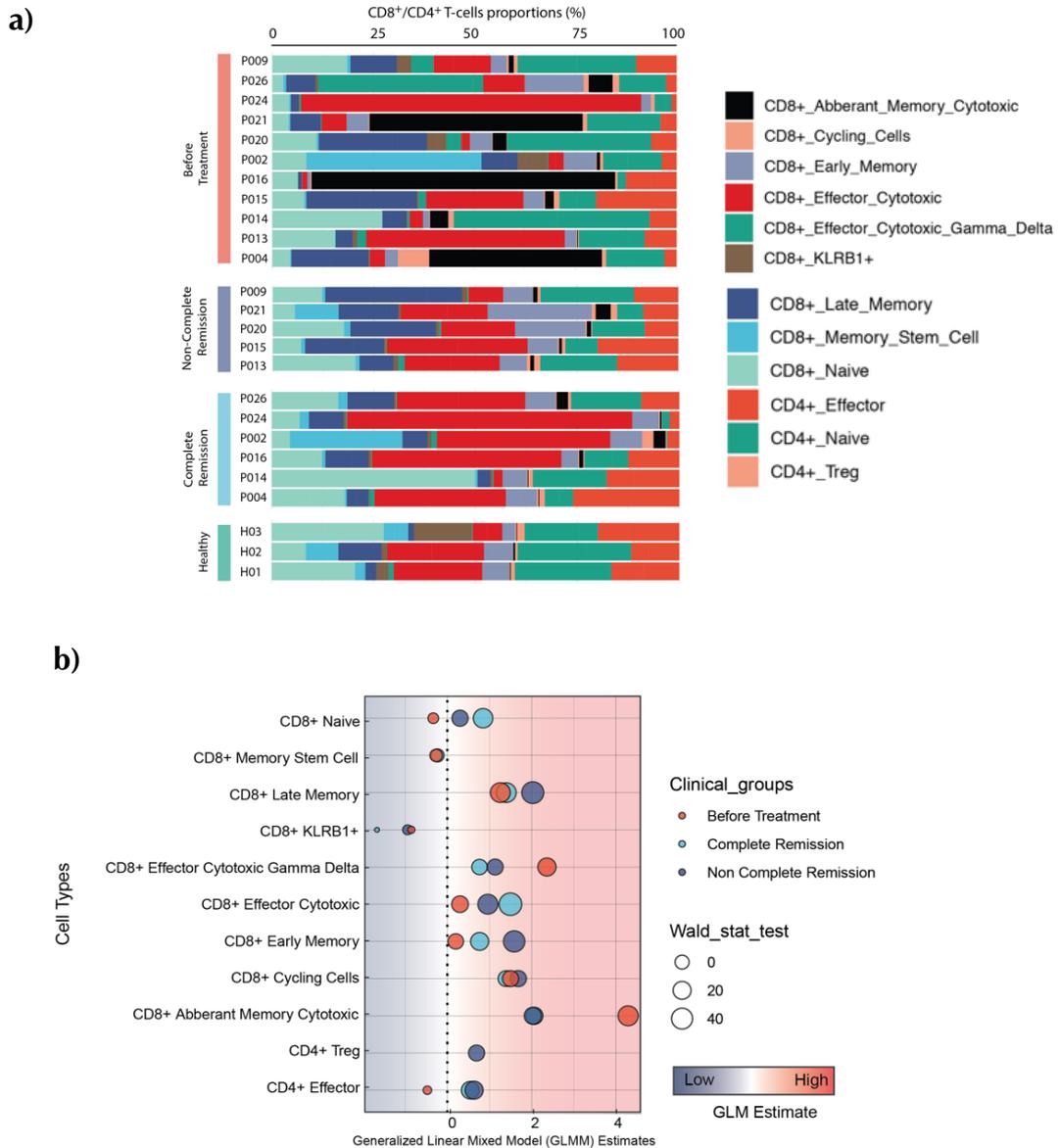
Figure 4.12: compositional shifts and alterations across the clinical groups in the T cell compartment subtypes

a) Proportions of the T cell subtypes in both healthy and MM patients before and after long-term survival for each sample. b) Estimates from the GLMM of the cell type abundance count data: The x-axis shows the GLMM model estimates (p-value <0.001), the y-axis shows the T cell subtypes, the circle color represents the clinical group, and the size of the circles reflects the Wald Test values (p-value < 0.001).

Both BT and non-CR groups showed significant upregulation of T cell dysfunction and exhaustion signature, TNFA signaling via NFKB, hypoxia pathways and apoptotic processes. In contrast, I observed an upregulation of T cell activation and the cytotoxic marker ITGB1 (CD29) (Nicolet et al. 2020) in the CR group's T cells (both CD8$^+$ and CD4$^+$ subtypes) and in the non-CR group (CD4$^+$ Subtype). Therefore, this finding supports the notion that the malignant plasma cells are under a strong immunosurveillance state, which controls the disease state, fate, and progression (Figure 6.9).

a)
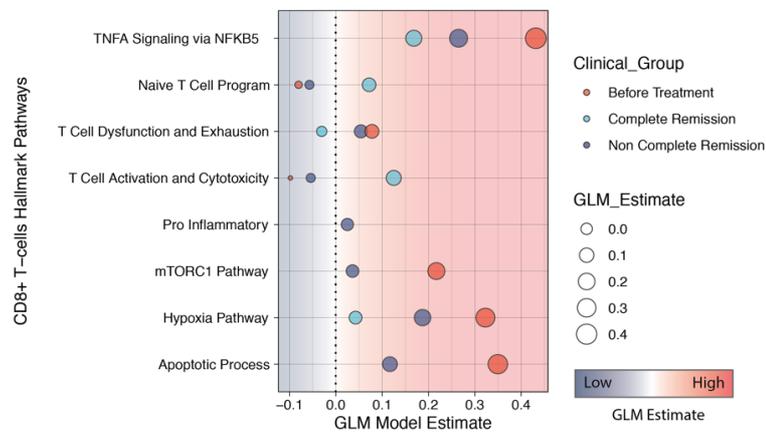
b)
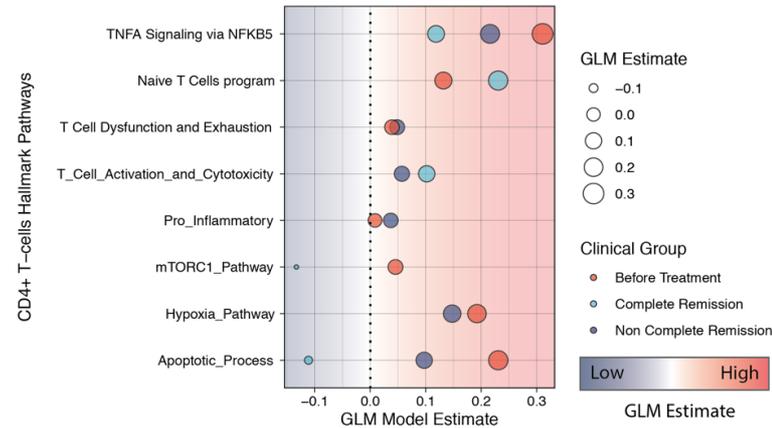
Figure 4.13: GLM model estimates of the T cell hallmarks pathways for single-cell scores per clinical group in the T cell compartment a) CD8$^+$ and b) CD4$^+$ T cell subtypes.

4.3.7 Naive CD8$^+$ T cells of the LTS group retain memory-like features in the active disease state

Naive CD8$^+$ T- cells have an astonishing capacity to interact with foreign antigens and pathogens and differentiate to memory and cytotoxic T cells, which can drive potent immunogenic responses. Underlying transcriptional programs with unique features are characterizing each of the clinical groups on the Naïve CD8$^+$ T cells landscape **(Figure 4.14)**. BT, non-CR, and CR groups showed similar major trajectory patterns as well as specific branching and "transcriptional clones" for each of these groups. Interestingly, I observed that CR and non-CR groups follow a trajectory like BT group, indicating that these cells retain footprints of memory-like features from the disease's history. These memory-like features are an interesting observation, since the CR and non-CR groups are sampled several years after the autologous stem cell transplantation (~ 9 years). Moreover, the healthy naive CD8$^+$ T cells showed two specific transcriptional clones which are mostly depleted from the other clinical groups.

I performed module analysis and detected specific modules of co-expressed genes that regulate the naïve transcriptional states between the clinical groups. The CR group showed upregulation of certain gene regulatory modules (Modules 2, 3, 6,7, and 13), and showed enrichment of mTORC1 signaling and IL2-STAT5 signaling indicating a T cell quiescence exit (QE) state (**Supplementary Figure 6.10 and Supplementary Figure 6.11**). In contrast, BT and non-CR groups showed downregulation of the CR group-specific modules and upregulation of certain modules (Modules 1,8,17,4,9,16,11 and 14) and specific surface markers and transcriptional factors indicating that these cells are primed towards a high activation state.
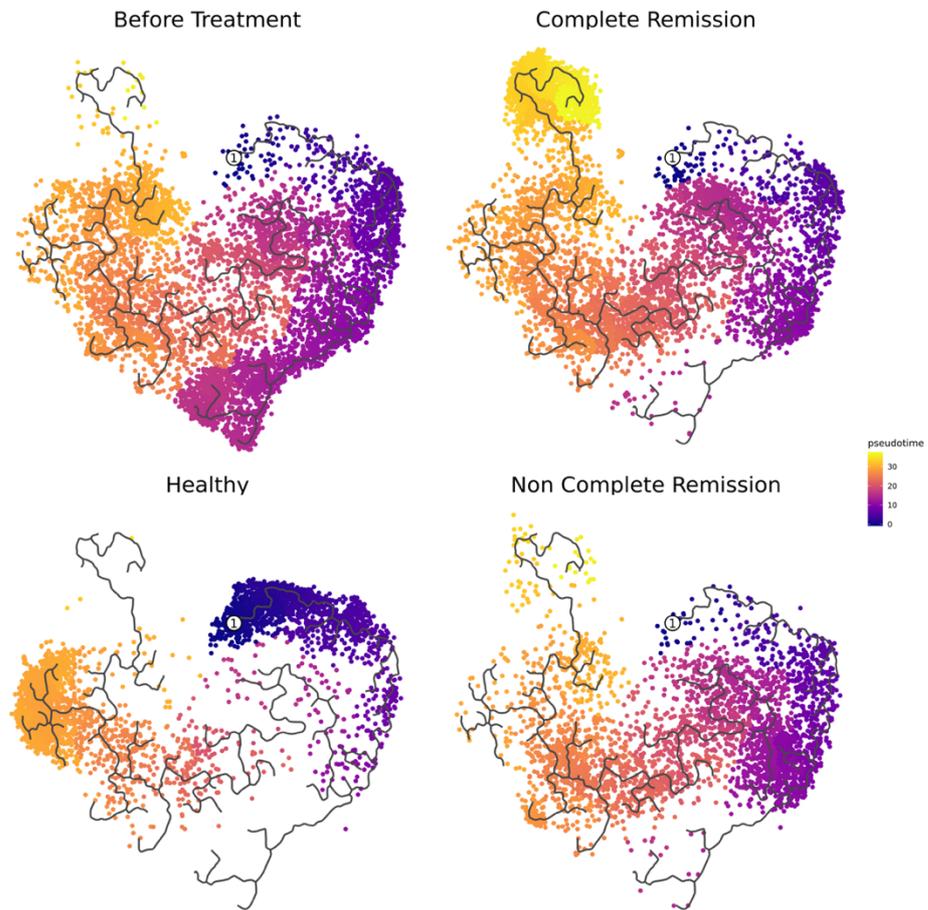
Figure 4.14: Pseudotemporal ordering of the naive CD8+ T cells transcriptional states of each clinical group.

I investigated the genes underlying the modules, and observed that there is a co-expression pattern of CCR7, ADGRE5 (CD97), and CD44 surface markers and many transcriptional factors that co-regulate the naïve T cell state in the BT group **(Supplementary Figure 6.10 c)**. ADGRE5 (CD97) is an early activation marker and its interaction with CD55 showed a strong costimulatory signal to the T cell (Spendlove and Sutavani 2010). Moreover, CD44 is upregulated after T cell activation (Baaten, Li, and Bradley 2010), suggesting that both expressions of CD97 and CD44 are indicative of an activation state in the BT and non-CR naïve CD8$^+$ T cell compartment. Many crucial TFs like ARID5A, CREM, HIF1A, and SON are regulating the BT and non-CR naïve T cell states. Upregulation of ARID5A, CREM, HIF1A, and SON supports the notion that these cells are in a highly activated state (Zaman et al. 2016, Fang et al. 2015, Phan and Goldrath 2015, Ahn et al. 2011). Besides, the non-CR group showed an activation state by upregulating many activation surface markers CD69 (Ziegler, Ramsdell, and Alderson 1994) and KLRD1 (CD94) (Spendlove and Sutavani 2010)).

I could argue that the non-CR group activation state is relatively different from the BT activation state. The non-CR naive state showed upregulation of many genes like GZMK and GZMA indicating an immune surveillance role (Arias et al. 2017). Furthermore, IL10RA expression in the non-CR naive state reduces any potential tissue damage as a side effect of inflammatory cytotoxic activity (Ouyang and O'Garra 2019), and the upregulation of PIM-1 expression promotes cell survival within this state (Peperzak et al. 2010).

4.3.8 CD8⁺ T cell global differentiation models in MM before and after long-term survival

Several models have been proposed to explain the emerging memory and effector cytotoxic CD8⁺ T cell populations after T cell activation (Kaech and Cui 2012b). In the MM field, we do not have such models which could explain the mechanisms of CD8⁺ T cell differentiation in the diseased bone marrow of MM patients. To explore this question, I ran trajectory analyses (see Trajectories and developmental process inference methods section) for CD8⁺ T cell subtypes for the clinical groups and I observed striking differentiation patterns.

In the healthy CD8⁺ T cell context, the differentiation processes follow a Linear Cell Fate Model where the trajectory starts with a naive state followed by a memory state and ends with cytotoxic states **(Figure 4.15).** However, in BT and non-CR groups, the differentiation processes can be best described by the Continuum Cell Fate Model where naïve CD8⁺ T cells have the capacity to differentiate simultaneously to memory and cytotoxic CD8⁺ T cells from an early differentiation point of the trajectory. Moreover, I observed a global developmental shift in the differentiation order, where BT and non-CR trajectories end with memory states in contrast to the healthy trajectory which ends in a cytotoxic state **(Figure 4.16 and Supplementary Figure 6.12).** In the CR CD8⁺ T cell context, the differentiation processes are in line with Partial-Linear Cell Fate Model which looks similar to the healthy trajectory and ends with a cytotoxic state **(Supplementary Figure 6.13).**

a)

b)

c)
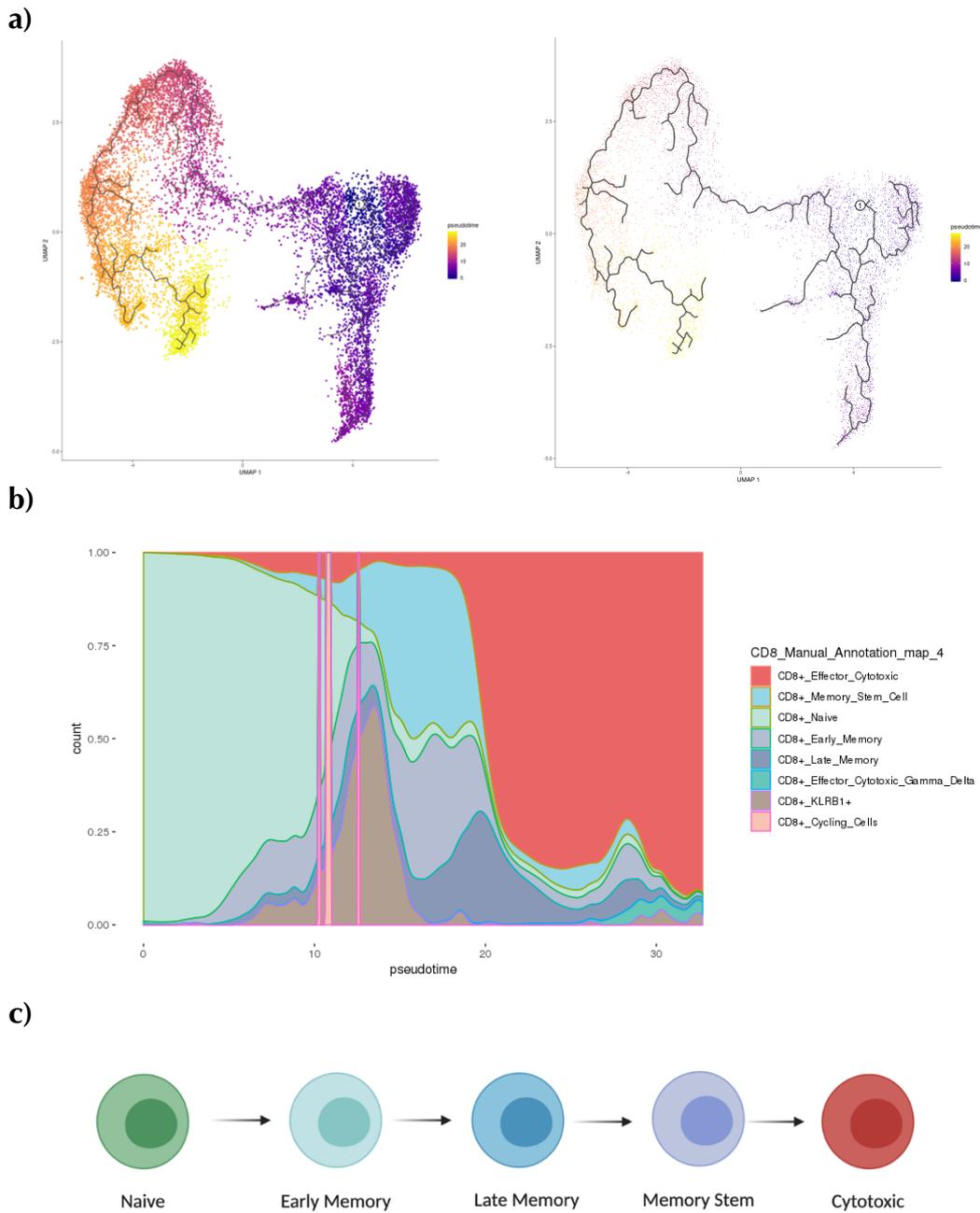
Figure 4.15: Healthy CD8+ T cells follow a linear cell-fate differentiation trajectory

a) UMAP representation shows the pseudotime assignment of the cells over the developmental trajectory. b) The proportions of the CD8+ T cell subtypes over the assigned pseudotime. c) Graphical summery represents the Linear Cell Fate Model.

a)

b)
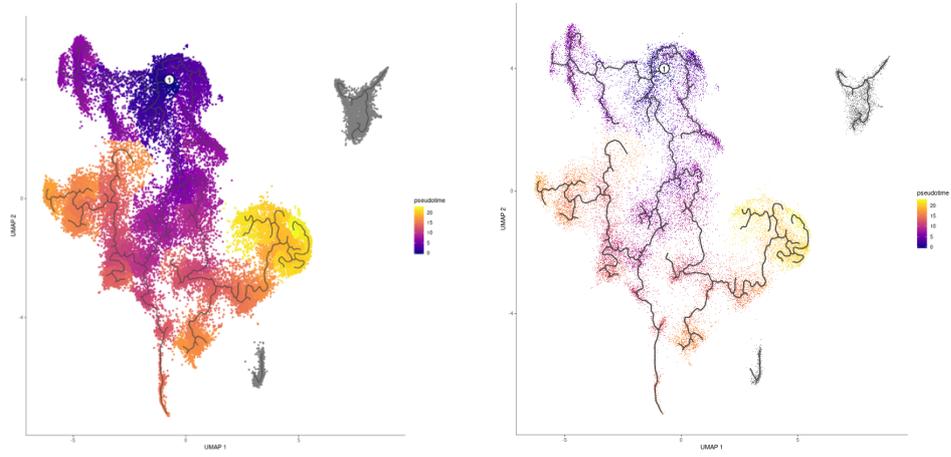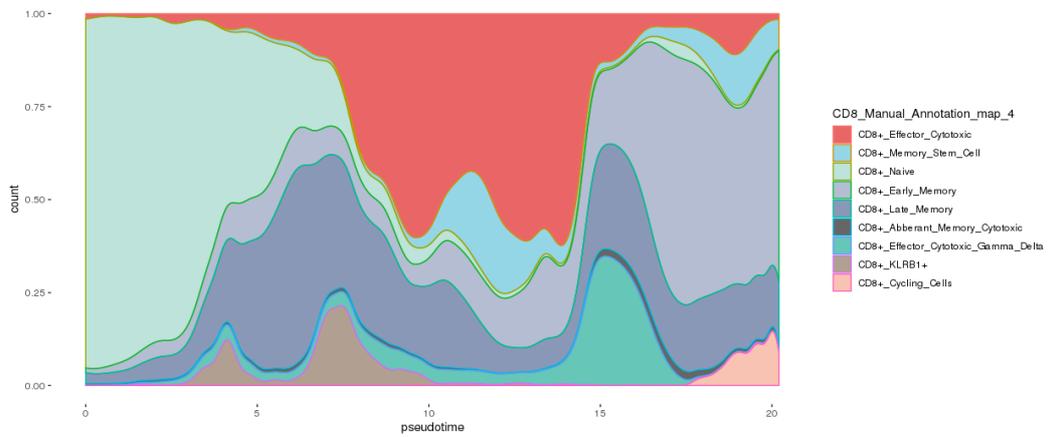
Legend next page …
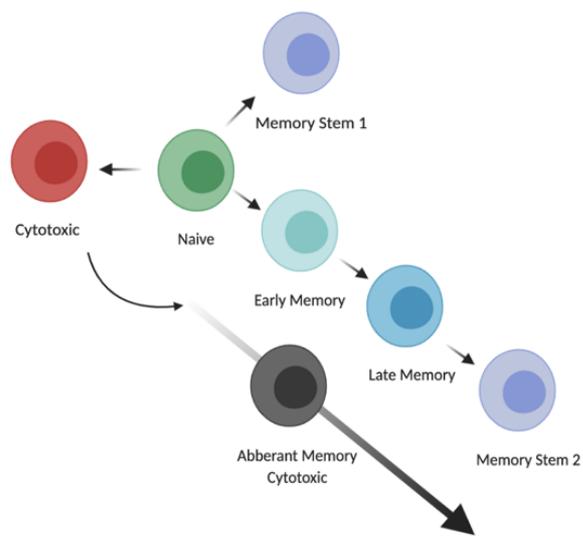
Figure 4.16: Before Treatment (BT) CD8$^+$ T cells follow a continuum cell-fate differentiation trajectory

a) UMAP representation shows the pseudotime assignment of the cells over the developmental trajectory. b) The proportions of the CD8+ T cell subtypes over the assigned pseudotime. c) Graphical summery represents the Continuum Cell Fate Model in BT group.

4.3.9 Global disease-state CD8+ T cell markers

I observed that there are certain surface markers, TFs, and pathways constitutively expressed in most of the CD8+ T cell subtypes within each of the clinical groups. These biological markers could be disease-State specific markers rather than being specific for each of the CD8+ T cell subtypes. These markers could have the potential to be used as a disease state predictive marker which needs to be biologically and clinically validated.
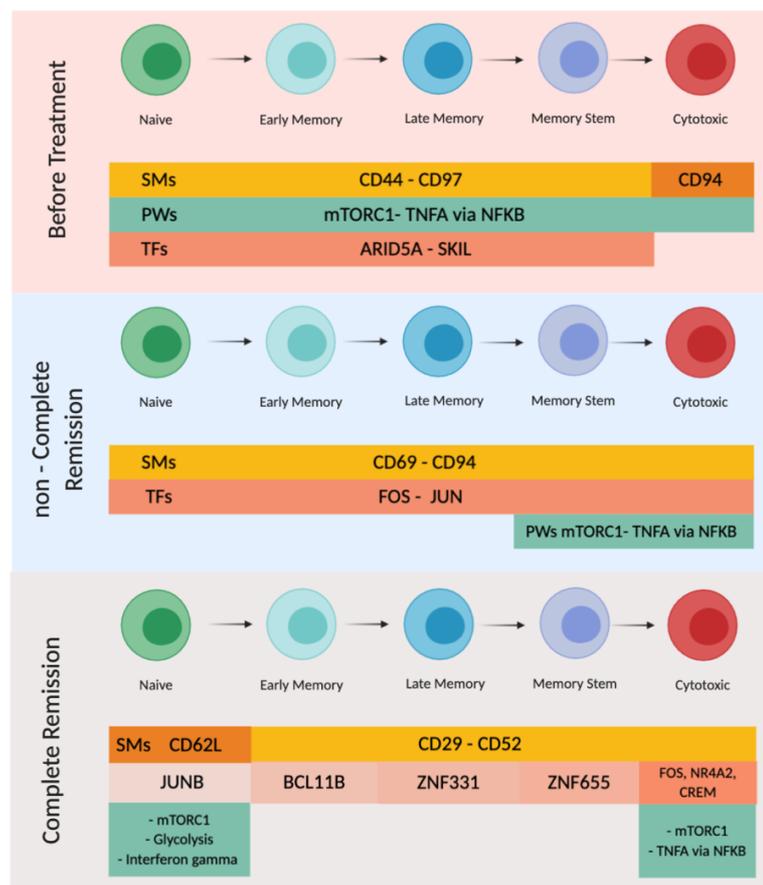
Figure 4.17: Immunophenotypic summary of CD8+ T cell compartments in Multiple Myeloma patients' BM before and after long-term survival.

SMs: Surface Markers - **PWs:** Pathways - **TFs:** Transcriptional Factors

4.3.10 The neutrophil heterogeneous transcriptional landscape in MM LTS and the definition of a new population: Myeloma Associated Neutrophils (MAN)

Neutrophils are the most abundant leukocytes in the circulation and have an important role in modulating the BM immune microenvironment and regulating the adaptive immune response (Rosales 2018). MM patients with high neutrophil/lymphocyte ratio (NLR) are more likely to have a poorer prognosis (Mu et al. 2018) ʻ(Onec et al. 2017). It has been shown that high-density neutrophils are dysfunctional and immunosuppressive in MGUS and MM (Romano et al. 2020). However, the role of neutrophils in MM long-term survival is unclear.

To define the phenotypic states and subpopulations in the neutrophil compartment, I extracted the neutrophil lineage single-cell RNA-seq data (n= 18118 cells) and performed dimensionality reduction, graph-based clustering and defined 11 clusters (resolution parameter = 0.5) distributed over the clinical groups **(Supplementary Figure 6.14)**. Furthermore, I defined two main neutrophils subtypes; normal neutrophils (NN) and myeloma associated neutrophils (MAN) **(Figure 4.18).** Interestingly, I found that NN clusters are highly enriched in the healthy and CR groups, while MAN cell clusters are highly enriched in the BT and partially enriched in the non-CR groups **(Figure 4.18 b and Supplementary Figure 6.15)**.

Additionally, MAN cells showed high positive GLMM estimates indicating the significant predictive power of this population abundance for the clinical group states **(Figure 4.2 b and Table 2).** I observed that MAN cellular states between patients show lower correlation structures in comparison to the NN cells patients' samples **(Figure 4.19)**; indicating that MAN cells are more phenotypically diverse between our patients' samples.

a)
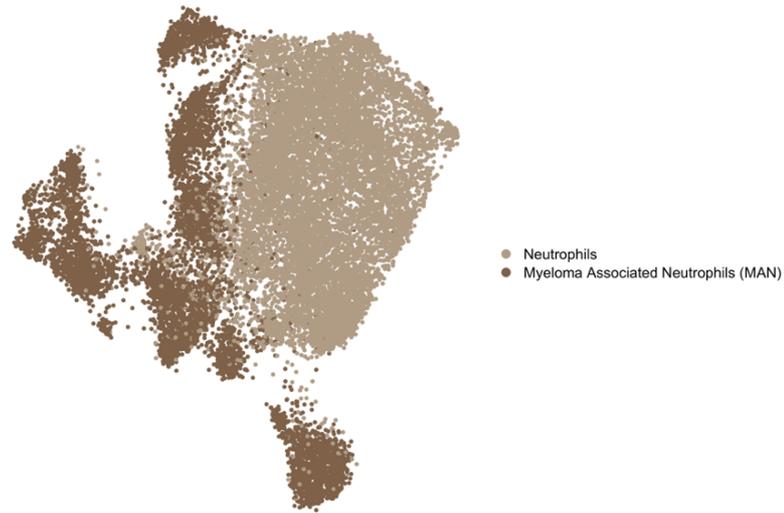
b)
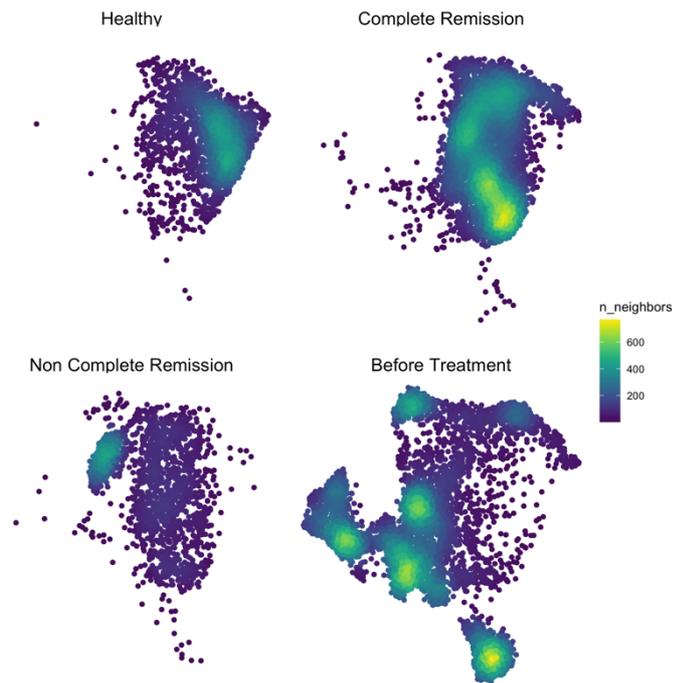
Figure 4.18: Neutrophil transcriptional landscape in all clinical groups and states.

a) UMAP representation of normal neutrophils (NN) and Myeloma Associated Neutrophils (MAN). **b)** Density plot* shows the enrichment and depletion of the neutrophil compartment across the clinical groups. *UMAP representation: the axes have been omitted for simplicity. The color scale represents number of neighbors

I developed a random forest (RF) classifier to quantify this phenotypic expansion and decipher the cellular states per clinical group (check **Figure 3.4** in the methods section for more details). Interestingly, I found that both non-CR and CR groups harbor BT-Like neutrophil cells **(Figure 4.20 a)**. The GLMM showed a significant positive estimate for BT, non-CR, and CR groups (p-value < 0.001) **(Figure 4.20 b and c)**; indicating that BT-like neutrophils have good predictive power with respect to the clinical state, suggesting a potentially important role in the disease state control and progression.

a)

b)
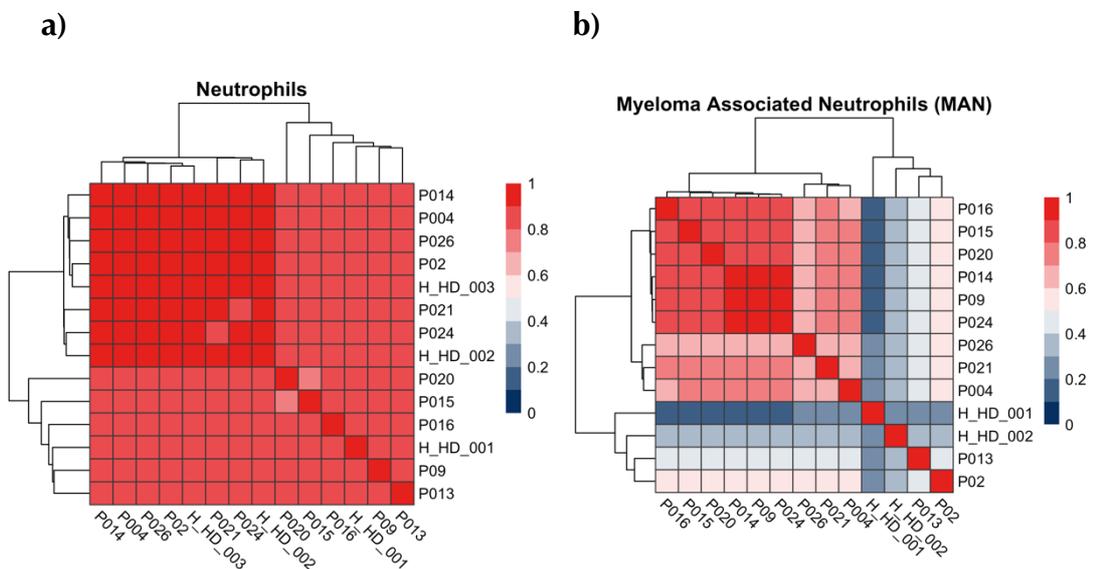
Figure 4.19: The Heatmap shows the correlation coefficient similarity between patients' samples in both a) NN and b) MAN cells.
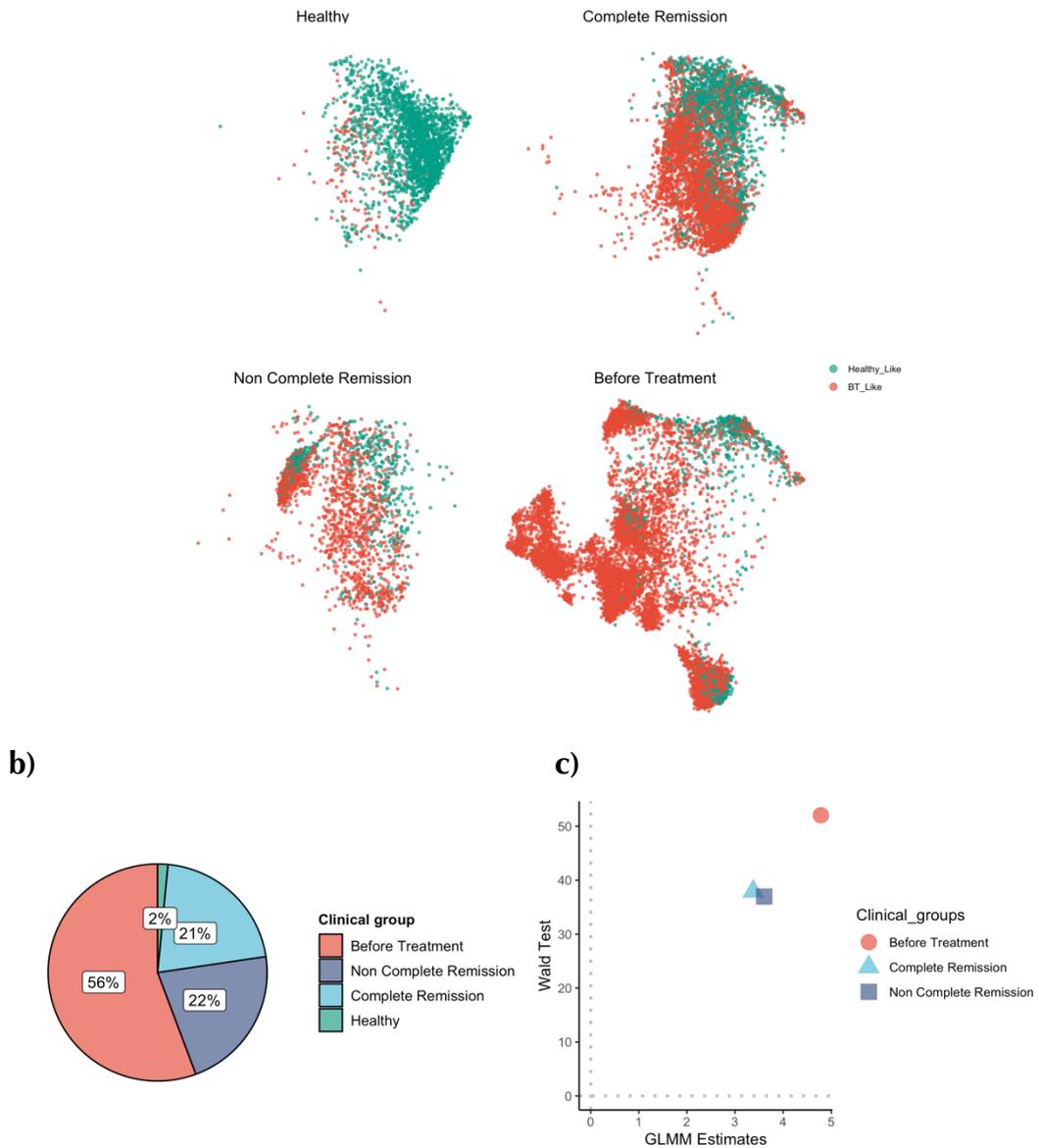
Figure 4.20: Random Forest Model prediction for BT-like and Healthy-like states across all clinical groups in the neutrophil compartment.

a) UMAP representation shows the random forest Model prediction for before treatment-like and healthy-like states across all clinical groups in the neutrophil compartment. **b)** The pie chart shows the proportions of BT-like cells in different clinical groups. **c)** GLMM estimates of BT-like cells for each clinical group (p-value <0.001).

4.3.11 MAN cells retain a migration phenotype and induce pro-inflammatory and immunosuppressive signals in both BT and non-CR groups.

I performed differential expression analysis between the neutrophils subtypes and found 118 upregulated genes which are specific for MAN cells. Furthermore, I performed GSEA **(Figure 4.21 and Supplementary Figure 6.16)** and observed an upregulation of many pathways and biological programs in MAN cells related to neutrophil migration, chemokine and cytokine signaling pathways and proinflammatory responses.

I found that IL1B is highly upregulated in MAN cells, a potent proinflammatory cytokine and upstream of IL-6, which supports myeloma cell growth (Lust and Donovan 1999) and induces an immunosuppressive signal in the BM microenvironment (Kaplanov et al. 2019). Besides, many other genes which induce proinflammatory signals (CD83, PLEK, IFNGR2) and activate the NFKB inflammatory pathway (BID, BCL2A1, and ANKRD28) (Aerts-Toegaert et al. 2007, 8, Lundmark et al. 2015, T. Liu et al. 2017). In-addition, I observed that CXCL8 family genes (CXCL8, CXCL2, CXCL3) (Oliveira, Rosowski, and Huttenlocher 2016), GPR183, SOD2 (Zhou et al. 2018), and MARCKS (Wang 2018) are upregulated in MAN cells which regulate and induce neutrophil migration. Furthermore, I used these genes to calculate the overall biological program score for every single cell and constructed a predictive GLM model for these scores (see Gene set enrichment analysis (GSEA) and biological program scoring: Methods section). I observed an upregulation of the neutrophil migration and proinflammatory signal program genes in BT and non-CR groups **(Figure 4.21 b)**. The GLMM model showed positive estimates for the BT and non-CR groups **(Figure 4.21 c)** indicating that both groups, harbor immunosuppressive and pro-inflammatory signals in comparison to the CR and healthy groups, which would be one of the key phenotypic differences between CR and non-CR group immune states.

a) Heatmap shows the neutrophils migration, pro-inflammatory and immunosuppressive responses of neutrophil compartment cells across the patients' samples. **b)** Median scores of the neutrophil migration and proinflammatory program across clinical groups. **c)** GLM estimates of the neutrophil migration and pro-inflammatory program across clinical groups (p-value < 0.001).

Figure 4.21: The MAN cells shows a migration phenotype and upregulation of pro-inflammatory and immunosuppressive signatures

4.3.12 Dendritic cells (DCs) states and subtypes across MM clinical groups

Dendritic cells (DCs) are potent antigen-presenting cells (APCs) that have an important role in activating and regulating T cell antitumor activity (Fu and Jiang 2018). However, their role is contradictory in MM context where they can activate CD8$^+$ T cells to act against myeloma cells, but also, DCs can protect myeloma cells from CD8$^+$ T cell cytotoxic effect (Leone et al. 2015)'(Vo et al. 2018). In our data, I detected and defined three DCs subpopulations (n=2834 cells); conventional dendritic cells (cDC), Myeloid-derived Dendritic cells (mdDCs), and plasmacytoid Dendritic cells (pDCs) **(Figure 4.22)**.
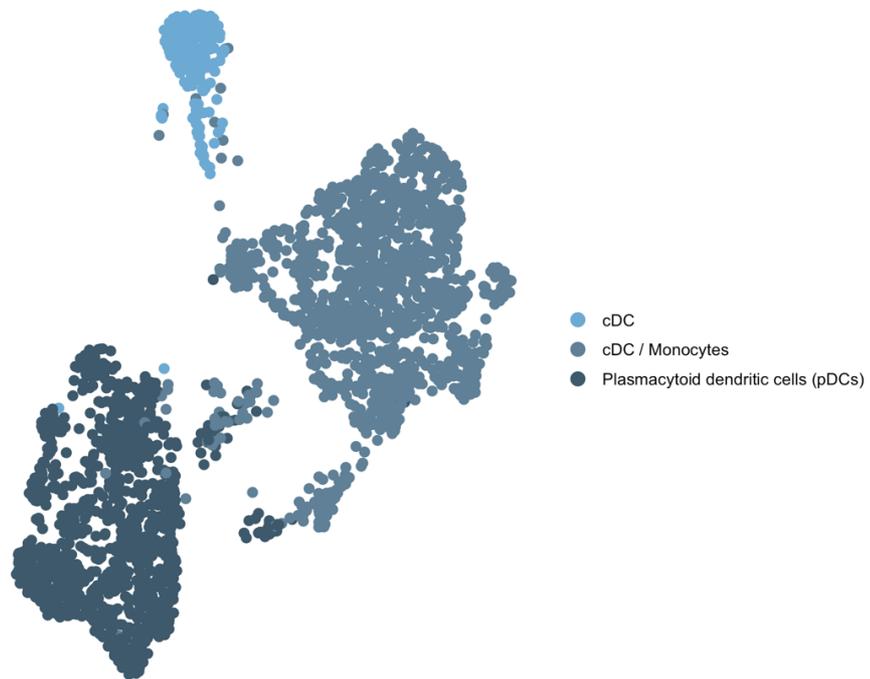
Figure 4.22: UMAP representation of DCs subtypes in all clinical groups and states.

4.3.13 mdDCs induce higher IFG signals in the BM and mediate immunosuppressive crosstalk with MAN cells in both BT and non-CR groups.

I observed significant enrichment of mdDCs in the CR group and significant enrichment of pDCs in BT and non-CR groups **(Figure 4.23 a and Figure 4.2)**. I performed DE analysis between the clinical groups in mdDCs and pDCs subtypes and I detected 30 differentially expressed genes (*adj. p-value* ≤ 0.05) between our clinical groups **(Figure 4.23 b and c).** In mdDCs, I observed the upregulation of interferon-alpha and gamma genes (ISG15, IFITM3 and LY6E). I observed high overall program scores in the BT and non-CR group **(Figure 4.23 b and Supplementary Figure 6.17).**

I performed receptor-ligand interactions analysis and observed that the interaction partner AXL-GAS6 in the BT group is expressed, forming cell-cell interaction between mdDCs and Plasma cells, monocytes, and pDCs **(Figure 4.24)** indicating that the mdDCs phenotype could potentially support tumor growth and has an immunosuppressive role in MM BM microenvironment (Yan et al. 2019)ʼ (Waizenegger et al. 2015).

Furthermore, I detected many other interaction partners that are expressed in the BT- group as IL1R-IL1B and IL1R-ILRN between mdDCs and MAN cells, supporting the potential immunosuppressive role of mdDCs and MAN cells in the BT- group.

a)

b)

cDC / Monocytes

c)

Plasmacytoid dendritic cells (pDCs)
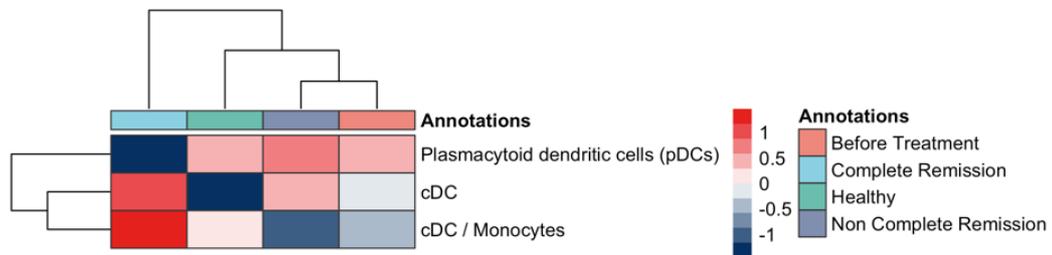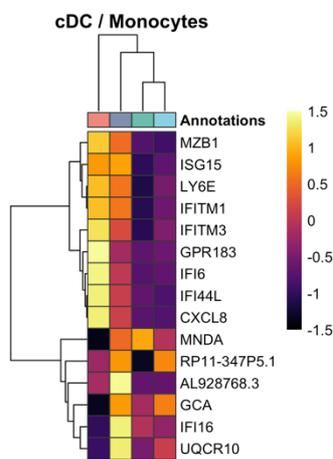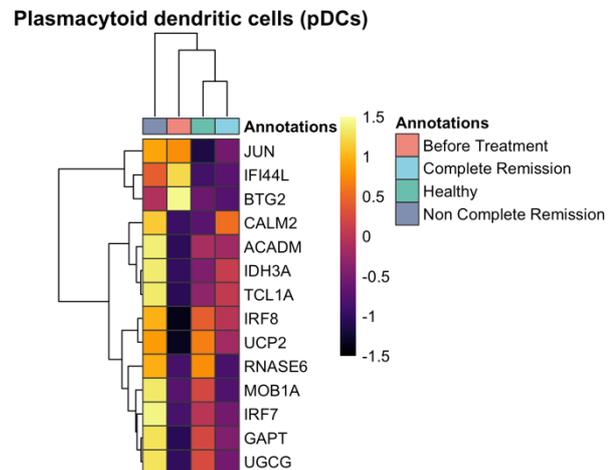
Figure 4.23: a) Heatmap shows the enrichment and depletion of DCs subtypes across the clinical groups. b) Heatmap shows the DE genes between the clinical groups in mdDCs. c) Heatmap shows the DE genes between the clinical groups in pDCs.

The color scale reflects scaled values of the gene's expression (z-score). The metadata is indicated by the color codes on the right-hand side

118
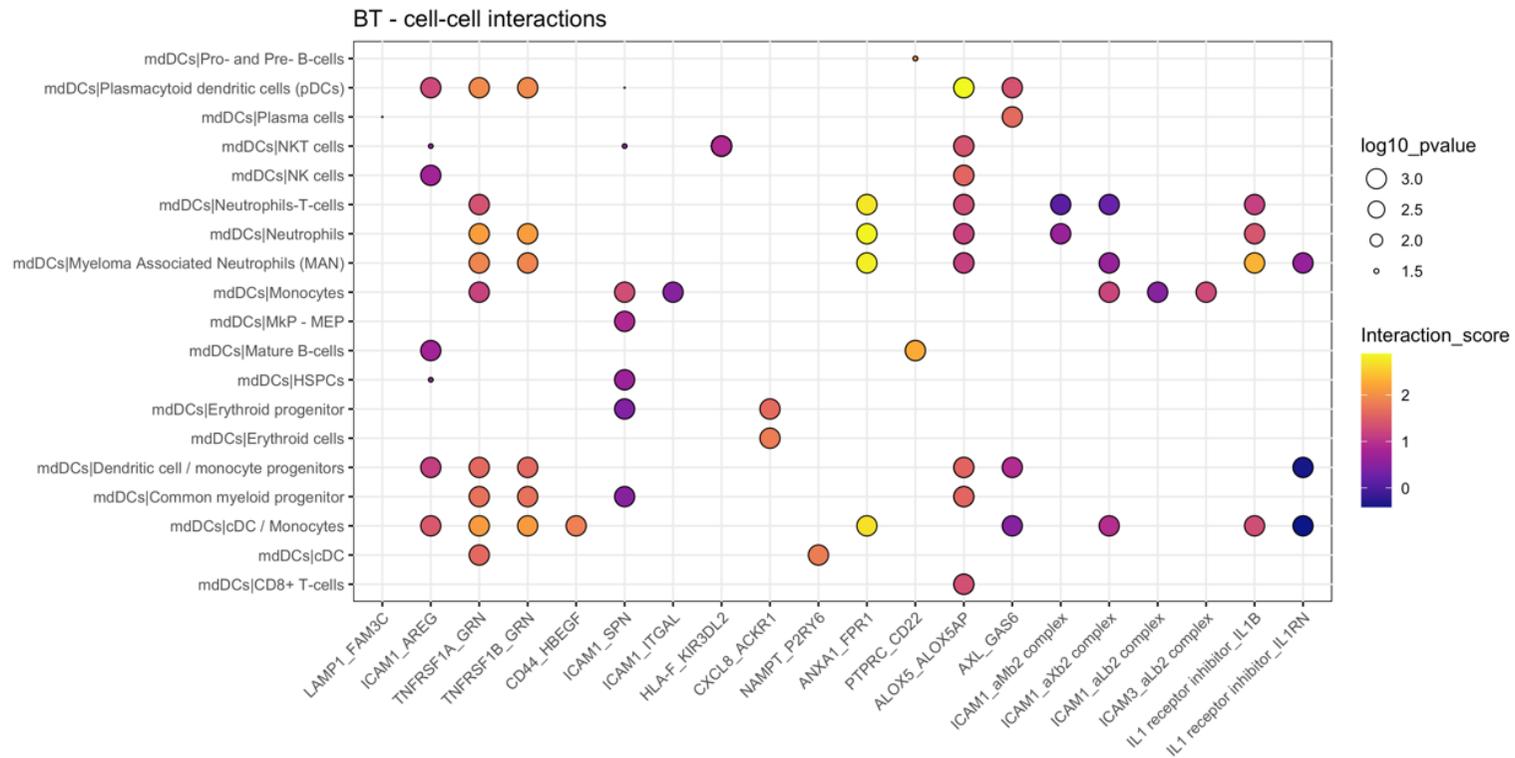
Figure 4.24: Balloon plot shows the Receptor-Ligand interaction patterns between mdDCs and other cell types in BT-group.

The color scale represents the mean receptor-ligand interaction scores. The size of the circle represents the (-Log10) of the p value

4.3.14   AMC CD8+ T cells enrichment in the BT-group

After T cell activation, new populations and subtypes of memory CD8+ T cells emerge, which have a quick capacity to proliferate and execute cytotoxicity functions and secret cytokines and live long after infection (Martin and Badovinac 2018).

To decipher the memory cellular states across the clinical groups; I performed in-depth analyses and defined four main memory CD8+ subtypes (n= 41718 cells). The aberrant memory-cytotoxic (AMC) CD8+ T cells were mainly enriched in MM-patients BT and partially enriched in the non-CR group **(Figure 4.25, Figure 4.26 and Supplementary Figure 6.18).**
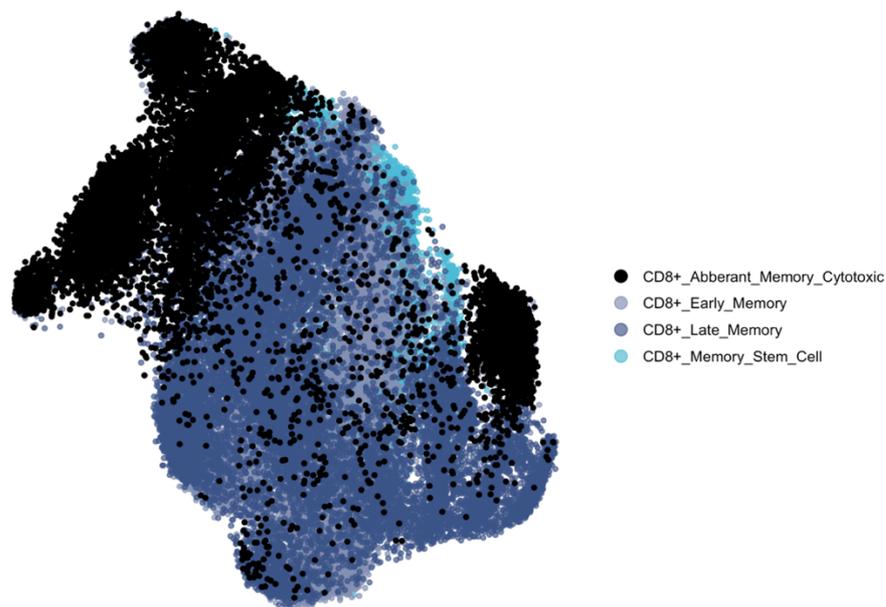
Figure 4.25: UMAP representation of the CD8+ memory T cells subtypes.
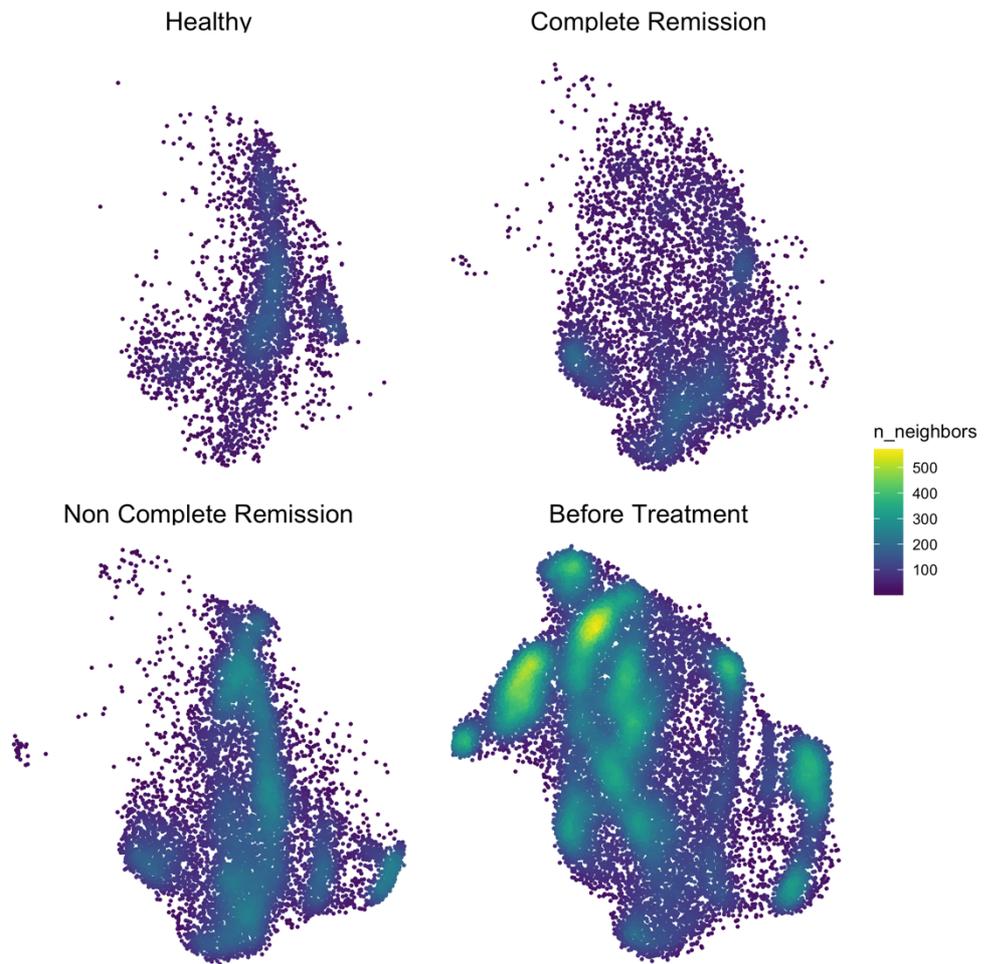
Figure 4.26: The density plot shows the memory CD8+ subtypes in all clinical groups and the enrichment of AMC CD8+ T cells in BT group.

* UMAP representation: the axes have been omitted for simplicity. The color scale represents number of neighbors

4.3.15 Velocity estimates and connectivity analyses predict multiple origins of AMC CD8⁺ T cells.

To investigate the differentiation trajectory of the AMC CD8⁺ T-cells, I performed RNA velocity analysis based on the ratios of spliced/unspliced mRNA to predict the potential sources of AMC CD8⁺ T cells for the memory CD8⁺ subtypes. I calculated the PAGA connectivity measure to quantify the cellular fate and transitions confidence of the memory CD8⁺ subtypes in BT group **(Figure 4.27 and Figure 4.28)**.

I found that the AMC CD8⁺ T cell population could originate from early memory and partially from late-memory and memory-stem CD8⁺ T- cells. I defined the top-likelihood genes representing the underlying dynamic behavior of the differentiation process toward the AMC CD8+ T cells in the BT group **(Supplementary Figure 6C)**.

Legend:
- CD8+ Abberant Memory Cytotoxic
- CD8+ Early Memory
- CD8+ Late Memory
- CD8+ Memory Stem Cell
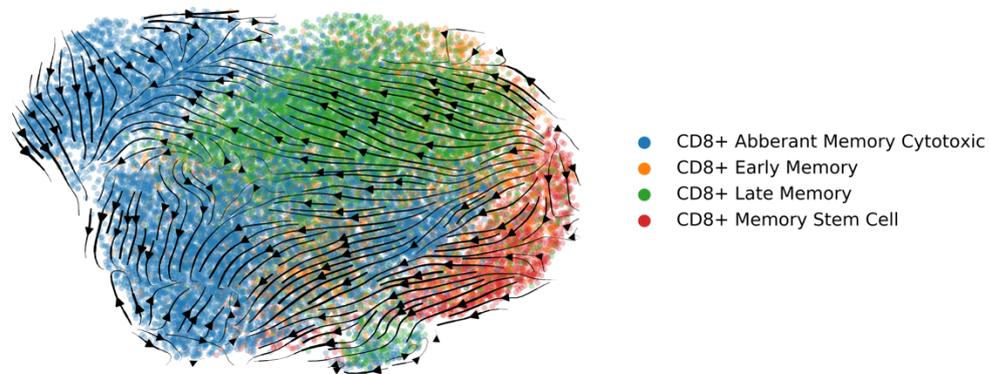
Figure 4.27: UMAP representation overlaid with RNA velocity vector field shows the differentiation trajectory directionality of the CD8⁺ memory subtypes in BT- group.

* The arrows represent vectors in the gene expression space and are estimated from the RNA velocity. The arrows show the directionality and speed of the underlying developmental process of the T cell memory subtypes.
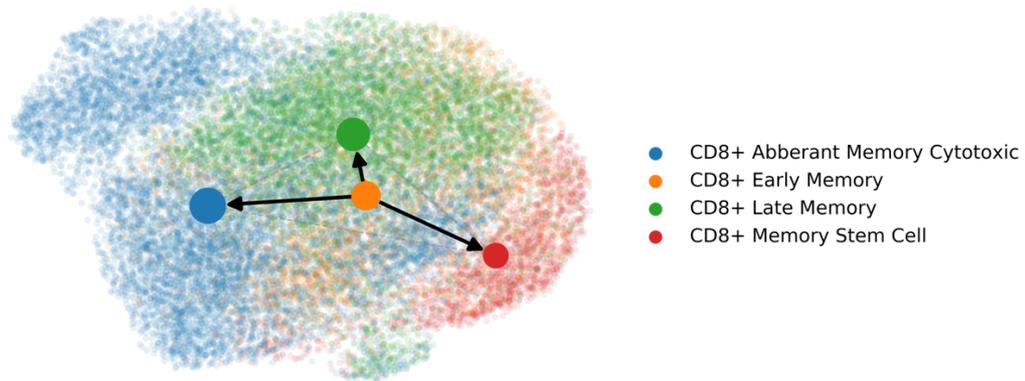
Figure 4.28: UMAP representation is overlaid with a directed graph which summarizes the transition confidence between the CD8⁺ memory subtypes in BT- group.

In-addition, I observed that the Phosphodiesterase enzyme subunit (PDE4B) which has an important role in T cell activation and suppression(Epstein 2017) is upregulated later in the development latent-time **(Figure 4.29).** To quantify the phenotypic expansion in the memory CD8⁺ T cell compartment across the clinical groups; I trained the RF model only on healthy and BT treatment cells and used the trained model to predict the memory CD8⁺T- cell phenotypic states in all clinical groups including CR and non-CR groups. I observed that the non-CR and CR groups retain memory CD8⁺ T cells which have a BT-like phenotype **(Supplementary Figure 6.19)** which is in line with similar observations in the analysis of both neutrophils and NK compartments.
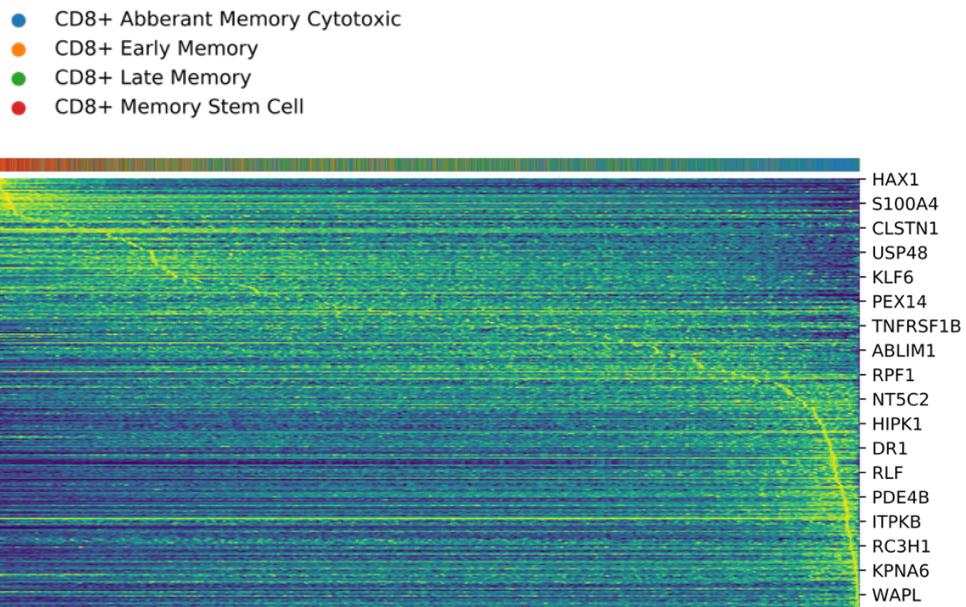
Figure 4.29: The heatmap shows the top-likelihood genes underlying the latent time of the CD8⁺ memory subtypes in BT- group.

* The yellow color's high Intensity represents high expression of the shown genes in rows

4.3.16 The AMC CD8⁺ T cell harbors an exhaustion phenotype

Furthermore, I performed DE analysis and found that AMC CD8⁺ T cells are expressing a unique group of surface markers and TFs, reflecting the exclusiveness of this aberrant memory state in the disease context. Upregulation of TFs like (ARID5A, JUND, BHLHE40, REL, ZEB2) (Roychoudhuri et al. 2016, Mognol et al. 2017, Scott and Omilusik 2019, C. Li et al. 2019, Visekruna, Volkov, and Steinhoff 2012), and other TFs like (ZNF331, CEBPZ, SKIL, and NR1H2) with an unclear role in T cell biology support the notion of the exclusiveness of these states in newly diagnosed multiple myeloma patients **(Supplementary Figure 6.21).** The upregulation of the surface markers ATP1B3 (CD298) and SLC7A5 (LAT1) indicates that these cells are metabolically active (Sinclair et al. 2013).

High expression of the memory surface marker CXCR3 (CD183) is observed. CXCR3 (CD183) has been proposed to have an important role in T cell trafficking and migration function (Groom and Luster 2011) (Hu et al. 2011) **(Figure 4.30).** However, a recent study showed that CXCR3 was not required for T cell migration and has an important role in enhancing the intratumoral CD8⁺ T cell response to PD-1 blockade (Chow et al. 2019). Another study showed that the Progenitors Exhausted T cells (T-PEX) express CXCR3 and CD44 as well as the known exhaustion markers like PDCD1, TIGIT, and TOX (Galletti et al. 2020b).

The upregulation of the transcription factor NR4A2, pre-dysfunctional markers (CXCR3, CXCR4, and CD44) and the downregulation of the activation markers ITGB1, CD52 and KLRD1 indicates a clear direction that AMC CD8⁺ states represent the "T cell exhaustion" phenotype in the multiple myeloma disease context (Mognol et al. 2017)(Seo et al. 2019).

CD8+ memory subtypes: DE Surface Markers and Exhaustion
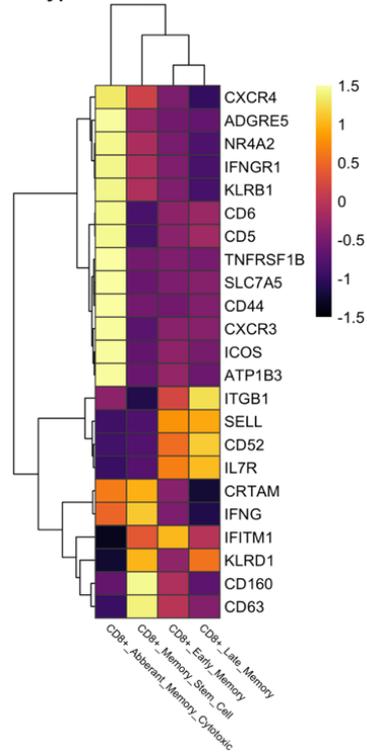
Figure 4.30: Heatmap shows the differentially expressed surface markers and exhaustion genes in the CD8⁺ T cell memory subtypes.

* The color scale reflects scaled values of the gene's expression (z-score). The metadata is indicated by the color codes on the right-hand side**.**

126

4.3.17   Experimental validation of AMC CD8$^+$ T cell population

To further characterize the AMC CD8$^+$ T-cells, I proposed many marker genes **(Figure 4.30)** to be tested via wet lab experiments. By using qPCR and FACs experiments (performed by Dr. med. Raphael Lutz), we found that CXCR3$^+$ expression shows specific FACs expression in a new patient cohort of newly diagnosed MM patients (n=30).

Further qPCR analyses have been performed for the sorted CXCR$^+$ CD8$^+$ T cell population and showed a high expression of the predicted and proposed markers from the single-cell analyses. Furthermore, the FACs data showed that there is a partial correlation between the CXCR3$^+$ T cell and plasma cell infiltration which has been confirmed by imaging staining on bone marrow samples.

4.3.18 Systems level understanding and constructing the global network of cell-cell interactions

To gain a global understanding of the cell-cell crosstalk interaction patterns in the bone marrow microenvironment, I constructed a global network between of potential receptor-ligand interactions based on the co-expression pattern. At the global level, I observed a substantial decrease in the receptor-ligand total interaction counts (R-L TIC) between the cell types in BT group and LTS group state, indicating an altered bone marrow microenvironment in comparison to the healthy bone marrow microenvironment state **(Figure 4.31).**

Additionally, I observed increased receptor-ligand core interaction patterns (R-L CO) between the MAN cells and other cell types in the BT group and non-CR groups **(Supplementary Figure 6.22)**. More specifically, I observed that in both BT group and non-CR groups, the stimulatory interaction partners LGALS9_CD44, LGALS9_CD47, CD74_MIF (Y. Zheng et al. 2016) and HLA-DPB1_TNFSF13B are upregulated between MAN cells and many other cell types. These co-expression pattern of receptor-ligand suggests that MAN cells mediate a pro-inflammatory crosstalk interaction with the other cell types, therefore inducing an immunosuppressive and inflammatory microenvironment.
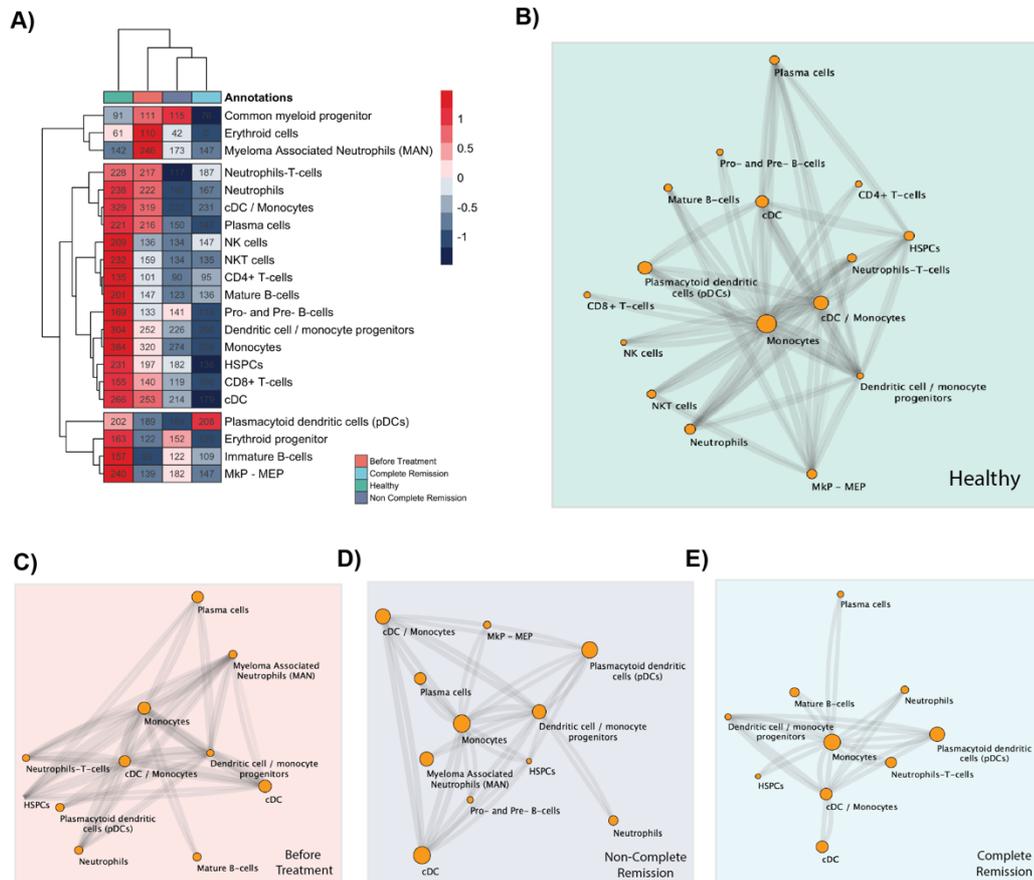
Figure 4.31: Bone marrow microenvironment global network construction between the cell types across all clinical groups.

a) The heatmap shows the receptor-ligand total interaction counts (R-L TIC) per cell type for each clinical group. Network construction for the receptor-ligand core interaction (R-L CO) between bone marrow cell types **(b)** healthy **(c)** before treatment **(d)** non-complete remission group **(e)** complete remission group. Node size and edge width reflect the interaction count per cell type (interaction count threshold > 15).

4.3.19 Global Hallmark pathway scores across all bone marrow cell types and states.

To determine which pathways have a global influence on the BM microenvironment remodeling and cellular state phenotypic shift; I performed global scoring for all GSEA Hallmark pathways for each single cell of all BM cell types (after excluding the plasma cell compartment). Then I conducted a correlation analysis between the pathway scores for each clinical group; I observed altered correlation structures between the hallmark pathways per clinical group in comparison to the healthy BM state **(Supplementary Figure 4.19).**

I constructed the GLM for the pathway scores for each clinical group and I observed statistically significant GLM estimates (P-value < 0.001) for many of these pathways to predict the clinical state **(Figure 4.32 and Supplementary Figure 6.24)**. I observed that there are many pathways related to the induction of proinflammatory and immunosuppressive signals in the BM microenvironment (TNF alpha signaling via NFKB and TGF beta signaling), in addition to apoptosis and hypoxia biological signals.

These pathways showed stepwise increasing GLM estimates starting from low estimates in the CR group and moving to higher estimate values in the non-CR group to the highest predictive estimates in the BT group **(Figure 4.32)**. Furthermore, I observe that there are other pathways specific to certain clinical states (e.g., interferon-alpha and gamma pathways in the non-CR group, and p53 and complement pathways in the CR group). These global analyses suggest that there is a potentially evolving landscape of immune states that govern the MM disease before and after long-term survival.
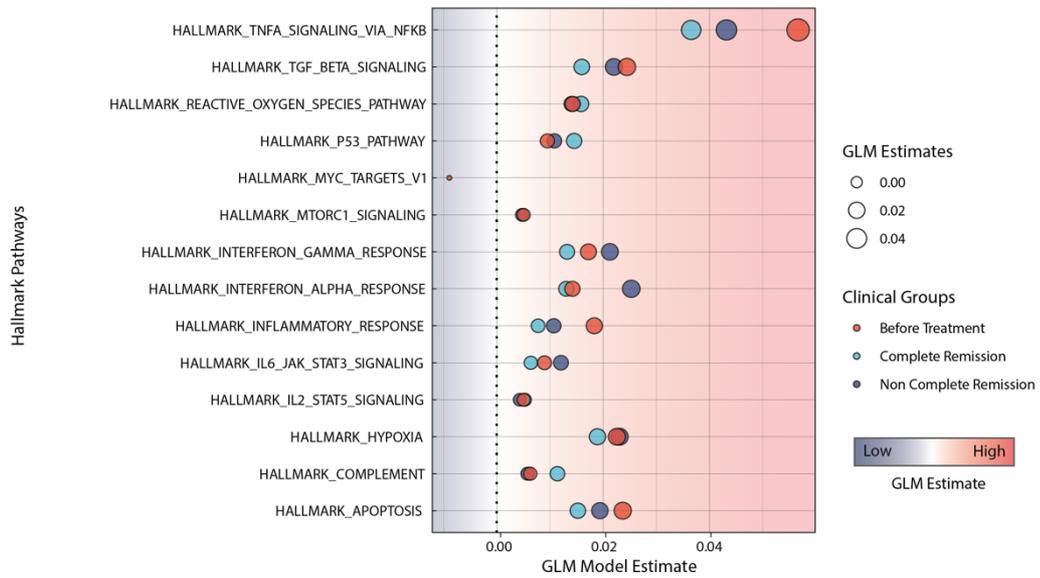
Figure 4.32: GLM estimates of the hallmark pathways of all immune bone marrow cell types across the clinical groups.

The circle colors represent the clinical groups and the size of the circle reflects the GLM estimate values (p-value < 0.001).

5 Discussion

Even though we started to gain more insights and understanding about myeloma biology and treatment, we are still missing the key solutions for the MM puzzle which is reflected in the relapse of patients after receiving current intensive therapy.

While I am conducting this study, I found few resources on the microenvironment and the immune landscape of the disease (Ghobrial et al. 2018), and rather more studies focused on the tumor compartment (myeloma cells) (Pawlyn and Morgan 2017). We are missing in-depth studies (from computational and wet lab sides) into the tumor microenvironment in the context of MM disease (Thorsson et al. 2018). More efforts are needed to gain deep insights into the tumor microenvironment of the disease, which will be a key factor to resolve the disease biology and find an ultimate cure.

Recently, two single-cell RNA-seq studies have been published on the precursor stages of MM. Ledergor et al. 2018 characterized malignant myeloma cells in asymptomatic individuals (precursor stages) and identified tumor cells that have a similar transcriptomic profile as the active disease state. Nevertheless, Ledergor et al. 2018 study focused on the tumor cells and did not report on the immune-microenvironment. Zavidij et al. 2020 study performed single-cell RNA sequencing for the precursor stages of MM with more focus on the tumor immune-microenvironment, where they characterized dysregulated cellular states and found compromised immune microenvironments.

5.1 MM long-term survivors go through a complex and evolving immune landscape

The research work on this thesis represents the first comprehensive investigation of the bone marrow immune microenvironment (paired-sample setting) at a single-cell resolution for the same MM patients before and in long-term survival (~ 7-15 years). Our study represents the first effort on linking the cellular states and immune phenotypes of MM patients to the long-term survival (LTS) using single-cell genomics approach and computational models.

I developed computational and machine learning approaches to analyze and dissect the cellular states and subtypes of the different clinical groups. I revealed the global landscape of the bone marrow microenvironment before and in long-term survival for the first time. I found that the MM patients go through a complex landscape of the immune states which control the disease. Over time, our patients lose this immune control that leads to the emergence of a malignancy state and, ultimately, they relapse. I proposed a new model, the Continuum Immune Landscape (CIL) Model, which explains the long-term survival phenomenon from the perspective of the cell types and states in the bone marrow microenvironment. The CIL model and the key findings of this study will be discussed next in this chapter.

5.1.1 The Continuum Immune Landscape (CIL) model: a new model explains long-term survival in MM

The CIL model can be summarized in <u>six</u> assumptions which are based on the extensive analyses and observations (showed in the Results section) of the cellular states and cell types before and after long-term survival **(Figure 5.1)**:

1- The long-term survival states (CR and non-CR) represent one of the disease stages and do not reflect a cure nor healthy-like state.

2- The immune states of multiple myeloma patients can be represented as a continuum of a complex immune landscape of cell types and states.

3- One of the main characteristics of the CIL model is that the patients follow a multi-stepwise evolution from a more cytotoxic state (high immune control) to a less cytotoxicity (medium immune control), and finally to an inflammatory cellular state (low immune control).

4- In the complete remission (CR) state, immune cells with high cytotoxicity potential are abundant, and mediate a strong immune control of the disease (high immune control).

5- The non-complete remission (non-CR) state harbor lower cytotoxicity functions in comparison to the CR state, and strong pro-inflammatory signal (medium immune control).

6- The MM initial diagnosis state (supposedly similar to the relapse state) is characterized by high inflammatory signals and exhausted cellular states (low immune control).

Figure 5.1: The Continuum Immune Landscape (CIL) Model explains multiple myeloma immune states before and in long-term survival.

X-axis: **High** (MGUS and CR), **Medium** (SMM and non-CR) and **Low** (initial diagnosis and relapse) **Immune Control.**

Y-axis: represent the immune control potential **and** MM patients are in the circles**.**

5.1.2 The global landscape representation and compositional shifts in the BM immune microenvironment before and after Long Term Survival (LTS)

At the global level, I defined and annotated 23 major cell types across all clinical groups representing the major known immune cell types in the BM microenvironment. I observed a high degree of immune cell types composition variation and shifts per patient's samples as well as across clinical groups in comparison to the healthy donors' samples.

I developed a Generalized Linear Mixed Model (GLMM) to further associate cell types abundance with the clinical states of the patients before and in long-term survival (LTS). In line with previous studies (Ledergor et al. 2018, Zavidij et al. 2020), the BT group showed an enrichment of malignant plasma cells and NK cells. I defined a new population termed myeloma associated neutrophils (MAN). The GLMM model showed that MAN cells have a predictive power for the clinical state; especially in BT and non-CR groups. In addition, the model showed interesting predictive associations between the abundance of NKT, T cell, Monocytes and DCs subtypes abundance and the LTS groups.

I developed a random forest (RF) model to quantify the cellular states per cell type across all clinical groups and to predict the cellular state in a probabilistic manner. The RF model showed that all LTS states harbor a phenotypic state similar to that of the before treatment-like (BT-like) states and lower proportions of healthy-like (H-like) cellular states. This means that all patients even after a complete remission, still in one of the disease states. They are not cured nor healthy individuals, and they harbor different cellular states in comparison to the healthy control.

5.1.3 Residual tumor cells (RTCs) have been detected in the non-CR and CR groups.

At the cell type level, I started to decipher each cell type one at a time. In the plasma cell compartment, I started to define the healthy and malignant plasma cells using graph-based clustering algorithm, which showed a heterogenous subpopulation of specific malignant plasma cells per patient. The malignant transcriptional clones have been confirmed using previously defined malignancy markers such as SCD1 and TNFRSF17. I detected malignant transcriptional clones in the non-CR state, too, representing residual tumor cells (RTCs).

I performed CNV analyses to define underlying chromosomal aberrations. The CNV analyses showed common aberrations like chr13 loss and chr22 duplication in all patients. I observed that each patient has a unique make-up of chromosomal aberrations, which persisted over time in the non-CR and CR groups. These analyses suggest an evolutionary trajectory of malignant transcriptional clones which co-adapted and co-evolved overtime with long-term immune cellular states.

5.1.4 The complete remission (CR) group represents the high immune control (HIC) state

The CIL model proposes that all MM patients are under a complete remission state after receiving the treatment protocol. Based on the in-depth analyses of the cellular states, more specifically the NK and T cells, I observed a striking pattern of upregulation of the effector cytotoxic molecules and programs in complete remission (CR) group. In the NK compartment, I observed NK phenotypic expansion in CR and non-CR groups, showing the highest diversity in the BT group.

To further quantify this phenotypic expansion, I developed a random forest (RF) model, which suggests that the LTS groups retain NK cells that share BT-like phenotype. Further analyses revealed that the CR group showed upregulation of cytolytic effector markers (GNLY, GZMB, NKG7, PRF1, and KLRD1); indicating that the CR group harbored higher NK cytotoxicity function **(Figure 5.2)**.
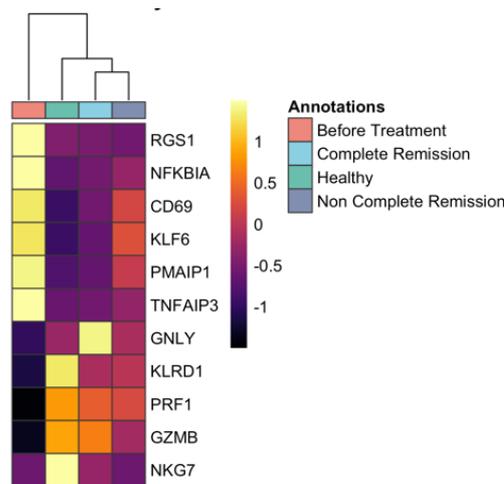
Figure 5.2: The heatmap shows the expression of NFKB and inflammatory pathway and NK cytotoxicity genes across the clinical groups

Moreover, I performed receptor-ligand interaction analysis and observed that the stimulatory interaction partner CD94:NKG2C heterodimer - HLA-E, which has a high interaction score between NK cells and many other cell types in the CR group, reflecting a high NK activation state in CR group. In the T- cell compartment of the CR group, I observed an enrichment of the effector cytotoxic CD8[+] T- cell and high overall global scores for the T cell activation signatures as well as upregulation of the superior cytotoxic marker ITGB1 (CD29) (Nicolet et al. 2020).

One critical observation is that I did not observe enrichment of malignant plasma cells' transcriptional clones in the CR group. I propose that the malignant plasma cells are under a strong immunosurveillance state controlling the disease state and progression. Therefore, we could not detect their presence. Building on these evidences, the CIL model suggests that the CR group represents the high immune control (**HIC**) state in the long-term survival immune landscape and fate-trajectory **(Figure 5.1)**.

5.1.5 The non-Complete Remission (non-CR) group represents medium immune control (MIC) state

In-depth analyses of the NK and T cell compartments of the non-CR group showed lower cytotoxicity functions in comparison to the CR group. In-contrast, I observed upregulation of TNFA and NFKB1 inflammatory pathway genes (TNFAIP3, CD69, RGS1, KLF6, and NFKBIA) indicating a higher proinflammatory signal in non-CR BM microenvironment.

The receptor-ligand interaction analysis reveal that non-CR group loses the stimulatory interaction partner CD94:NKG2C heterodimer - HLA-E and keeps upregulating the inhibitory interaction partner CD94:NKG2A heterodimer - HLA-E suggesting an inhibitory phenotype in the non-CR BM microenvironment.

In the T cell compartment of the non-CR group, I observed an enrichment of both early and late memory CD8$^+$ T cells as well as T$_{regs}$ and effector CD4$^+$ T cells with high overall scores of the T cell dysfunction and exhaustion signature. Furthermore, I observed partial enrichment of a new population termed myeloma associated neutrophils (MAN), which induce pro-inflammatory and immunosuppressive signals. Furthermore, in the non-CR group's mdDCs, I observed an upregulation of interferon-alpha and gamma genes and signatures.

139

The enrichment of malignant plasma cells in the non-CR group without clear clinical signs suggests that the non-CR group is under relatively lower immune control than the CR group, which represents an early phase of potential relapse. Based on these observations, the CIL model suggests that the non-CR group represents the medium immune control (**MIC**) state in the long-term survival immune landscape **(Figure 5.1)**.

5.1.6 Before treatment (BT) group represents the low immune control (LIC) state

While LTS in both groups showed high cytotoxicity signals, the before treatment (BT), on the other hand, showed downregulation of the cytotoxic functions and a significant increase in inflammatory signals.

In the NK compartment, the BT group showed significant upregulation of TNFA and NFKB1 inflammatory pathway genes (TNFAIP3, CD69, RGS1, KLF6, and NFKBIA) and downregulation of the main cytolytic effector marker molecules (GNLY, GZMB, NKG7, PRF1, and KLRD1). Furthermore, the receptor-ligand interaction showed an upregulation of (TNFRSF1B - GRN) and (IFNG Type II - IFNR) interaction partners in the NK compartment of the BT group, which potentially mediate an immunosuppression state; allowing myeloma cell proliferation.

Strikingly, I observed a high abundance and enrichment of MAN cells which showed upregulation of IL1B, a potent proinflammatory cytokine and upstream of IL-6, which supports myeloma cell growth in addition to other pro-inflammatory signatures (Rosean et al. 2014). Adding to the upregulation of interferon-alpha and gamma genes in mdDCs; the immunosuppressive interaction partner (AXL-GAS6) is expressed in the BT group, forming cell-cell interaction between mdDCs and malignant plasma cells, monocytes, and pDCs.

Regarding the T cell compartment, the BT group showed significant upregulation of T cell dysfunction and exhaustion signatures as well as TNFA signaling via the NFKB pathway, hypoxia pathway, and apoptotic processes.

I described a novel population termed AMC CD8$^+$ T cells (CXCR3$^+$) population which showed high abundance in the BT group. RNA velocity estimates showed that the AMC CD8$^+$ T cells would potentially originate from several memory CD8$^+$ T cell subtypes. The AMC CD8$^+$ T cells showed upregulation of T cell dysfunctional and exhaustion markers (CXCR3, CXCR4, CD44, and NR4A2) along with downregulation of activation markers ITGB1, CD52 and KLRD1. This population represents the hard to define T cell exhaustion state in MM disease context.

All these pieces of evidence support the CIL model and correctly position the initial diagnosis state (BT-group) at the end of the immune-control trajectory. I hypothesize that the BT state is a proxy for the relapse state and I therefore propose that the BT group represents the low immune control (**LIC**) state in the immune landscape model leading to an immune escape state of myeloma cells and a full-blown disease **(Figure 5.1)**.

5.1.7 Disease-associated trajectories and dysregulated cellular states and phenotypes

The CIL model poses the hypothesis that the disease can be represented as a deviated trajectory from the healthy cellular state trajectory. Such deviations showed a unique molecular profile, gene regulatory program, and networks of cell-cell interaction patterns.

The disease-state trajectories (DST) hypothesis **(Figure 5.3)** is based on <u>three</u> main observations from the analyses:

1- The rise of new populations and cellular states (e.g., MAN cells, AMC CD8$^+$ T cells and NK phenotypic expansion).

2- altered receptor-ligand interactions and cell-cell interaction network patterns in the diseased states.

3- altered global developmental trajectories (e.g., CD8$^+$ T cell differentiation trajectories across the clinical groups)
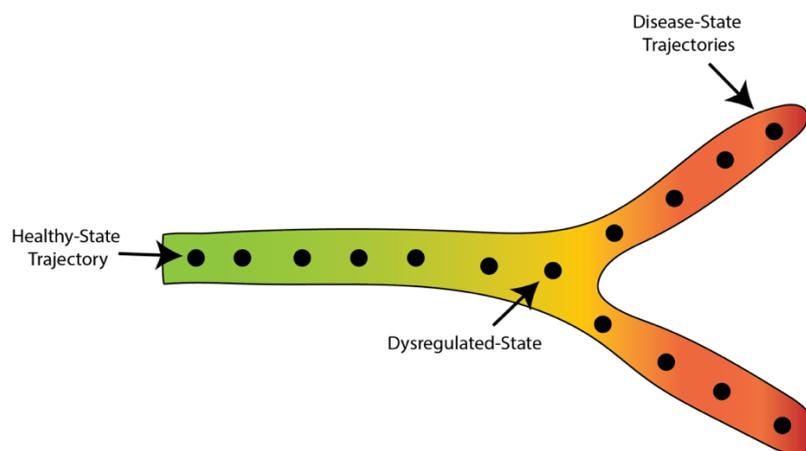
Figure 5.3: The disease-state trajectories (DST) model and dysregulated cellular states.

The greenish colors represent the cellular state in healthy individuals. A dysregulated cellular state (yellowish color) may start to perturb the healthy trajectory that would lead to the emergence of the disease-state trajectories (DST) (red colors).

One striking observation that supports the Disease-State Trajectories (DST) hypothesis is the observation of completely new cellular states in the neutrophil compartment (MAN cells), which showed a unique transcriptional state in the BT and non-CR groups. I found an altered cellular state associated with the disease progression in the CD8$^+$ T compartment (AMC CD8$^+$ T cells) which are expressing a unique group of surface markers and TFs, reflecting the exclusiveness of this aberrant state in the MM disease context. The cell of origin of these new cellular states could be from the same cell type or the emergence of the cellular state from hydride origins of several cell types. A recent study (Ahmed et al. 2019) found an unknown lymphocyte that showed a dual expression of TCR and BCR and key lineage markers of both B and T cells in the context of type 1 diabetes (T1D).

At the cell-cell communication level, I found an altered pattern of receptor-ligand interactions associated with the disease state. For example, I found that MAN cells started to initiate new cell-cell interaction patterns with other cell types in the non-CR and BT groups; mediating a pro-inflammatory crosstalk interaction with the other cell types. Such observations alongside others in many cell types (NK, NKT, mdDCs and others); support that disease trajectories induce new type of cell-cell interactions to ultimately induce cellular state shifts and phenotypic changes, as a mean for mediating a new disease trajectory.

At the global developmental trajectory level, I observed that the CD8$^+$ T cells started to adapt a variety of differentiation strategies associated with the disease state. In-depth analyses showed that in the healthy context, CD8$^+$ T cell differentiation processes follows linear cell fate model. However, in BT and non-CR groups, the differentiation processes was better described by a continuum cell fate model. In the CR group, the differentiation processes adapted partial-linear cell fate model, which is more similar to the healthy trajectory order. Such observations suggest that through the disease trajectories, cell types start to adapt to different developmental strategies according to the disease state.

This observation is confirmed by the different transcriptional state of the Naïve CD8$^+$ T cells. The module analysis showed that the naïve CD8$^+$ T cell states are shaped and influenced by the immune state of the disease, and this could further alter the global T cell differentiation process of the naïve CD8$^+$ T cells into memory and cytotoxic CD8$^+$ phenotypes and give rise to more dysregulated CD8$^+$ states and phenotypes.

5.2 New therapeutic targets and predictive prognostic markers

The analyses open new possibilities for clinical applications including new therapeutics options and predictive prognostic markers for the disease state. Future efforts are needed to test and validate these markers via in-vitro, in-vivo and clinical approaches and assays.

5.2.1 CXCR3 and NR4A2: new therapeutic targets to reverse the T cell exhaustion state in newly diagnosed MM patients

I found that the AMC CD8$^+$ T cells represent the exhaustion state in MM context. I found that CXCR3 surface marker and NR4A2 transcriptional factor (TF) expression defines and regulates this T cell exhaustion state in multiple myeloma context. Therefore, I propose that by targeting CXCR3 and NR4A2, we could reverse the *exhaustion* state of the T cells to a more *effector cytotoxic* state **(Figure 5.4)**.
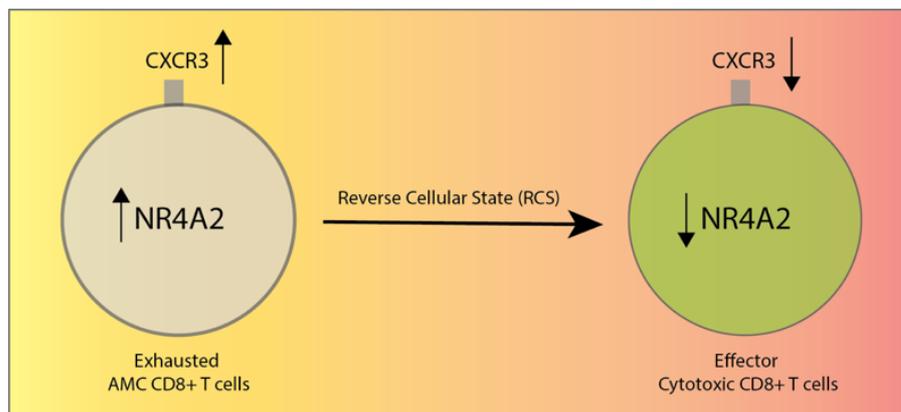
Figure 5.4: Reversing T cell exhaustion state by targeting CXCR3$^+$ CD8$^+$ T cells and its NR4A2 transcriptional factor in MM context.

This notion is supported by recent studies on the role of NR4A transcriptional factor in T cell exhaustion biology, which has been shown in a mouse model (X. Liu et al. 2019) and human CAR T cells (Chen et al. 2019). Furthermore, CXCR3 has been found recently to be upregulated in the newly defined Progenitors Exhausted T cell (T_{PEX}) state (Galletti et al. 2020a). $CXCR3^+$ has been defined to have an important role in enhancing the intra-tumoral $CD8^+$ T cell response to PD-1 blockade (Chow et al. 2019).

5.2.2 Global disease-state $CD8^+$ T cell prognostic markers

One important application of the analyses is the definition of certain surface markers, TFs, and pathways constitutively expressed in the $CD8^+$ T cell compartment subtypes, which are specifically expressed in each of the clinical groups **(Figure 4.17)**. These markers could serve as disease-state specific markers, and the prognostic value of these markers needs to be tested in larger clinical cohorts.

5.3 Limitations and future Directions

Through the following section, I will try to summarize the limitations of this study and the field as a whole. In addition, I will propose new directions which could resolve these challenges

5.3.1 mRNA represents just one layer of the biological regulation processes

Through this study, I tried to define cell types and states based on one layer of information (mRNA). The biological systems are rather more complex; multiple sources of information orchestrate and regulate these systems. The genetic and epigenetic states of the cell regulate the cell type identity and phenotypic states (Kundaje et al. 2015).

Recently, new approaches have been developed to capture more layers of information. For example, single-cell ATAC-seq is available now to define the chromatin accessibility and capture the open and closed chromatin regions across the whole genome at a single-cell resolution (Satpathy et al. 2019). More technologies have been developed to capture the methylation state combined with gene expression at single-cell resolution (Linker et al. 2019). Other technologies have been developed to detect surface protein expression e.g, CITE-seq (Stoeckius et al. 2017) and spatial transcriptomics techniques to map the co-localization of cell types in different tissues e.g., Slide-Seq (Rodriques et al. 2019). Such technologies have been developed to capture more layers of information (Butler et al. 2018), which could be used to gain in-depth characterization of the immune landscape of the long-term survival of MM patients.

These technologies and new data types would need to be integrated into one joint representation of cell types and states, which could open new possibilities for understanding the true nature of the MM disease and its immune landscape during the disease evolution. These new data types will pose challenges for the computational community. We would need to develop new methods to handle increased data size, integrate data types, deal with technical variations, inherited sampling noise, and the asynchrony of the multiple layers of biological information processing. One promising approach is the use of Bayesian deep learning models e.g. the variational autoencoder (VAE) model. VAE has been shown recently that it can process vast amounts of data and integrate different data types in an efficient manner (Lopez et al. 2018).

5.3.2 Cell of origin: the rise of new populations and cellular states.

It has been challenging to define the cell of origin of the new populations and cellular states. For the AMC CD8$^+$ T cells, I tried to infer the cell of origin using trajectory inference and RNA velocity methods to define potential cell of origins. However, it would be important to perform TCR sequencing for AMC CD8$^+$ to trace the differentiation process using their inner cellular marks. Regarding the MAN cell origins, it would be important to use lineage tracing techniques (Weinreb et al. 2020) to define its cell of origin, whether it gives arise from the same cell type or from hydride origins and states.

5.3.3 Systems Immunology: building a holistic view of the immune system

Through this study, I observed lots of contradictory information in the literature about the role of markers or cellular state in different disease contexts. I think that the source of this confusion is due to <u>two critical points</u>:

1- **Cell type definition:** different groups and labs use different marker genes and surface markers, as a definition for the same cell types and subtypes (Günther and Schultze 2019).

2- **Context-dependent findings:** due to the focused research approach on a certain cell type in a certain disease or biological context, we lost the global view of the immune system and introduced contradictory statements about the same marker and cell type (Blank et al. 2019).

The current state is challenging, however, it is encouraging us to find new ways to understand the immune system and to come up with a holistic view of the immune system by defining the underlying mechanistic regulatory networks (Binnewies et al. 2018).

Systems immunology approach with the aid of the current bioinformatics methods, machine learning models, and integrated single-cell atlases could help us reach the next wave of understanding the immune system despite the context specificities (Davis, Tato, and Furman 2017). Moreover, such systems biology models and approaches could be helpful to reach universal definitions of cell types and states across diseases and biological contexts.

5.3.4 New therapeutic paradigm: cellular-state targets and reversers

The current therapeutic approaches are trying to find new targets that are based on the "one-gene-target" hypothesis. This process starts with screening for many genes that could be associated with the disease. This knowledge then is translated to in vivo experimental systems; trying to perturb such genes (one at a time) and observe the downstream effects of such perturbations. Then, as soon as an approximate understanding of this gene in a certain context is achieved, finding the therapeutic option starts in place; this gene is then addressed as a potential therapeutic target to hopefully inhibit its downstream effects.

The "one-gene-target" approach assumes that a single gene could be a key regulator for the tumorigenic processes. The current large-scale genomics studies showed the existence of 43,778,859 single-nucleotide variants, 2,418,247 indels, and 288,416 structural variant events which represent somatic variants in 2,583 cancer patients (Campbell et al. 2020). In addition, early cancer genomics studies showed that hundreds to thousands of mutations can be detected per tumor type (Lawrence et al. 2013). This realization reflects the complexity of alterations in real cancer patients' samples. We need to rethink the "one-gene-target" hypothesis to ultimately think of cancer cure rather than enhancing the survival for several months in the majority of our cancer patients.

Through the analyses, I observed that the clinical states of multiple myeloma patients are associated with the emergence of new cellular states (e.g., MAN cells, AMC CD8$^+$ T cells, NK cell phenotypic expansion and other). Such observations suggest one key realization; which is that the disease induces complete cellular state changes and shifts.

Possibly, we could think about new therapeutic approaches that could target such global cellular state changes, and reversing therefore the cellular states **(Figure 5.5)** to more healthy-like states via Cellular-States Reversers (**CSRs**).
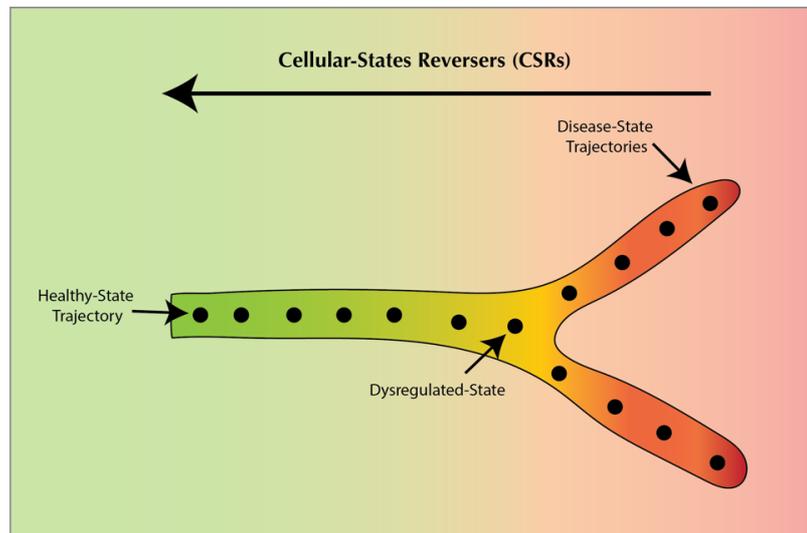
Figure 5.5: New therapeutic paradigm to target and reverse cellular states

To reverse an entire cellular state means that we should think about a new way of altering the cell-cell interaction patterns, which are mediated by the chemokine-cytokine signaling networks, and reversing the transcriptional state of the altered cellular state.

I propose that by designing newly programed cell types, which can function as CSRs via sensing the altered microenvironment states in the disease context, and secreting biological factors to reverse back the entire cellular state to a healthy-like state accordingly. This could seem hard to reach, however, there is a current synthetic biology approach that has been developed to design new biological circuits with programmed biological functions (Purnick and Weiss 2009). Furthermore, the current CAR T cells therapeutics approach has been developed based on cell-based therapeutics and synthetic biology approaches (Caliendo, Dukhinova, and Siciliano 2019).

More research efforts are needed from the computational and experimental sides, to reach fully programmed cell types which can induce a reversing cellular states (RCS) process and ultimately reach healthy-like cellular states or freeze current preferential cellular states as the high immune control (**HIC**) state of our patients.

5.3.5 Causal Inference and Reinforcement learning (RL)

Current machine learning methods are based on the independent and identically distributed (IID) data assumption and on finding a correlation between the observed variables. Such an approach cannot help us in defining the causal links between the observed variables. Incorporating causal knowledge to our models would be an important aspect that needs to be further investigated; to gain mechanistic insights about the underlying biological causal structures between cell types and states in the MM and various disease contexts.

One attempt to move statistical learning to causal representation learning is the current use of reinforcement learning (RL) methods. RL methods showed surprising success in video game scenarios (Silver et al. 2017), however, they have some trouble dealing with real-world high-dimensional data (Schölkopf 2019), and with shaping the right reward function (sparse rewards) to gain desirable outcomes (Hare 2019).

Despite the current challenges in the RL field, the idea of having multiagent learning (MAL) from the environment and implementing policies depending on a reward function would be a good start. We could develop such models to resolve cell-cell interaction scenarios and to gain mechanistic insights through inducing global perturbations to the tumor microenvironments from different tumor entities.

6 Appendices

6.1 Supplementary tables

Number	Gender (M/F)	age	MM Type	CRAB criteria	Time after ASCT (years)	Stage (ISS)	cytogenetics	Cytology: % PC in BM	Induction treatment	Pre-ASCT response	conditioning	Maintenance, duration (in years after ASCT)	Post-ASCT response
1	M	68	IgG kappa	bone disease	14	I	standard	5%	3x VAD	NA	2x Mel 200	interferon, 8	CR
2	M	73	IgG lambda	bone disease, anemia	11	I	standard	30%	3x VAD	PR	2x Mel 200	thalidomide, 2	CR
3	F	71	IgA/IgG lambda	bone disease, anemia	10	III	standard	90%	3x VAD	VGPR	2x Mel 200	thalidomide, 1	CR
4	F	69	IgG kappa	bone disease	9.5	I	standard	15%	3x VAD	PR	1x Mel 200	thalidomide, 2	CR
5	F	73	IgK kappa	bone disease	9	II	high risk (del17p)	50%	3x PAD	VGPR	2x Mel 200	bortezomib, 2	CR
6	M	77	IgG kappa	bone disease	10	I	standard	10%	3x VAD	PR	1x Mel 200	thalidomide, 1	CR
7	M	73	IgG kappa	anemia	9	I	standard	20%	3x PAD	VGPR	2x Mel 200	none	CR
8	F	56	IgA lambda	bone disease	9	I	high risk (del 17p)	60%	3x PAD	VGPR	2x Mel 200	bortezomib, 2	CR
9	F	67	IgA kappa	anemia	9	I	standard	80%	3x TAD	VGPR	1x Mel 200	None	VGPR
10	M	54	BJ kappa	bone disease	15	NA	NA	NA	3x VAD	NA	2x Mel 200	interferon, 2	CR
11	F	58	IgG kappa	bone disease	14	II	NA	60%	3x TAD	NA	2x Mel 200	thalidomide, 4	CR
12	M	69	IgG kappa	bone disease	14	III	standard	80%	3x TAD	PR	2x Mel 200	thalidomide, 4	CR
13	M	70	IgA lambda	bone disease	9	II	high risk (del 17p)	100%	3x PAD	nCR	2x Mel 200	bortezomib, 2	CR
14	F	65	BJ kappa	bone disease	11	I	standard	20%	3x PAD	VGPR	2x Mel 200	bortezomib, 2	CR
15	F	79	IgA lambda	bone disease	14	I	standard	30%	3x VAD	CR	2x Mel 200	interferon, NA	CR
16	F	58	IgG kappa	bone disease	17	NA	NA	80%	4x VID	NA	2x Mel 200	interferon, 13	CR
17	M	59	IgG kappa	renal failure, anemia	12	II	standard	20%	3x VAD	PR	2x Mel 200	interferon, NA	CR
18	M	65	IgG lambda	bone disease	11	I	standard	70%	3x VAD	VGPR	2x Mel 200	thalidomide, 2	CR
19	M	75	IgA lambda	bone disease	11	I	standard	80%	3x PAD	VGPR	2x Mel 200	bortezomib, 2	CR
20	F	61	IgG kappa	bone disease	11	III	standard	30%	3x VAD	PR	1x Mel 200	thalidomide, 3	CR
21	M	55	IgG kappa	bone disease	10	II	high risk (gain 1q21)	50%	3x PAD	VGPR	2x Mel 200	bortezomib, 2	CR
22	M	68	IgG lambda	renal failure	9.5	III	high risk (t4;14)	80%	3x PAD	VGPR	2x Mel 200	bortezomib, 2	CR
23	M	60	IgG kappa	bone disease, hypercalcemia	9	I	standard	30%	3x TAD	CR	1x Mel 200	None	CR
24	F	46	IgG lambda	bone disease	7	II	standard	50%	3x VCD	PR	2x Mel 200	lenalidomide, 1	CR

Table 1: Characteristics of patients with Multiple Myeloma in LTR: Abbreviations: ASCT= autologous stem cell transplantation; BM= bone marrow; CR= Complete remission; Mel = melphalan; NA = not available; n.a. not applicable; PC = plasma cells; PR= partial response; PAD= bortezomib – doxorubicin- dexamethasone; TAD = thalidomide- doxorubicin- dexamethasone; VAD= vincristine – doxorubicin – dexamethasone; VCD= bortezomib- cyclophosphamide- dexamethasone; VID= vincristine – ifosfamide – dexamethasone; VGPR= very good partial response

Table 1: Clinical characterization of Multiple Myeloma patients in long-term survival cohort.

Source (collaborators: Dr. med. Raphael Lutz and Prof. Dr. med. Michael Hundemer)

cellTypes	Clinical_groups	Estimate	Std..Error	Wald_stat_test	P_Value
Pro- and Pre- B-cells	Complete Remission	-0,775814032	0,143021319	-5,424464266	5,81286E-08
Plasmacytoid dendritic cells (pDCs)	Complete Remission	0,515817173	0,130682221	3,947110537	7,91E-05
Plasmacytoid dendritic cells (pDCs)	Non Complete Remission	1,076135715	0,119950195	8,971521215	2,92447E-19
Plasmacytoid dendritic cells (pDCs)	Before Treatment	1,017122693	0,120565977	8,436233168	3,27729E-17
Plasma cells	Non Complete Remission	2,35955881	0,094817843	24,88517717	1,0767E-136
Plasma cells	Before Treatment	3,992772879	0,092436578	43,19472823	0
NKT cells	Non Complete Remission	-0,982142562	0,081129531	-12,10585775	9,8336E-34
NK cells	Complete Remission	0,637037688	0,052409976	12,15489377	5,4032E-34
NK cells	Non Complete Remission	0,378905864	0,051813996	7,312809167	2,61614E-13
NK cells	Before Treatment	1,382804007	0,048599248	28,45319734	4,4488E-178
Neutrophils-T-cells	Complete Remission	1,357969832	0,09443066	14,38060297	6,84952E-47
Neutrophils-T-cells	Non Complete Remission	-0,444635732	0,11902298	-3,735713307	0,000187184
Neutrophils-T-cells	Before Treatment	0,774388367	0,097122679	7,973301106	1,54491E-15
Neutrophils	Complete Remission	0,665782556	0,046540593	14,30541611	2,02423E-46
Neutrophils	Non Complete Remission	-1,020856056	0,051859639	-19,68498195	2,9004E-86
Neutrophils	Before Treatment	-0,828277452	0,049211847	-16,83085472	1,44994E-63
Myeloma Associated Neutrophils (MAN)	Complete Remission	1,73449992	0,468898031	3,699098323	0,000216367
Myeloma Associated Neutrophils (MAN)	Non Complete Remission	4,234558317	0,450776418	9,393921579	5,78069E-21
Myeloma Associated Neutrophils (MAN)	Before Treatment	6,157117968	0,44949448	13,69787225	1,04543E-42
Monocytes	Complete Remission	1,947112141	0,089446315	21,7685003	4,6148E-105
Monocytes	Before Treatment	0,841420051	0,09342722	9,006155262	2,13407E-19
MkP - MEP	Complete Remission	0,756984172	0,133613232	5,665488069	1,46607E-08
MkP - MEP	Non Complete Remission	0,626328502	0,132862804	4,714099684	2,42782E-06
MkP - MEP	Before Treatment	1,451127081	0,123125545	11,78575156	4,62278E-32
Mature B-cells	Complete Remission	0,527997249	0,048504656	10,88549621	1,35161E-27
Mature B-cells	Non Complete Remission	-0,200709862	0,049457218	-4,058252181	4,94414E-05
Immature B-cells	Complete Remission	-1,257479028	0,162943805	-7,717255821	1,18861E-14
Immature B-cells	Before Treatment	-0,498594043	0,122543819	-4,068700039	4,72762E-05
HSPCs	Non Complete Remission	-0,743469473	0,119573096	-6,217698595	5,04499E-10
HSPCs	Before Treatment	0,741163435	0,09149275	8,100788677	5,4604E-16
Erythroid cells	Before Treatment	2,873727025	0,19182643	14,98087111	9,79251E-51
Dendritic cell / monocyte progenitors	Complete Remission	0,64425913	0,126540071	5,091344777	3,55533E-07
Dendritic cell / monocyte progenitors	Non Complete Remission	0,451591939	0,126838324	3,56037453	0,000370326
Dendritic cell / monocyte progenitors	Before Treatment	0,516215404	0,126684509	4,074810778	4,60518E-05
Common myeloid progenitor	Complete Remission	0,581409927	0,084909689	6,847392038	7,52084E-12
Common myeloid progenitor	Non Complete Remission	-0,426504235	0,094411935	-4,517482187	6,25793E-06
Common myeloid progenitor	Before Treatment	0,40917171	0,083211904	4,917225639	8,77794E-07
cDC / Monocytes	Complete Remission	1,145353742	0,109516861	10,45824116	1,34326E-25
cDC / Monocytes	Non Complete Remission	0,769820093	0,11148042	6,905428715	5,0052E-12
cDC / Monocytes	Before Treatment	0,798442602	0,109667913	7,280548914	3,32465E-13
cDC	Complete Remission	2,224744204	0,428244525	5,195032446	2,04684E-07
cDC	Non Complete Remission	2,057002281	0,428356782	4,802077068	1,57028E-06
cDC	Before Treatment	2,034548235	0,432089471	4,708627202	2,49391E-06
CD8+ T-cells	Complete Remission	0,575225635	0,045582251	12,61950929	1,64845E-36
CD8+ T-cells	Before Treatment	0,817160466	0,043056086	18,97897718	2,54502E-80
CD4+ T-cells	Non Complete Remission	-0,469412214	0,049809594	-9,424132502	4,33675E-21
CD4+ T-cells	Before Treatment	0,23902489	0,04610373	5,184502177	2,16593E-07

**Table 2: GLMM model results of the whole bone marrow cell types across the clinical groups.**

cellTypes	Clinical_groups	Estimate	Std..Error	Wald_stat_test	P_Value
CD8+_Cycling_Cells	Complete Remission	1,39266653	0,187150228	7,441436438	9,96E-14
CD8+_KLRB1+	Complete Remission	-1,666495772	0,085227497	-19,55349887	3,85E-85
CD4+_Effector	Complete Remission	0,543798939	0,032364133	16,80251812	2,34E-63
CD8+_Abberant_Memory_Cytotoxic	Complete Remission	2,063250239	0,148294719	13,91317403	5,27E-44
CD8+_Early_Memory	Complete Remission	0,766509206	0,041802603	18,33639889	4,24E-75
CD8+_Effector_Cytotoxic	Complete Remission	1,496065606	0,029967107	49,92359148	0
CD8+_Effector_Cytotoxic_Gamma_Delta	Complete Remission	0,764093228	0,149080426	5,125375954	2,97E-07
CD8+_Late_Memory	Complete Remission	1,398739852	0,047095009	29,7003843	7,6E-194
CD8+_Memory_Stem_Cell	Complete Remission	-0,231267695	0,050627169	-4,568055081	4,92E-06
CD8+_Naive	Complete Remission	0,848195759	0,03106437	27,30445717	3,8E-164
CD4+_Treg	Non Complete Remission	0,693675336	0,088669318	7,823172103	5,15E-15
CD8+_Cycling_Cells	Non Complete Remission	1,68258525	0,18362611	9,16310457	5,04E-20
CD8+_KLRB1+	Non Complete Remission	-0,942081371	0,067933579	-13,86768349	9,94E-44
CD4+_Effector	Non Complete Remission	0,63494728	0,032960269	19,26402021	1,08E-82
CD8+_Abberant_Memory_Cytotoxic	Non Complete Remission	2,027756474	0,149967966	13,5212641	1,17E-41
CD8+_Early_Memory	Non Complete Remission	1,585492302	0,039440993	40,19909664	0
CD8+_Effector_Cytotoxic	Non Complete Remission	0,961825501	0,031699581	30,34189953	3,2E-202
CD8+_Effector_Cytotoxic_Gamma_Delta	Non Complete Remission	1,134041849	0,145458638	7,796318337	6,37E-15
CD8+_Late_Memory	Non Complete Remission	2,026481907	0,045738286	44,30603114	0
CD8+_Naive	Non Complete Remission	0,301824683	0,033752935	8,94217591	3,82E-19
CD8+_Cycling_Cells	Before Treatment	1,500161484	0,170218234	8,813165579	1,22E-18
CD8+_KLRB1+	Before Treatment	-0,853705508	0,046770602	-18,25303663	1,96E-74
CD4+_Effector	Before Treatment	-0,472019336	0,028461325	-16,58458747	9,01E-62
CD8+_Abberant_Memory_Cytotoxic	Before Treatment	4,287677327	0,137635648	31,15237498	4,7E-213
CD8+_Early_Memory	Before Treatment	0,200427645	0,036724413	5,45761323	4,83E-08
CD8+_Effector_Cytotoxic	Before Treatment	0,301330007	0,026411567	11,40901664	3,77E-30
CD8+_Effector_Cytotoxic_Gamma_Delta	Before Treatment	2,362177685	0,120801396	19,55422512	3,8E-85
CD8+_Late_Memory	Before Treatment	1,256978612	0,04197153	29,94836275	4,6E-197
CD8+_Memory_Stem_Cell	Before Treatment	-0,265321603	0,039683146	-6,686002231	2,29E-11
CD8+_Naive	Before Treatment	-0,33217998	0,027571346	-12,04801454	1,99E-33

**Table 3: GLMM model results of the T cells subtypes across the clinical groups**

6.2 Supplementary figures
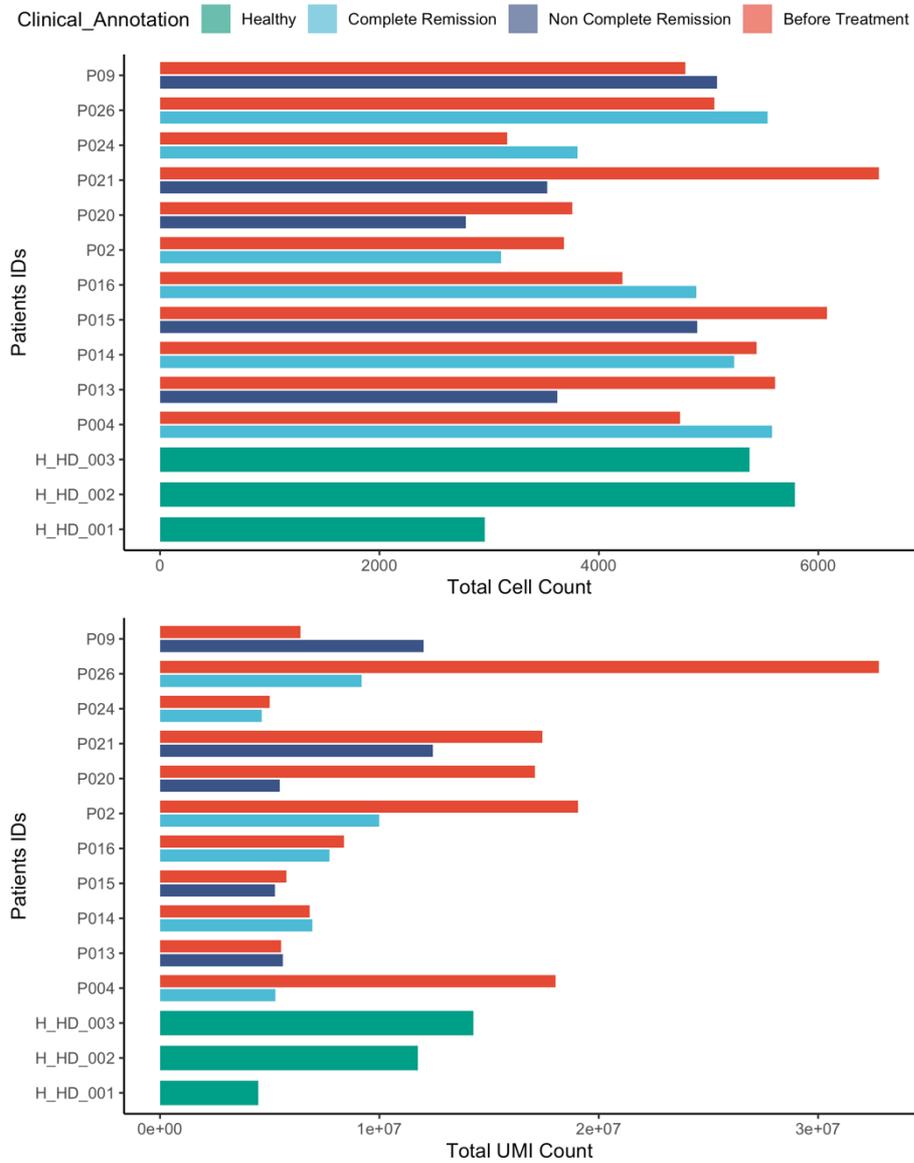
Figure 6.1: Quality control barplots show the total number of cells and total UMI counts per sample for each patient and donor.
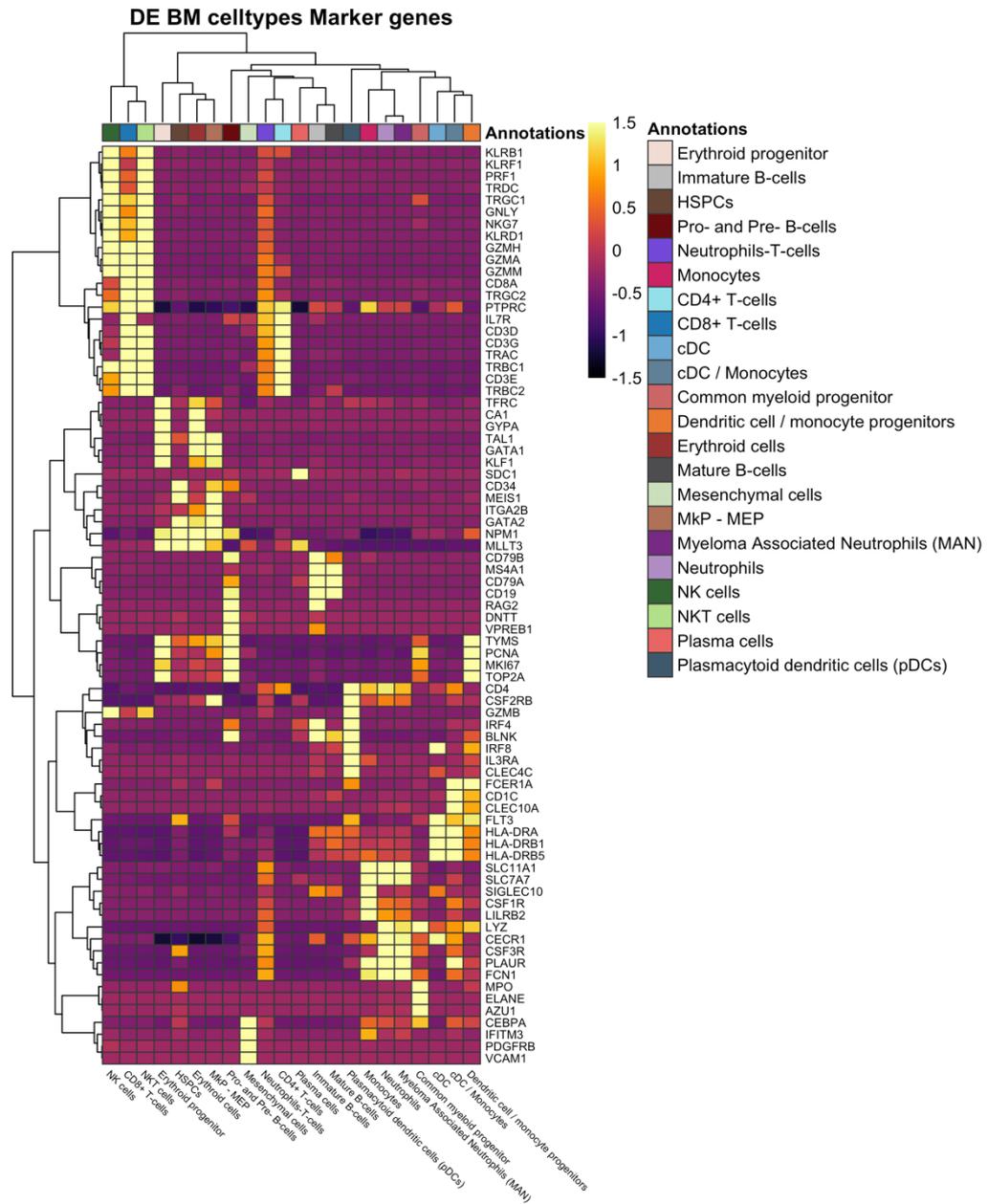
Figure 6.2: The heatmap shows the gene expression of differentially expressed genes of the known cell type markers used for the manual annotations.
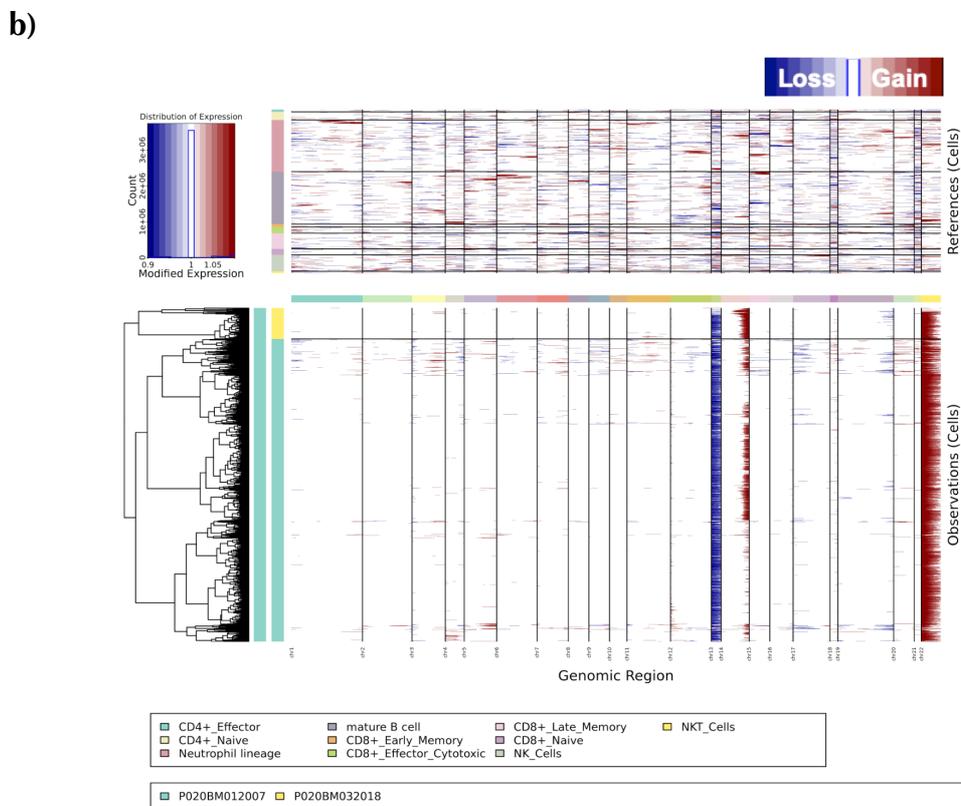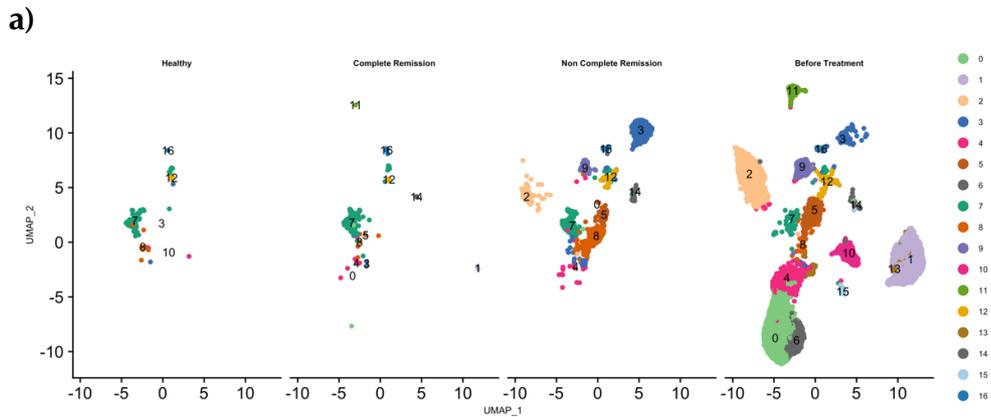
Figure 6.3: The inter- and intra- patient heterogeneity in the plasma cell compartment.

a) UMAP representation of the plasma cell compartment shows different clusters of distribution across clinical groups. **b)** The heatmap shows the Single cell RNA-seq CNV status of patient number 20 at two time points (BT and after LTS).

a)

P020

Before Treatment   Non Complete Remission

Legend:
- CD4+_Effector
- CD4+_Naive
- CD8+_Early_Memory
- CD8+_Effector_Cytotoxic
- CD8+_Late_Memory
- CD8+_Naive
- mature B cell
- Neutrophil lineage
- NK_Cells
- NKT_Cells
- Plasma cells

b)

Before Treatment   Non Complete Remission

proportion_scaled_loss_chr13

Legend next page

c)

d)
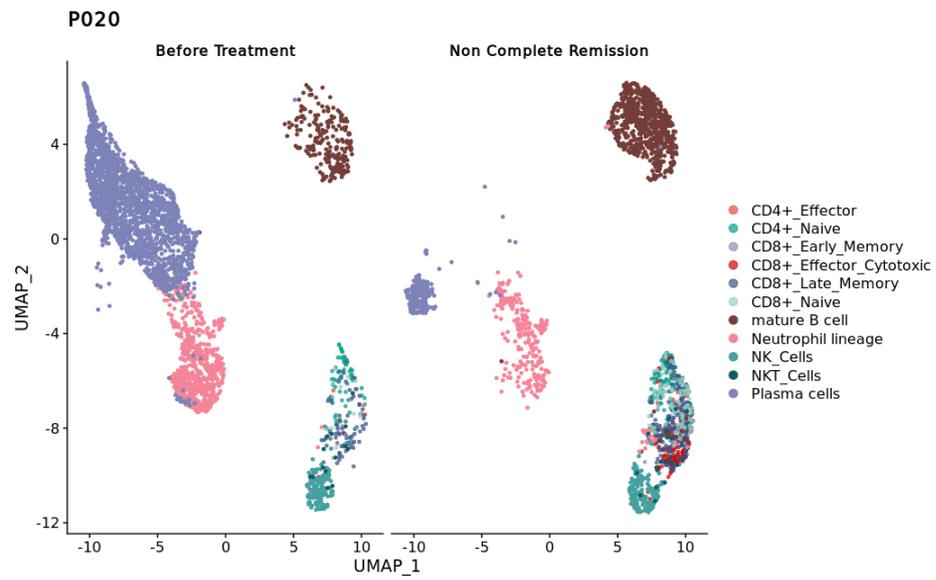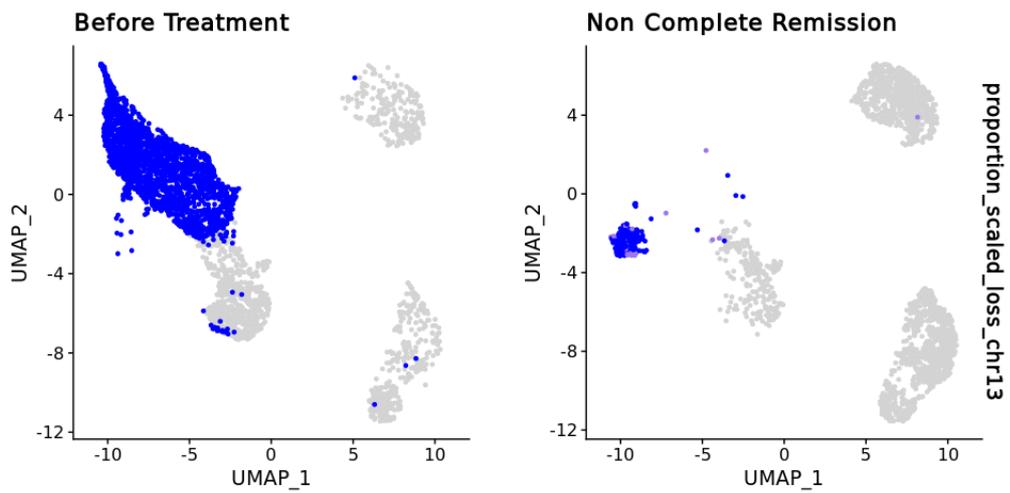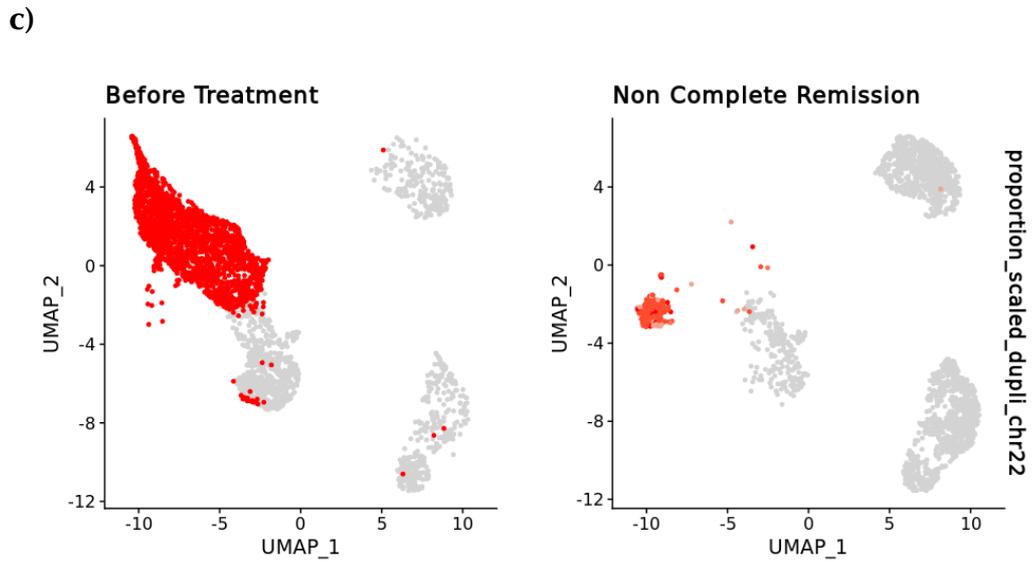
Figure 6.4: UMAP representation of patient 20 samples plasma cells and normal immune cells

a) Cell type annotation. **B)** Chromosome 13 loss in plasma cell compartment. **c)** Chromosome 22 duplication in the plasma cell compartment. **d)** The heatmap shows the DE genes between P20's plasma cells clusters and subpopulations.
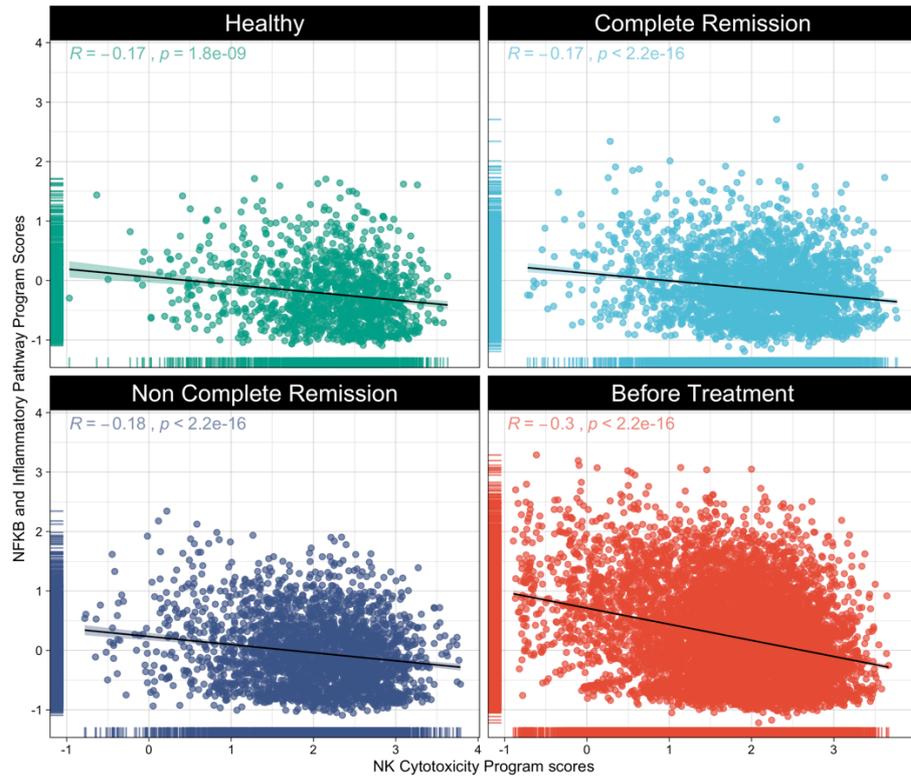
Figure 6.5: Correlation analysis between NFKB and inflammatory pathway scores and NK cytotoxicity program scores.

a)

b)

Legend next page …

162

c)
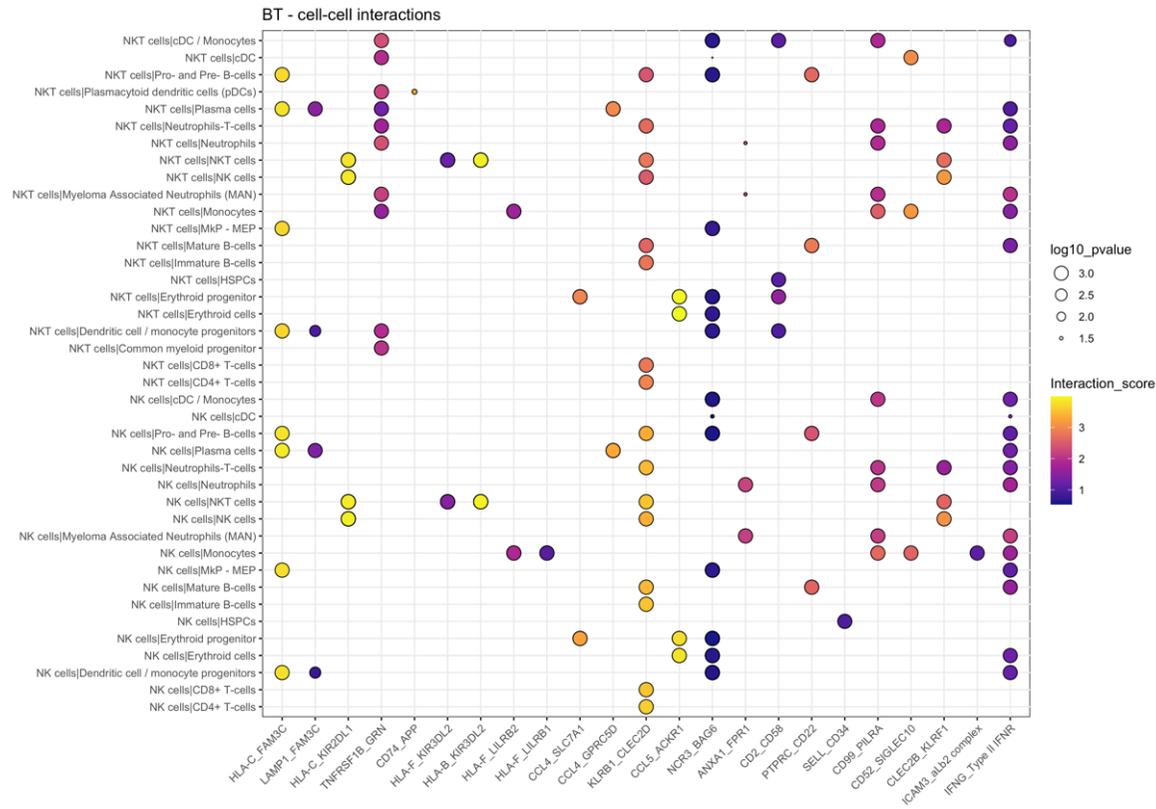
Figure 6.6: Receptor-Ligand interactions between NK cells and other cell types in the BM microenvironment across the clinical groups. a) healthy group b) non-CR c) BT groups.
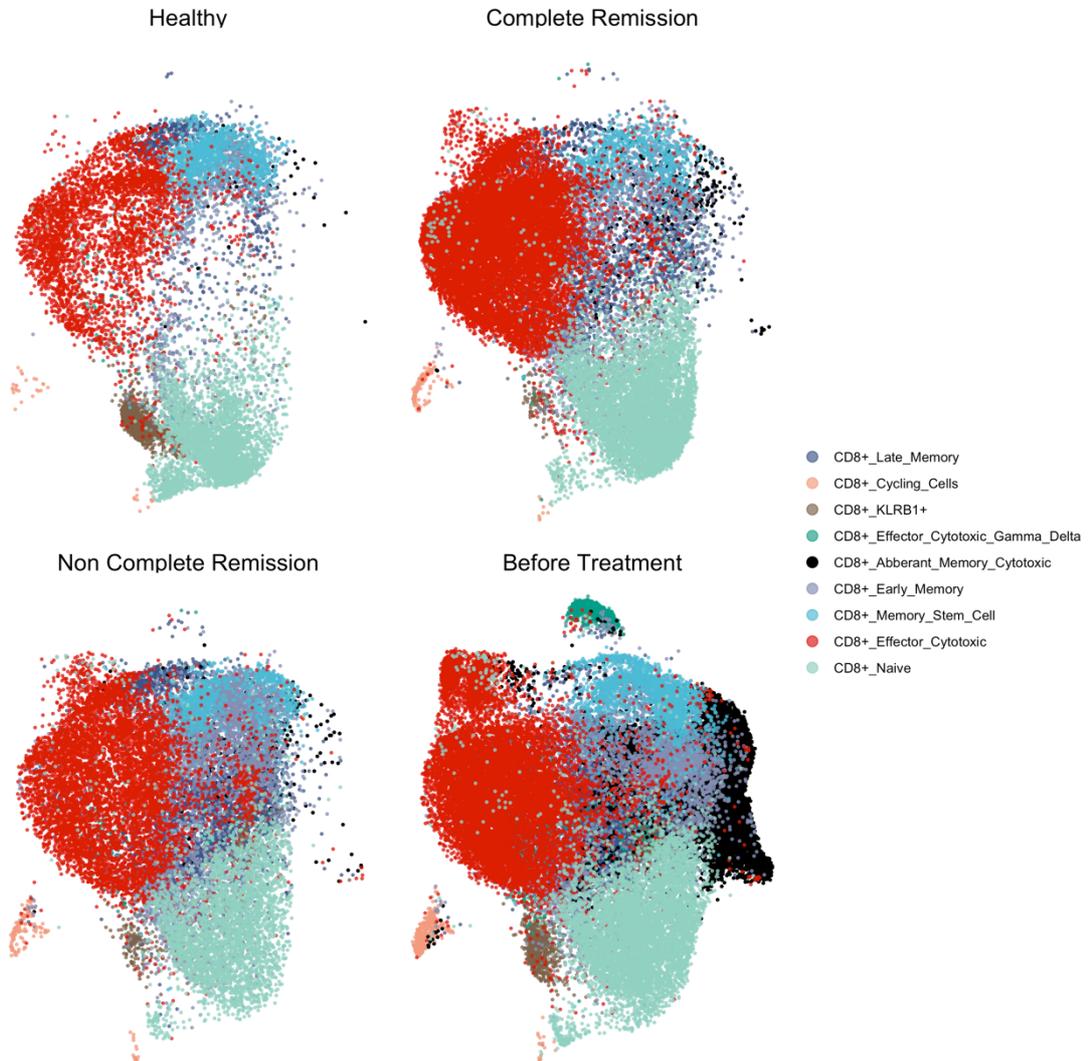
Figure 6.7: UMAP representation shows the exclusive existence of the aberrant Memory Cytotoxic CD8+ T cell population (black colored population) in the BT group.
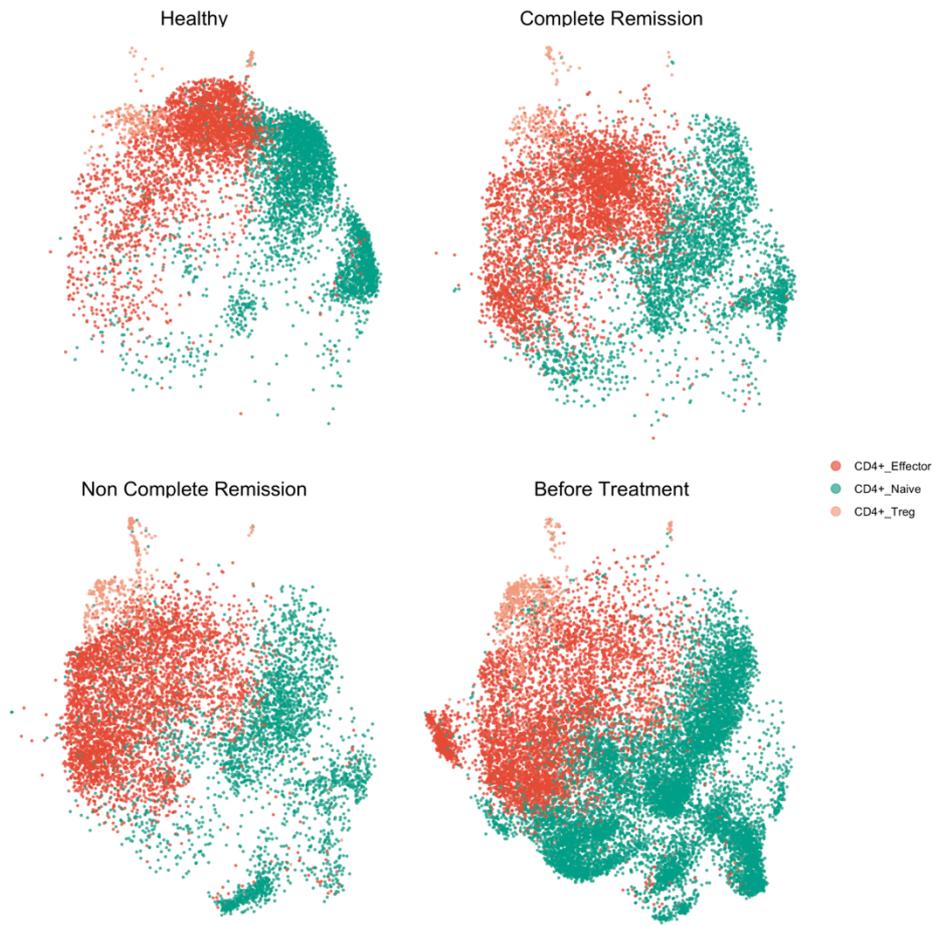
Figure 6.8: UMAP representation of the CD4+ subpopulations across clinical groups.
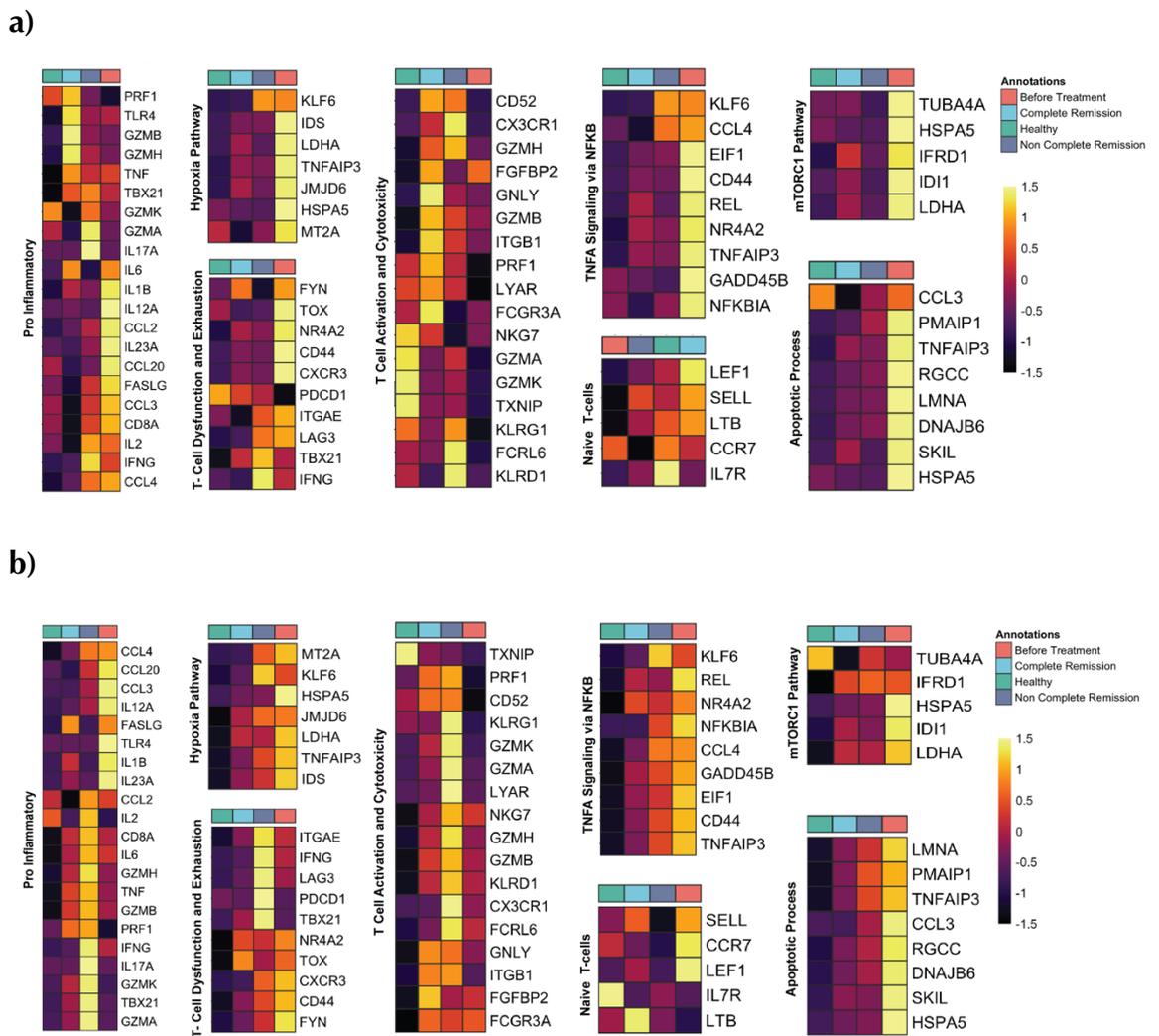
Figure 6.9: The heatmaps show the T cell hallmark pathways and underlying gene expression in the a) CD8⁺ b) CD4⁺ compartment across the clinical groups.

a)

b)

c)
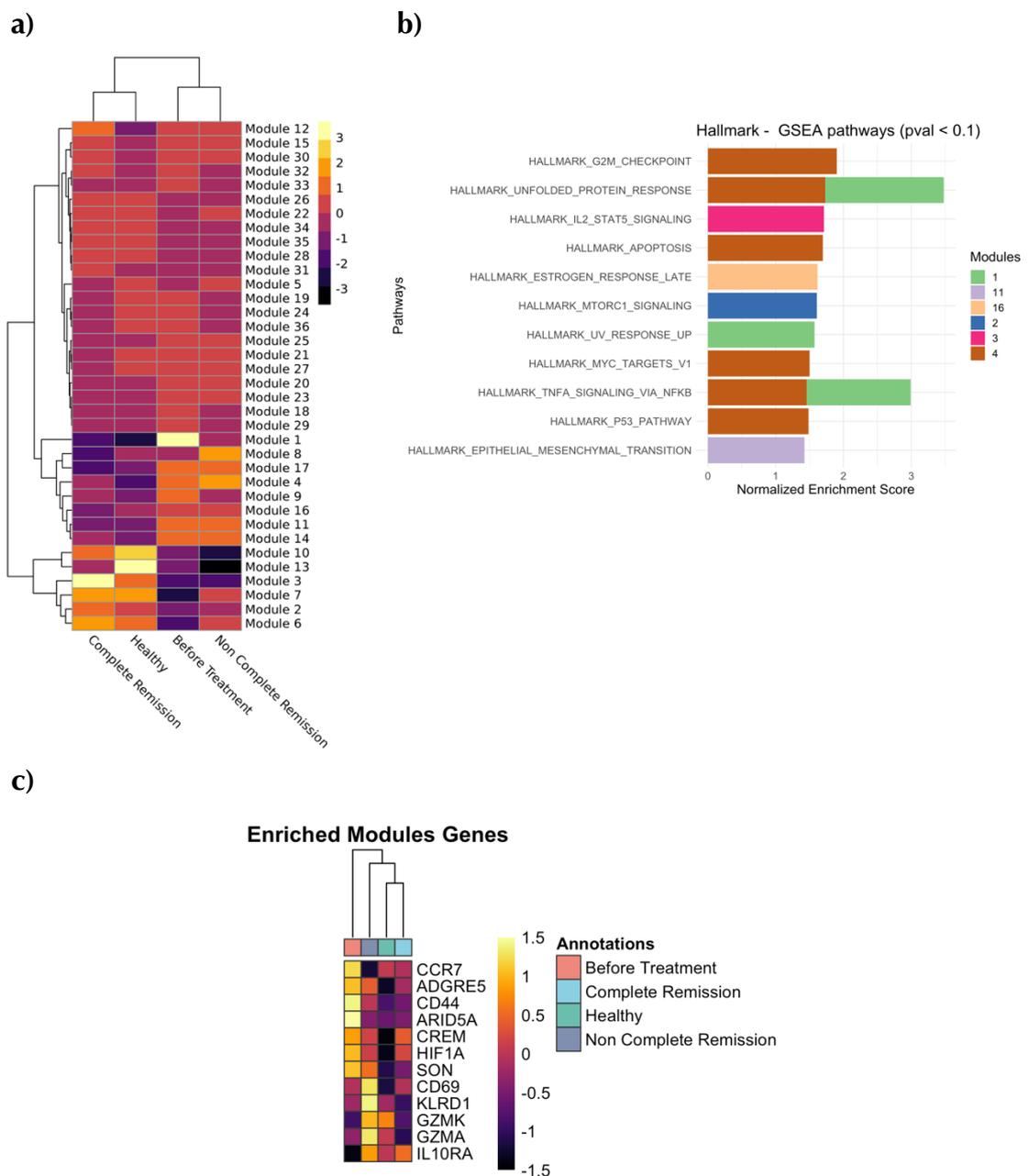
Figure 6.10: Naïve CD8⁺ T cell modules analyses.

a) Upregulated modules in the naïve C8⁺ T cell compartment in each clinical.
b) GSEA enrichment pathways for the naïve CD8⁺ T cell modules. **c)** The heatmap shows many specific underlying modules' gene expression across the clinical groups.
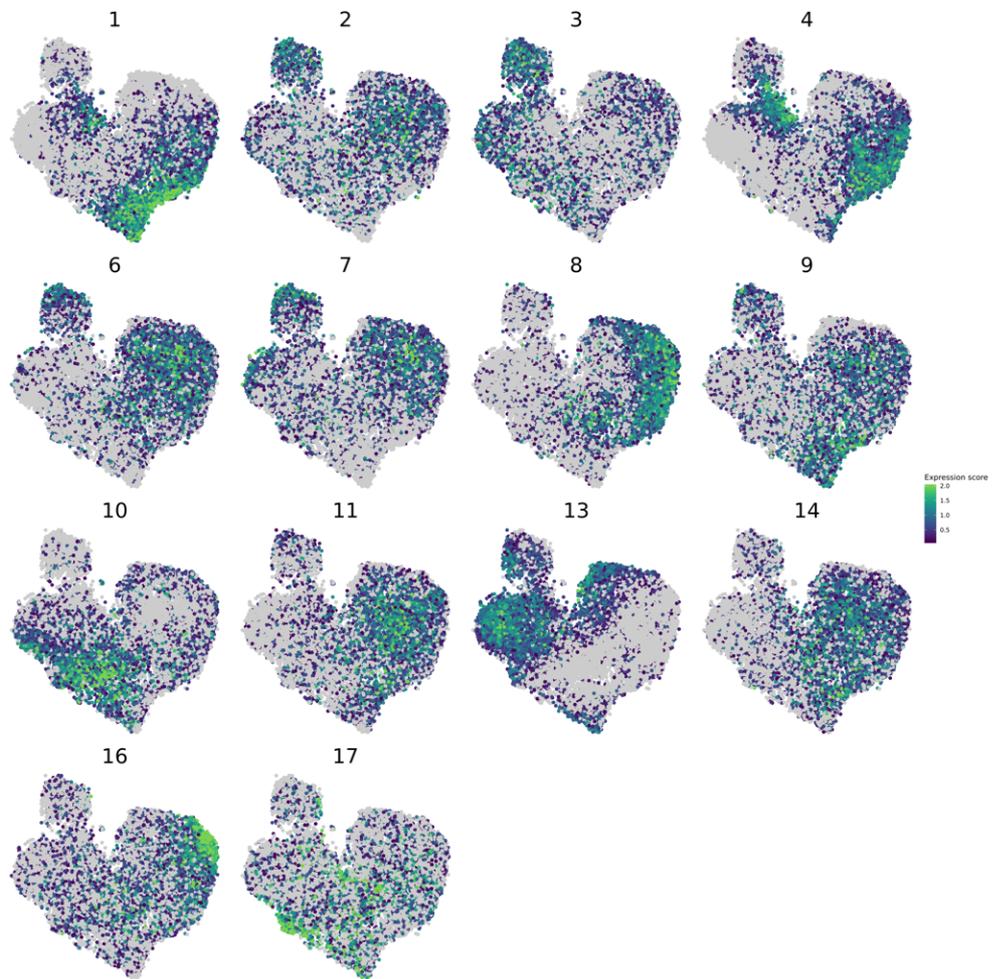
Figure 6.11: UMAP representation shows specific modules expressions which are specific for the clinical groups.

a)

b)

c)
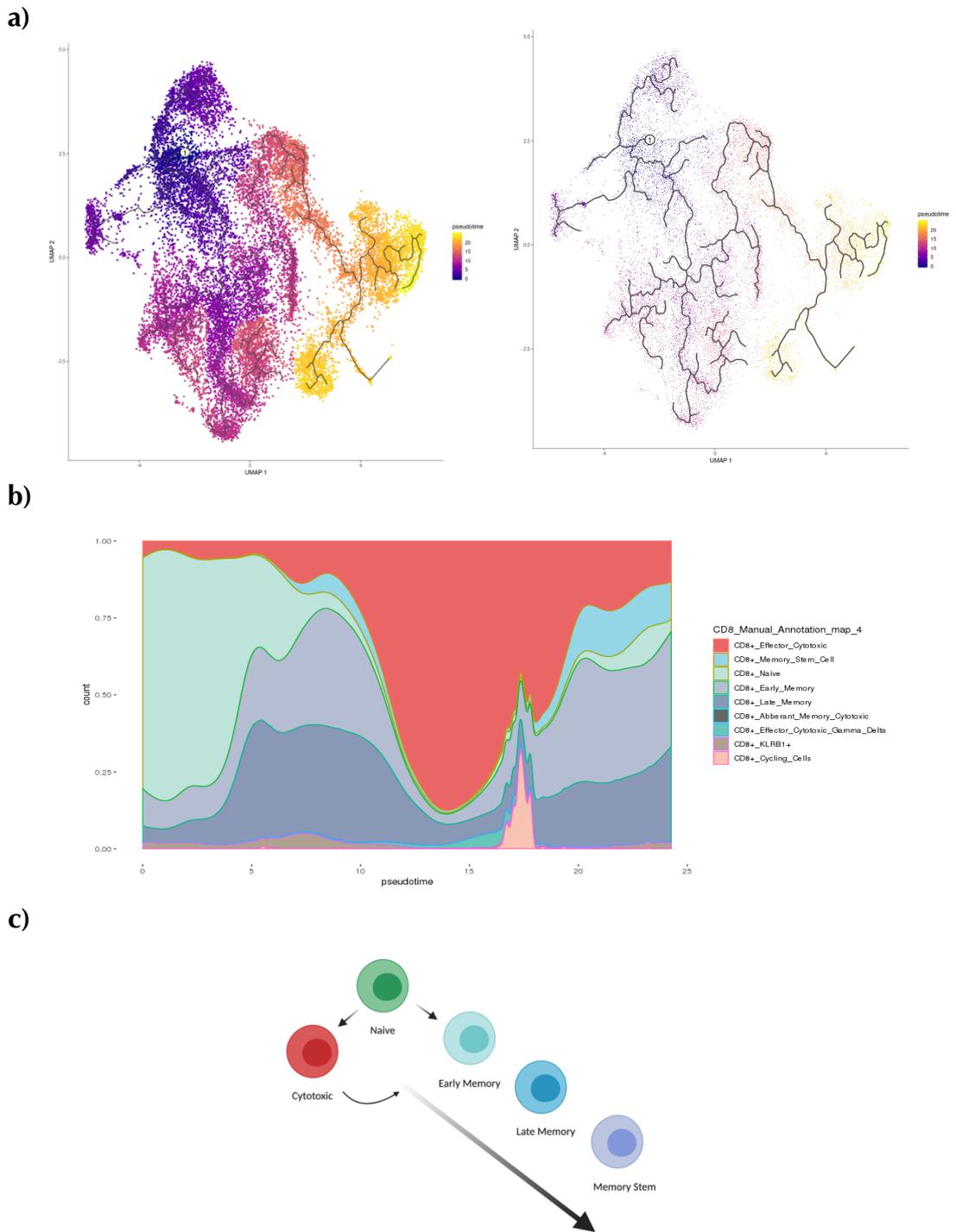
Figure 6.12: Non-Complete Remission CD8+ T cells follow a continuum cell fate differentiation Trajectory

a)

b)

c)
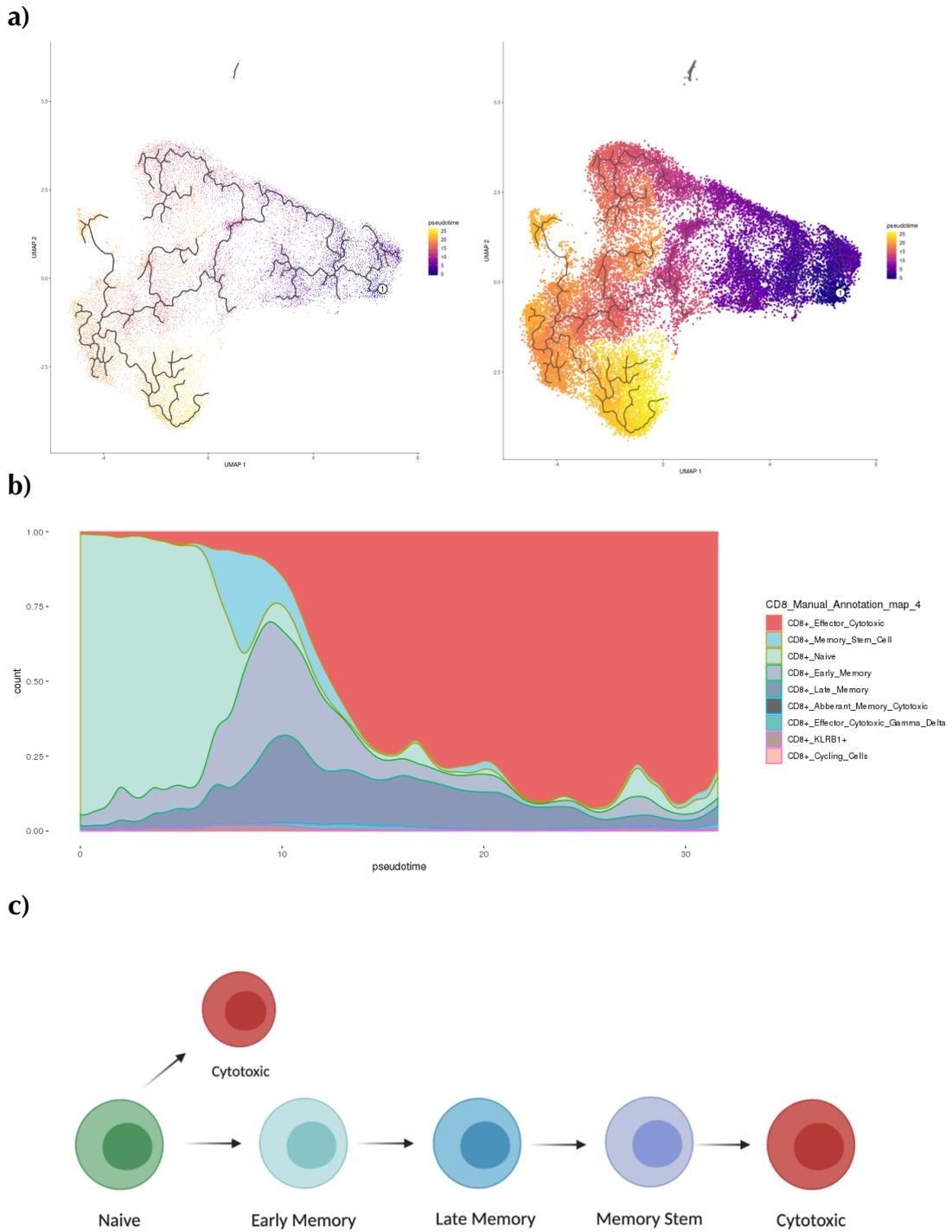
Figure 6.13: Complete Remission CD8⁺ T cells follow a Partial-Linear cell fate differentiation Trajectory.
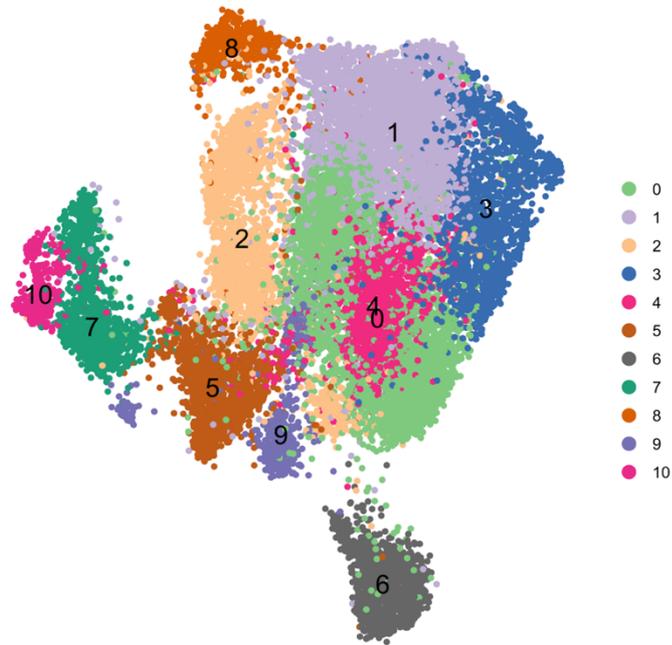
Figure 6.14: UMAP representation of neutrophil clusters across the clinical groups.
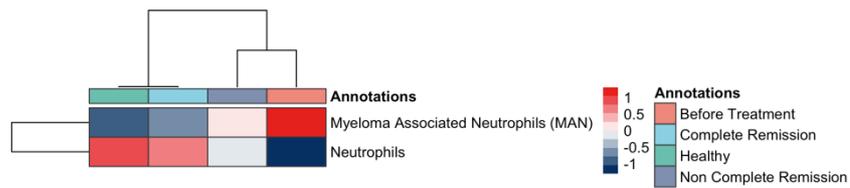
Figure 6.15: The heatmap shows the abundance of the neutrophil's subtypes across the clinical groups

a)
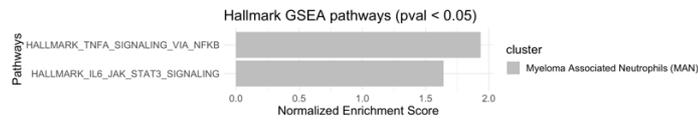
b)
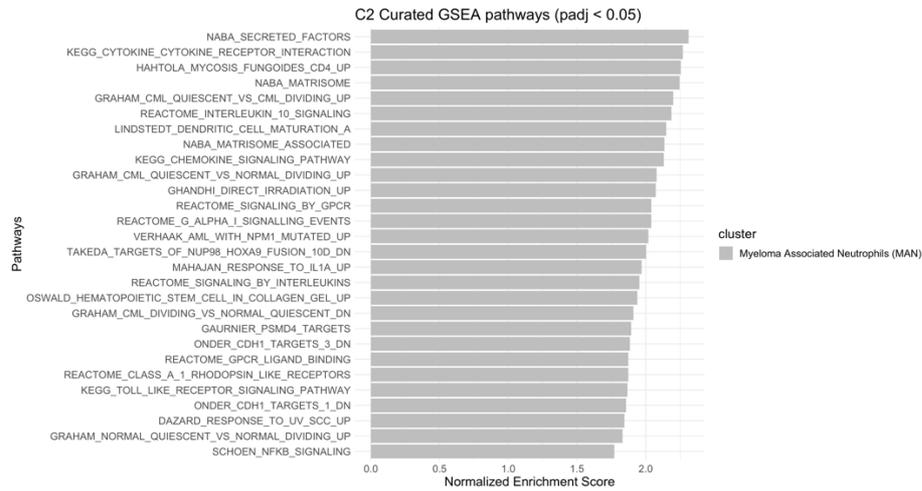
c)
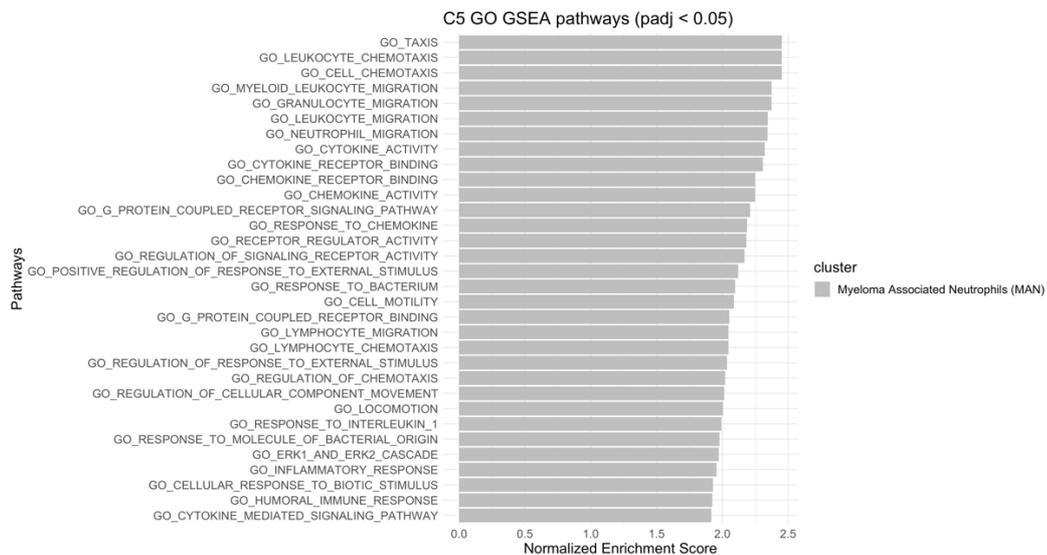
**Figure 6.16: GSEA analysis enriched pathways which are upregulated in MAN cells.**

a)
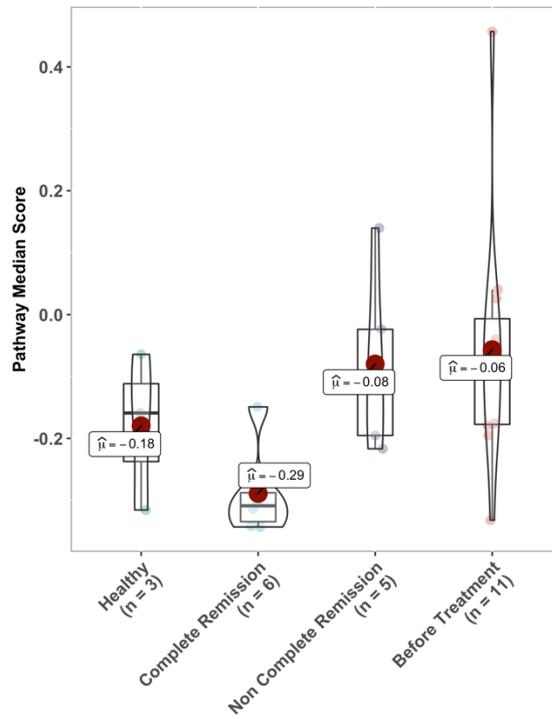
b)
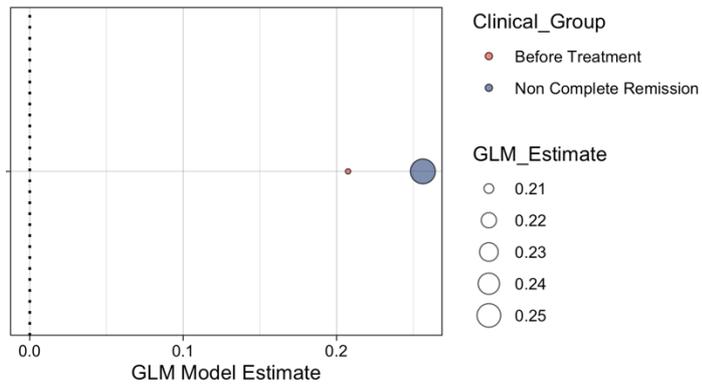
Figure 6.17: a) Median scores of the interferon-alpha and gamma genes across clinical groups. b) GLM model estimate of interferon alpha and gamma genes across clinical groups (p-value < 0.001).
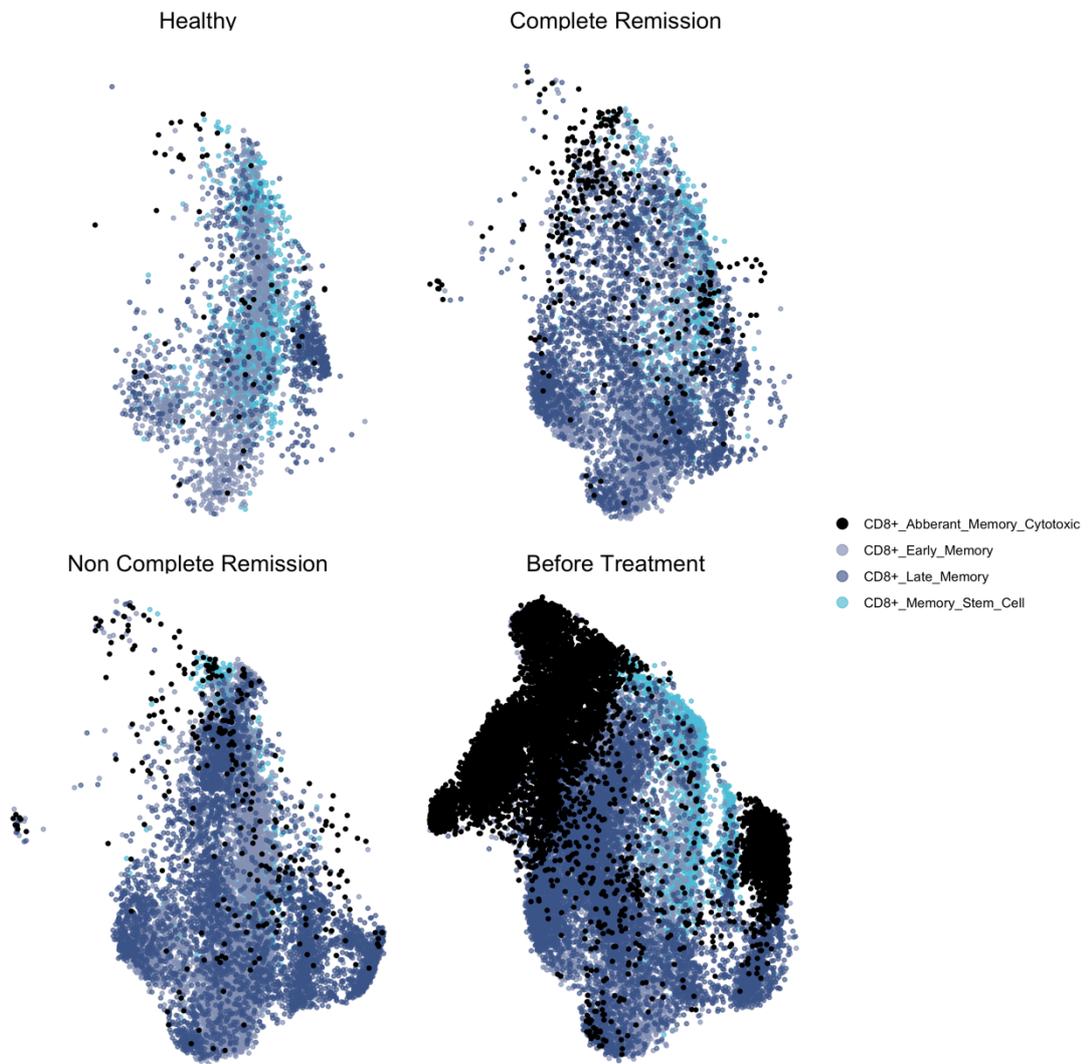
Figure 6.18: UMAP representation of the CD8+ memory T cells subtypes before and after long-term survival
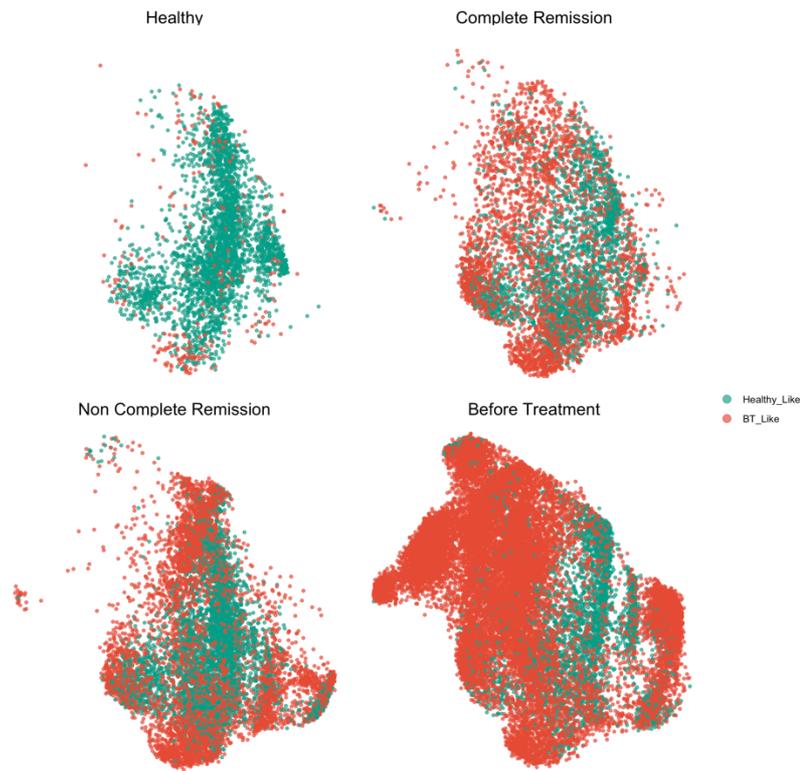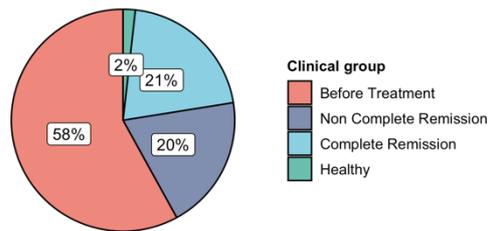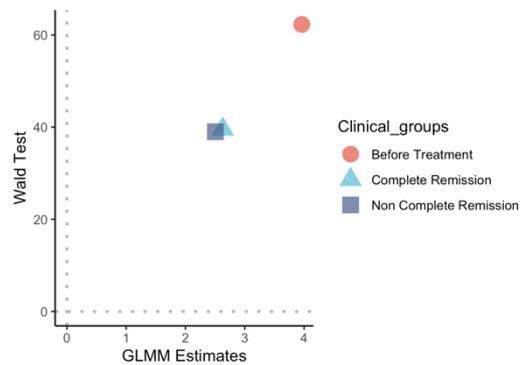
**Figure 6.19: Random forest model prediction for cellular states in the memory CD8+ T cell subtypes.**

a) UMAP representation of the Random Forest Model prediction for Before Treatment and Healthy-like states across all clinical groups **b)** The pie chart shows the proportions of BT-like cells in different clinical groups. c) GLMM estimates of BT-like cells for each clinical group (p-value <0.001).
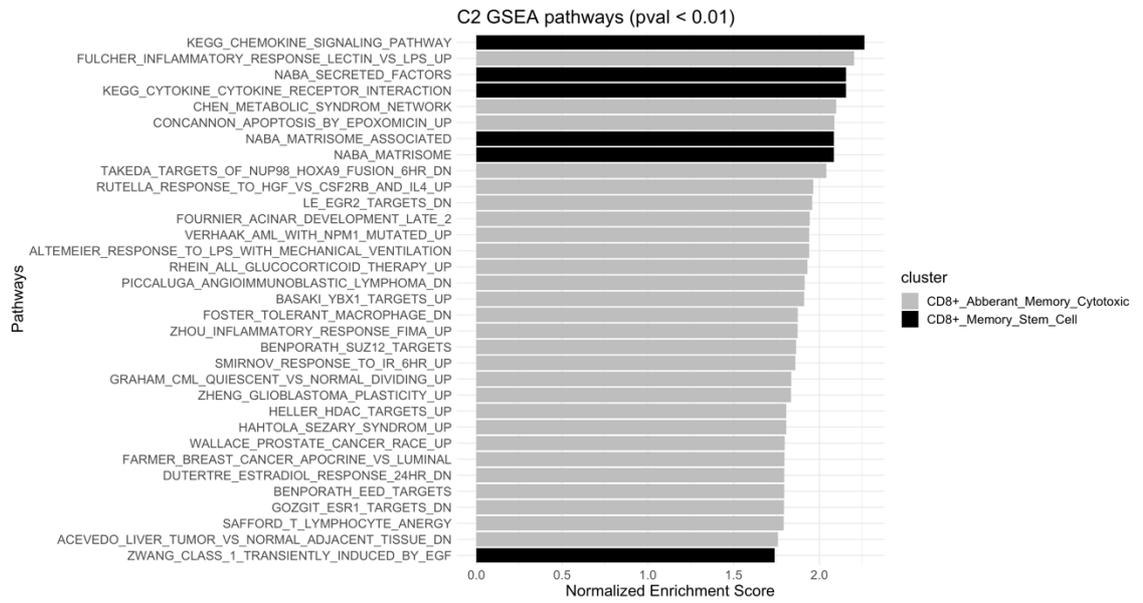
Figure 6.20: GSEA analysis enriched pathways for the CD8+ memory subtypes.
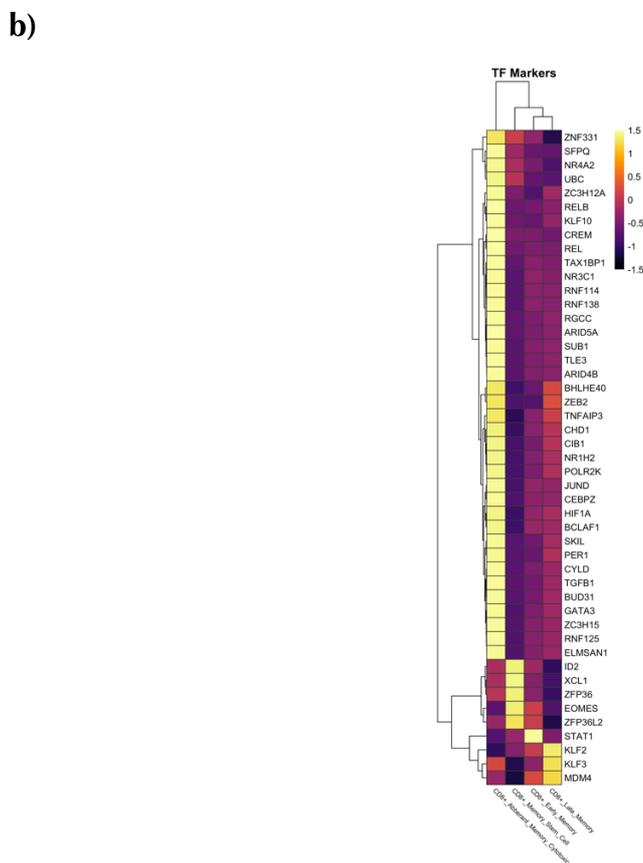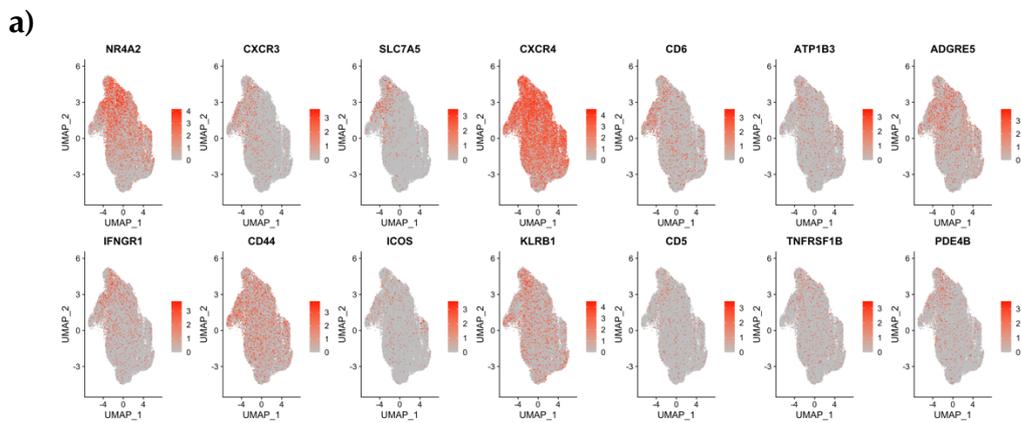
Figure 6.21: AMC CD8⁺ T cell specific surface marker genes and TFs

(I) UMAP representation shows gene expression (red color indicates high gene expression). **(J)** The heatmap shows the differentially expressed transcriptional factors for CD8⁺ memory subtypes.

a)
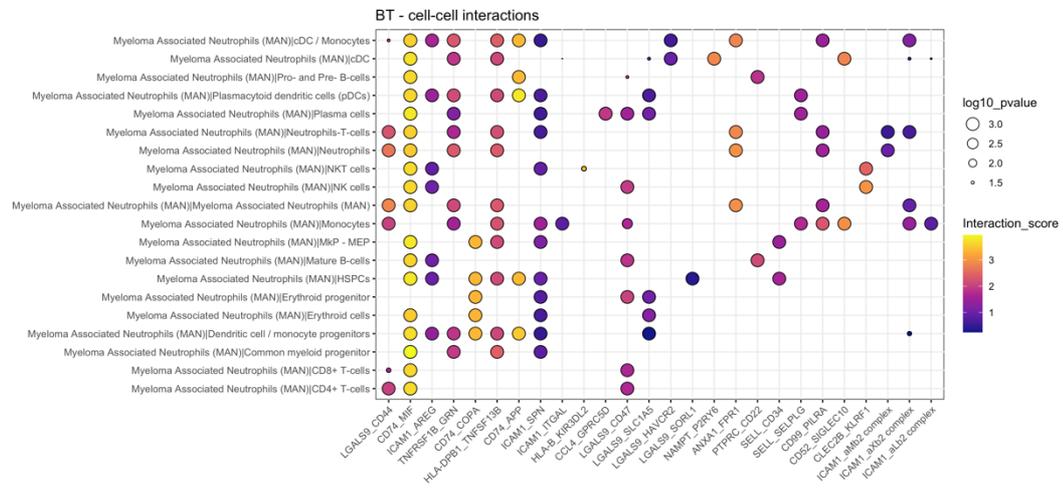
b)
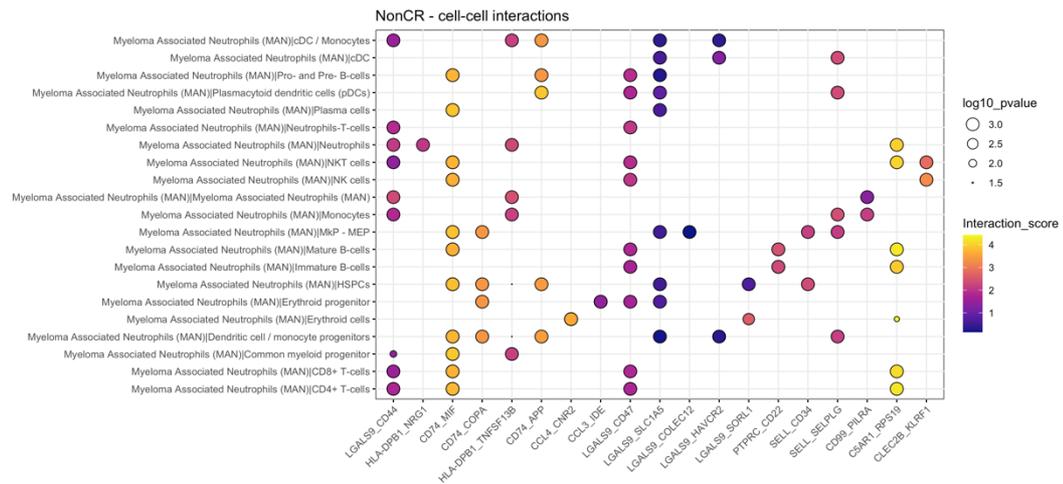

Figure 6.22: Balloon plot shows the receptor-ligand interaction patterns between MAN cells and other cell types in (A) BT group and (B) non-CR group.

The circles' colors represent the mean receptor-ligand interaction scores. The circle size represents the (-Log10) of p-values.
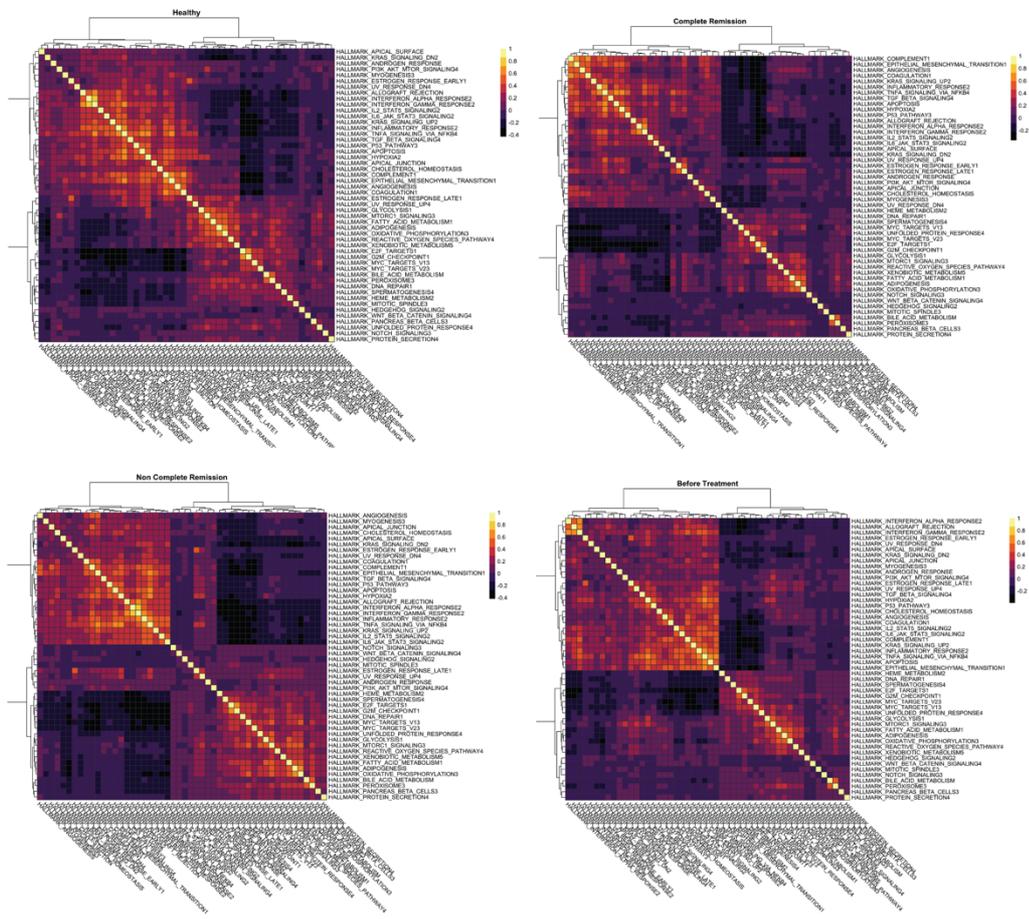
Figure 6.23: Heatmap shows the correlation patterns between the GSEA hallmark pathways and single-cell scores per clinical group.
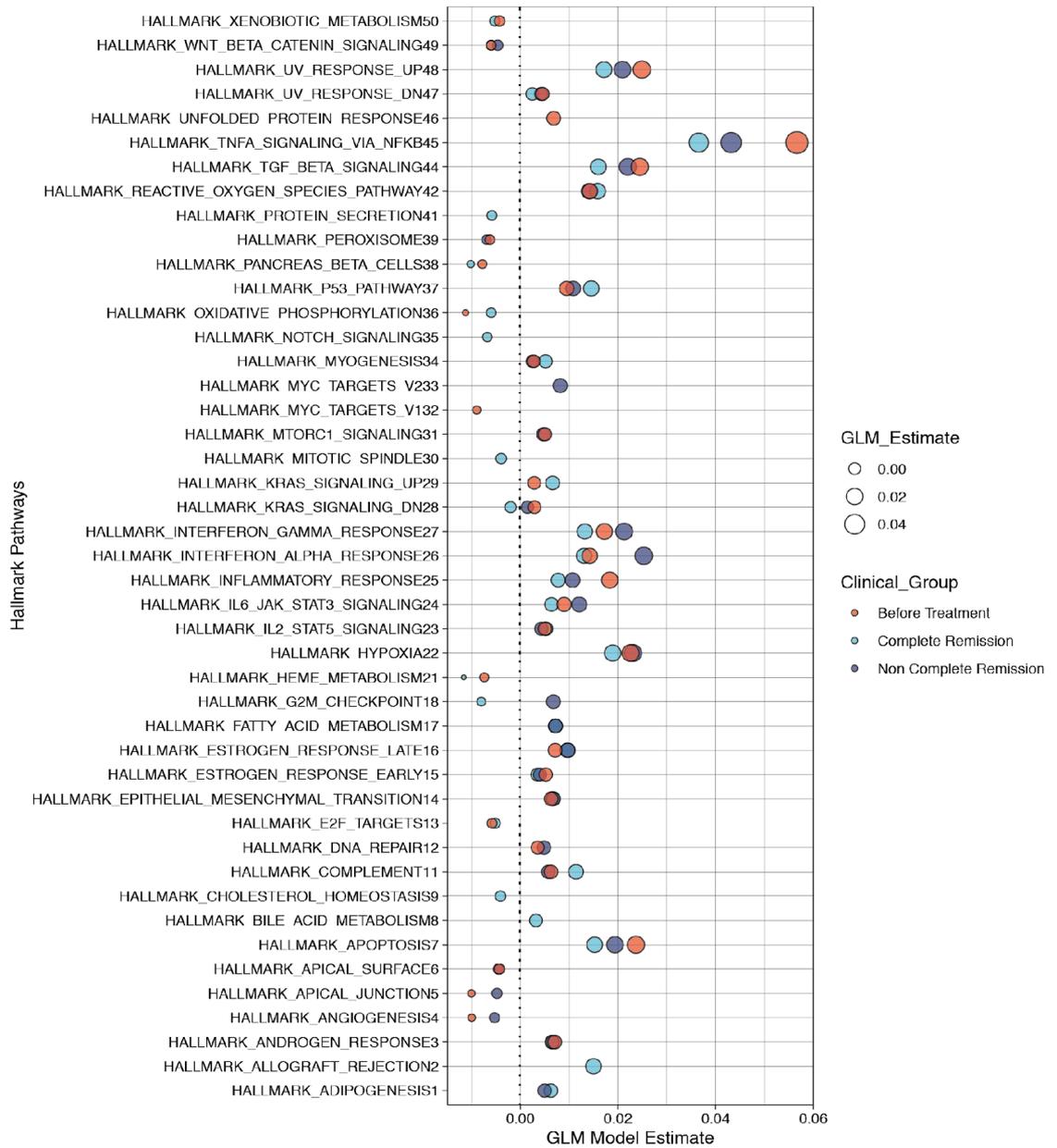
Figure 6.24: GLM model estimates of the hallmark pathways single-cell scores per clinical group.

The circle color is coded by the clinical groups and the size of the circle reflects the GLM model estimate values (p-value < 0.001).

6.3   Software versions and code availability

Hmisc_4.4-0   Formula_1.2-3   survival_3.1-8   lattice_0.20-38   writexl_1.3 wesanderson_0.3.6 Monocle 3.0.1.3 UpSetR_1.4.0   randomForest_4.6-14 cowplot_1.0.0 ggpointdensity_0.1.0 ggpubr_0.3.0   msigdbr_7.0.1   data.table_1.12.8   fgsea_1.12.0 Rcpp_1.0.4.6 reshape2_1.4.4   matrixStats_0.56.0   GSA_1.03.1   plyr_1.8.6 scales_1.1.0 ggsci_2.9   plotly_4.9.2.1   gplots_3.0.3   pheatmap_1.0.12   readxl_1.3.1 forcats_0.5.0   stringr_1.4.0   purrr_0.3.4   readr_1.3.1   tidyr_1.0.2   tibble_3.0.1 ggplot2_3.3.0   tidyverse_1.3.0   viridis_0.5.1 viridisLite_0.3.0   RColorBrewer_1.1-2 dplyr_0.8.5   Seurat_3.1.5.999   backports_1.1.6   fastmatch_1.1-0   igraph_1.2.5 lazyeval_0.2.2   splines_3.6.3   BiocParallel_1.20.1 listenv_0.8.0   digest_0.6.25 htmltools_0.4.0   gdata_2.18.0   fansi_0.4.1   checkmate_2.0.0   magrittr_1.5 cluster_2.1.0   ROCR_1.0-11   openxlsx_4.1.4   globals_0.12.5   modelr_0.1.7 jpeg_0.1-8.1   colorspace_1.4-1   rvest_0.3.5   ggrepel_0.8.2   xfun_0.13 haven_2.2.0   crayon_1.3.4 jsonlite_1.6.1   zoo_1.8-8   ape_5.3   glue_1.4.0 gtable_0.3.0 leiden_0.3.3   car_3.0-7   future.apply_1.5.0 abind_1.4-5   DBI_1.1.0 rstatix_0.5.0   htmlTable_1.13.3   reticulate_1.15   foreign_0.8-75   rsvd_1.0.3 tsne_0.1-3   htmlwidgets_1.5.1   httr_1.4.1 acepack_1.4.1   ellipsis_0.3.0   ica_1.0-2 pkgconfig_2.0.3   nnet_7.3-12   uwot_0.1.8   dbplyr_1.4.3   tidyselect_1.0.0   rlang_0.4.6 munsell_0.5.0   cellranger_1.1.0   tools_3.6.3   cli_2.0.2   generics_0.0.2   broom_0.5.6 ggridges_0.5.2   npsurv_0.4-0   knitr_1.28   fs_1.4.1   fitdistrplus_1.0-14 zip_2.0.4 caTools_1.18.0   RANN_2.6.1 pbapply_1.4-2 future_1.17.0 nlme_3.1-144   xml2_1.3.2 compiler_3.6.3   rstudioapi_0.11   curl_4.3   png_0.1-7   lsei_1.2-0   ggsignif_0.6.0 reprex_0.3.0   stringi_1.4.6   Matrix_1.2-18   vctrs_0.2.4   pillar_1.4.4 lifecycle_0.2.0   lmtest_0.9-37   RcppAnnoy_0.0.16   bitops_1.0-6   irlba_2.3.3 patchwork_1.0.0   R6_2.4.1   latticeExtra_0.6-29 KernSmooth_2.23-16 gridExtra_2.3 rio_0.5.16   codetools_0.2-16   MASS_7.3-51.5   gtools_3.8.2   assertthat_0.2.1 withr_2.2.0   sctransform_0.2.1 parallel_3.6.3   hms_0.5.3 rpart_4.1-15   carData_3.0-3   Rtsne_0.15   lubridate_1.7.8   base64enc_0.1-3 Python 03.07.06 Cell Ranger 3.0.1 CellPhoneDB 2.0 Cytoscape 03.08.00 scVelo 0.2.2 R version 3.6.3

Code availability (GitHub): https://github.com/AMA111/PhD-Thesis-2020

7 Bibliography

Aerts-Toegaert, Cindy, Carlo Heirman, Sandra Tuyaerts, Jurgen Corthals, Joeri L. Aerts, Aude Bonehill, Kris Thielemans, and Karine Breckpot. 2007. "CD83 Expression on Dendritic Cells and T Cells: Correlation with Effective Immune Responses." *European Journal of Immunology* 37 (3): 686–95. https://doi.org/10.1002/eji.200636535.

Agnarelli, Alessandro, Tim Chevassut, and Erika J. Mancini. 2018. "IRF4 in Multiple Myeloma—Biology, Disease and Therapeutic Target." *Leukemia Research* 72 (September): 52–58. https://doi.org/10.1016/j.leukres.2018.07.025.

Ahmed, Rizwan, Zahra Omidian, Adebola Giwa, Benjamin Cornwell, Neha Majety, David R. Bell, Sangyun Lee, et al. 2019. "A Public BCR Present in a Unique Dual-Receptor-Expressing Lymphocyte from Type 1 Diabetes Patients Encodes a Potent T Cell Autoantigen." *Cell* 177 (6): 1583-1599.e16. https://doi.org/10.1016/j.cell.2019.05.007.

Ahn, Eun-Young, Russell C. DeKelver, Miao-Chia Lo, Tuyet Ann Nguyen, Shinobu Matsuura, Anita Boyapati, Shatakshi Pandit, Xiang-Dong Fu, and Dong-Er Zhang. 2011. "SON Controls Cell-Cycle Progression by Coordinated Regulation of RNA Splicing." *Molecular Cell* 42 (2): 185–98. https://doi.org/10.1016/j.molcel.2011.03.014.

Akkaya, Munir, Kihyuck Kwak, and Susan K. Pierce. 2020. "B Cell Memory: Building Two Walls of Protection against Pathogens." *Nature Reviews Immunology* 20 (4): 229–38. https://doi.org/10.1038/s41577-019-0244-2.

Alder, Matthew N., Igor B. Rogozin, Lakshminarayan M. Iyer, Galina V. Glazko, Max D. Cooper, and Zeev Pancer. 2005. "Diversity and Function of Adaptive Immune Receptors in a Jawless Vertebrate." *Science (New York, N.Y.)* 310 (5756): 1970–73. https://doi.org/10.1126/science.1119420.

Almishri, Wagdi, Tania Santodomingo-Garzon, Tyson Le, Danuta Stack, Christopher H. Mody, and Mark G. Swain. 2016. "TNFα Augments Cytokine-Induced NK Cell IFNγ Production through TNFR2." *Journal of Innate Immunity* 8 (6): 617–29. https://doi.org/10.1159/000448077.

Altman, N. S. 1992. "An Introduction to Kernel and Nearest-Neighbor Nonparametric Regression." *The American Statistician* 46 (3): 175–85. https://doi.org/10.1080/00031305.1992.10475879.

Amezquita, Robert A., Aaron T. L. Lun, Etienne Becht, Vince J. Carey, Lindsay N. Carpp, Ludwig Geistlinger, Federico Marini, et al. 2020. "Orchestrating Single-Cell Analysis with Bioconductor." *Nature Methods* 17 (2): 137–45. https://doi.org/10.1038/s41592-019-0654-x.

Andrews, Tallulah S., and Martin Hemberg. 2018. "False Signals Induced by Single-Cell Imputation." *F1000Research* 7: 1740. https://doi.org/10.12688/f1000research.16613.2.

Aran, Dvir, Agnieszka P. Looney, Leqian Liu, Esther Wu, Valerie Fong, Austin Hsu, Suzanna Chak, et al. 2019. "Reference-Based Analysis of Lung Single-Cell Sequencing Reveals a Transitional Profibrotic Macrophage." *Nature Immunology* 20 (2): 163–72. https://doi.org/10.1038/s41590-018-0276-y.

Arias, Maykel, Luis Martínez-Lostao, Llipsy Santiago, Angel Ferrandez, David J. Granville, and Julián Pardo. 2017. "The Untold Story of Granzymes in Oncoimmunology: Novel Opportunities with Old Acquaintances." *Trends in Cancer* 3 (6): 407–22. https://doi.org/10.1016/j.trecan.2017.04.001.

Baaten, Bas J. G., Cheng-Rui Li, and Linda M. Bradley. 2010. "Multifaceted Regulation of T Cells by CD44." *Communicative & Integrative Biology* 3 (6): 508–12. https://doi.org/10.4161/cib.3.6.13495.

Baccin, Chiara, Jude Al-Sabah, Lars Velten, Patrick M. Helbling, Florian Grünschläger, Pablo Hernández-Malmierca, César Nombela-Arrieta, Lars M. Steinmetz, Andreas Trumpp, and Simon Haas. 2020. "Combined Single-Cell and Spatial Transcriptomics Reveal the Molecular, Cellular and Spatial Bone Marrow Niche Organization." *Nature Cell Biology* 22 (1): 38–48. https://doi.org/10.1038/s41556-019-0439-6.

Bacher, Rhonda, Li-Fang Chu, Ning Leng, Audrey P. Gasch, James A. Thomson, Ron M. Stewart, Michael Newton, and Christina Kendziorski. 2017. "SCnorm: Robust Normalization of Single-Cell RNA-Seq Data." *Nature Methods* 14 (6): 584–86. https://doi.org/10.1038/nmeth.4263.

Bergen, Volker, Marius Lange, Stefan Peidli, F. Alexander Wolf, and Fabian J. Theis. 2020. "Generalizing RNA Velocity to Transient Cell States through Dynamical Modeling." *Nature Biotechnology*, August, 1–7. https://doi.org/10.1038/s41587-020-0591-3.

Binnewies, Mikhail, Edward W. Roberts, Kelly Kersten, Vincent Chan, Douglas F. Fearon, Miriam Merad, Lisa M. Coussens, et al. 2018. "Understanding the Tumor Immune Microenvironment (TIME) for Effective Therapy." *Nature Medicine* 24 (5): 541–50. https://doi.org/10.1038/s41591-018-0014-x.

Blank, Christian U., W. Nicholas Haining, Werner Held, Patrick G. Hogan, Axel Kallies, Enrico Lugli, Rachel C. Lynn, et al. 2019. "Defining 'T Cell Exhaustion.'" *Nature Reviews Immunology* 19 (11): 665–74. https://doi.org/10.1038/s41577-019-0221-9.

Breiman, Leo. 2001. "Random Forests." *Machine Learning* 45 (1): 5–32. https://doi.org/10.1023/A:1010933404324.

Browaeys, Robin, Wouter Saelens, and Yvan Saeys. 2020. "NicheNet: Modeling Intercellular Communication by Linking Ligands to Target Genes." *Nature Methods* 17 (2): 159–62. https://doi.org/10.1038/s41592-019-0667-5.

Butler, Andrew, Paul Hoffman, Peter Smibert, Efthymia Papalexi, and Rahul Satija. 2018. "Integrating Single-Cell Transcriptomic Data across Different Conditions, Technologies, and Species." *Nature Biotechnology* 36 (5): 411–20. https://doi.org/10.1038/nbt.4096.

Caliendo, Fabio, Marina Dukhinova, and Velia Siciliano. 2019. "Engineered Cell-Based Therapeutics: Synthetic Biology Meets Immunology." *Frontiers in Bioengineering and Biotechnology* 7. https://doi.org/10.3389/fbioe.2019.00043.

Cambier, John C., Stephen B. Gauld, Kevin T. Merrell, and Barbara J. Vilen. 2007. "B-Cell Anergy: From Transgenic Models to Naturally Occurring Anergic B Cells?" *Nature Reviews Immunology* 7 (8): 633–43. https://doi.org/10.1038/nri2133.

Campbell, Peter J., Gad Getz, Jan O. Korbel, Joshua M. Stuart, Jennifer L. Jennings, Lincoln D. Stein, Marc D. Perry, et al. 2020. "Pan-Cancer Analysis of Whole Genomes." *Nature* 578 (7793): 82–93. https://doi.org/10.1038/s41586-020-1969-6.

Cao, Junyue, Malte Spielmann, Xiaojie Qiu, Xingfan Huang, Daniel M. Ibrahim, Andrew J. Hill, Fan Zhang, et al. 2019. "The Single-Cell Transcriptional Landscape of Mammalian Organogenesis." *Nature* 566 (7745): 496–502. https://doi.org/10.1038/s41586-019-0969-x.

Cerwenka, Adelheid, and Lewis L. Lanier. 2016. "Natural Killer Cell Memory in Infection, Inflammation and Cancer." *Nature Reviews Immunology* 16 (2): 112–23. https://doi.org/10.1038/nri.2015.9.

Chen, Joyce, Isaac F. López-Moyado, Hyungseok Seo, Chan-Wang J. Lio, Laura J. Hempleman, Takashi Sekiya, Akihiko Yoshimura, James P. Scott-Browne, and Anjana Rao. 2019. "NR4A Transcription Factors Limit CAR T Cell Function in Solid Tumours." *Nature* 567 (7749): 530–34. https://doi.org/10.1038/s41586-019-0985-x.

Chi, Xiying, Yue Li, and Xiaoyan Qiu. 2020. "V(D)J Recombination, Somatic Hypermutation and Class Switch Recombination of Immunoglobulins: Mechanism and Regulation." *Immunology* 160 (3): 233–47. https://doi.org/10.1111/imm.13176.

Choi, Kwangbom, Yang Chen, Daniel A. Skelly, and Gary A. Churchill. 2020. "Bayesian Model Selection Reveals Biological Origins of Zero Inflation in Single-Cell Transcriptomics." *Genome Biology* 21 (1): 183. https://doi.org/10.1186/s13059-020-02103-2.

Chow, Melvyn T., Aleksandra J. Ozga, Rachel L. Servis, Dennie T. Frederick, Jennifer A. Lo, David E. Fisher, Gordon J. Freeman, Genevieve M. Boland, and Andrew D. Luster. 2019. "Intratumoral Activity of the CXCR3 Chemokine System Is Required for the Efficacy of Anti-PD-1 Therapy." *Immunity* 50 (6): 1498-1512.e5. https://doi.org/10.1016/j.immuni.2019.04.010.

Collin, Matthew, and Venetia Bigley. 2018. "Human Dendritic Cell Subsets: An Update." *Immunology* 154 (1): 3–20. https://doi.org/10.1111/imm.12888.

Cooper, Max D., and Matthew N. Alder. 2006. "The Evolution of Adaptive Immune Systems." *Cell* 124 (4): 815–22. https://doi.org/10.1016/j.cell.2006.02.001.

Cooper, Max D., Raymond D. A. Peterson, and Robert A. Good. 1965. "Delineation of the Thymic and Bursal Lymphoid Systems in the Chicken." *Nature* 205 (4967): 143–46. https://doi.org/10.1038/205143a0.

Crowell, Helena L., Charlotte Soneson, Pierre-Luc Germain, Daniela Calini, Ludovic Collin, Catarina Raposo, Dheeraj Malhotra, and Mark D. Robinson. 2020. "On the Discovery of Subpopulation-Specific State Transitions from Multi-Sample Multi-Condition Single-Cell RNA Sequencing Data." *BioRxiv*, August, 713412. https://doi.org/10.1101/713412.

Datlinger, Paul, André F. Rendeiro, Thorina Boenke, Thomas Krausgruber, Daniele Barreca, and Christoph Bock. 2019. "Ultra-High Throughput Single-Cell RNA Sequencing by Combinatorial Fluidic Indexing." *BioRxiv*, December, 2019.12.17.879304. https://doi.org/10.1101/2019.12.17.879304.

Davis, Mark M., Cristina M. Tato, and David Furman. 2017. "Systems Immunology: Just Getting Started." *Nature Immunology* 18 (7): 725–32. https://doi.org/10.1038/ni.3768.

Dobin, Alexander, Carrie A. Davis, Felix Schlesinger, Jorg Drenkow, Chris Zaleski, Sonali Jha, Philippe Batut, Mark Chaisson, and Thomas R. Gingeras. 2013. "STAR: Ultrafast Universal RNA-Seq Aligner." *Bioinformatics* 29 (1): 15–21. https://doi.org/10.1093/bioinformatics/bts635.

Doulatov, Sergei, Faiyaz Notta, Kolja Eppert, Linh T. Nguyen, Pamela S. Ohashi, and John E. Dick. 2010. "Revised Map of the Human Progenitor Hierarchy Shows the Origin of Macrophages and Dendritic Cells in Early Lymphoid Development." *Nature Immunology* 11 (7): 585–93. https://doi.org/10.1038/ni.1889.

Duhan, Vikas, Thamer A. Hamdan, Haifeng C. Xu, Prashant Shinde, Hilal Bhat, Fanghui Li, Yahya Al-Matary, et al. 2019. "NK Cell–Intrinsic

FcɛRIγ Limits CD8+ T cell Expansion and Thereby Turns an Acute into a Chronic Viral Infection." *PLOS Pathogens* 15 (6): e1007797. https://doi.org/10.1371/journal.ppat.1007797.

Duò, Angelo, Mark D. Robinson, and Charlotte Soneson. 2018. "A Systematic Performance Evaluation of Clustering Methods for Single-Cell RNA-Seq Data." *F1000Research* 7 (September). https://doi.org/10.12688/f1000research.15666.2.

Efremova, Mirjana, Miquel Vento-Tormo, Sarah A. Teichmann, and Roser Vento-Tormo. 2020. "CellPhoneDB: Inferring Cell–Cell Communication from Combined Expression of Multi-Subunit Ligand–Receptor Complexes." *Nature Protocols* 15 (4): 1484–1506. https://doi.org/10.1038/s41596-020-0292-x.

Epstein, Paul M. 2017. "Different Phosphodiesterases (PDEs) Regulate Distinct Phosphoproteomes during CAMP Signaling." *Proceedings of the National Academy of Sciences* 114 (30): 7741–43. https://doi.org/10.1073/pnas.1709073114.

Fan, Jean, Hae-Ock Lee, Soohyun Lee, Da-eun Ryu, Semin Lee, Catherine Xue, Seok Jin Kim, et al. 2018. "Linking Transcriptional and Genetic Tumor Heterogeneity through Allele Analysis of Single-Cell RNA-Seq Data." *Genome Research*, June, gr.228080.117. https://doi.org/10.1101/gr.228080.117.

Fang, Fang, Mangju Wang, Quan Qiu, Dingfang Bu, Wei Liu, Ying Zhang, Yongjin Shi, Xuzhen Lv, and Ping Zhu. 2015. "Human Transcription Factor KLF3 Maintains T Lymphocyte Quiescent Phenotype Via Inhibiting SHP-1 Expression." *Blood* 126 (23): 3426–3426. https://doi.org/10.1182/blood.V126.23.3426.3426.

Finak, Greg, Andrew McDavid, Masanao Yajima, Jingyuan Deng, Vivian Gersuk, Alex K. Shalek, Chloe K. Slichter, et al. 2015. "MAST: A Flexible Statistical Framework for Assessing Transcriptional Changes and Characterizing Heterogeneity in Single-Cell RNA Sequencing Data." *Genome Biology* 16 (1): 278. https://doi.org/10.1186/s13059-015-0844-5.

F.R.S, Karl Pearson. 1901. "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 (11): 559–72. https://doi.org/10.1080/14786440109462720.

Fu, Chunmei, and Aimin Jiang. 2018. "Dendritic Cells and CD8 T Cell Immunity in Tumor Microenvironment." *Frontiers in Immunology* 9 (December). https://doi.org/10.3389/fimmu.2018.03059.

Gajewski, Thomas F., Hans Schreiber, and Yang-Xin Fu. 2013. "Innate and Adaptive Immune Cells in the Tumor Microenvironment." *Nature Immunology* 14 (10): 1014–22. https://doi.org/10.1038/ni.2703.

Galletti, Giovanni, Gabriele De Simone, Emilia M. C. Mazza, Simone Puccio, Claudia Mezzanotte, Timothy M. Bi, Alexey N. Davydov, et al. 2020a. "Two Subsets of Stem-like CD8+ Memory T Cell Progenitors with Distinct Fate Commitments in Humans." *Nature Immunology*, October. https://doi.org/10.1038/s41590-020-0791-5.

———. 2020b. "Two Subsets of Stem-like CD8 + Memory T Cell Progenitors with Distinct Fate Commitments in Humans." *Nature Immunology*, October, 1–11. https://doi.org/10.1038/s41590-020-0791-5.

Galli, Stephen J., Niels Borregaard, and Thomas A. Wynn. 2011. "Phenotypic and Functional Plasticity of Cells of Innate Immunity: Macrophages, Mast Cells and Neutrophils." *Nature Immunology* 12 (11): 1035–44. https://doi.org/10.1038/ni.2109.

Gerlach, Carmen, Jeroen W.J. van Heijst, Erwin Swart, Daoud Sie, Nicola Armstrong, Ron M. Kerkhoven, Dietmar Zehn, Michael J. Bevan, Koen Schepers, and Ton N.M. Schumacher. 2010. "One Naive T Cell, Multiple Fates in CD8+ T Cell Differentiation." *The Journal of Experimental Medicine* 207 (6): 1235–46. https://doi.org/10.1084/jem.20091175.

Germain, Ronald N. 2019. "The Cellular Determinants of Adaptive Immunity." *New England Journal of Medicine* 381 (11): 1083–85. https://doi.org/10.1056/NEJMcibr1909387.

Ghannam, Soufiane, Carine Bouffi, Farida Djouad, Christian Jorgensen, and Danièle Noël. 2010. "Immunosuppression by Mesenchymal Stem Cells: Mechanisms and Clinical Applications." *Stem Cell Research & Therapy* 1 (1): 2. https://doi.org/10.1186/scrt2.

Ghobrial, Irene M., Alexandre Detappe, Kenneth C. Anderson, and David P. Steensma. 2018. "The Bone-Marrow Niche in MDS and MGUS: Implications for AML and MM." *Nature Reviews. Clinical Oncology* 15 (4): 219–33. https://doi.org/10.1038/nrclinonc.2017.197.

Giladi, Amir, Franziska Paul, Yoni Herzog, Yaniv Lubling, Assaf Weiner, Ido Yofe, Diego Jaitin, et al. 2018. "Single-Cell Characterization of Haematopoietic Progenitors and Their Trajectories in Homeostasis and Perturbed Haematopoiesis." *Nature Cell Biology* 20 (7): 836–46. https://doi.org/10.1038/s41556-018-0121-4.

Goodnow, Christopher C. 2007. "Multistep Pathogenesis of Autoimmune Disease." *Cell* 130 (1): 25–35. https://doi.org/10.1016/j.cell.2007.06.033.

Görgens, André, Stefan Radtke, Michael Möllmann, Michael Cross, Jan Dürig, Peter A. Horn, and Bernd Giebel. 2013. "Revision of the Human Hematopoietic Tree: Granulocyte Subtypes Derive from Distinct Hematopoietic Lineages." *Cell Reports* 3 (5): 1539–52. https://doi.org/10.1016/j.celrep.2013.04.025.

Granja, Jeffrey M., Sandy Klemm, Lisa M. McGinnis, Arwa S. Kathiria, Anja Mezger, M. Ryan Corces, Benjamin Parks, et al. 2019. "Single-Cell Multiomic Analysis Identifies Regulatory Programs in Mixed-Phenotype Acute Leukemia." *Nature Biotechnology* 37 (12): 1458–65. https://doi.org/10.1038/s41587-019-0332-7.

Groom, Joanna R., and Andrew D. Luster. 2011. "CXCR3 in T Cell Function." *Experimental Cell Research* 317 (5): 620–31. https://doi.org/10.1016/j.yexcr.2010.12.017.

Günther, Patrick, and Joachim L. Schultze. 2019. "Mind the Map: Technology Shapes the Myeloid Cell Space." *Frontiers in Immunology* 10. https://doi.org/10.3389/fimmu.2019.02287.

Hafemeister, Christoph, and Rahul Satija. 2019. "Normalization and Variance Stabilization of Single-Cell RNA-Seq Data Using Regularized Negative Binomial Regression." *Genome Biology* 20 (1): 296. https://doi.org/10.1186/s13059-019-1874-1.

Haghverdi, Laleh, Aaron T. L. Lun, Michael D. Morgan, and John C. Marioni. 2018. "Batch Effects in Single-Cell RNA-Sequencing Data Are Corrected by Matching Mutual Nearest Neighbors." *Nature Biotechnology* 36 (5): 421–27. https://doi.org/10.1038/nbt.4091.

Hajdu, Steven I. 2011. "A Note from History: Landmarks in History of Cancer, Part 1." *Cancer* 117 (5): 1097–1102. https://doi.org/10.1002/cncr.25553.

Hare, Joshua. 2019. "Dealing with Sparse Rewards in Reinforcement Learning." *ArXiv:1910.09281 [Cs, Stat]*, November. http://arxiv.org/abs/1910.09281.

Hie, Brian, Hyunghoon Cho, Benjamin DeMeo, Bryan Bryson, and Bonnie Berger. 2019. "Geometric Sketching Compactly Summarizes the Single-Cell Transcriptomic Landscape." *Cell Systems* 8 (6): 483-493.e7. https://doi.org/10.1016/j.cels.2019.05.003.

Hie, Brian, Joshua Peters, Sarah K. Nyquist, Alex K. Shalek, Bonnie Berger, and Bryan D. Bryson. 2020. "Computational Methods for Single-Cell RNA Sequencing." *Annual Review of Biomedical Data Science* 3 (1): 339–64. https://doi.org/10.1146/annurev-biodatasci-012220-100601.

Hooke, Robert. 1667. *Micrographia: Or, Some Physiological Descriptions of Minute Bodies Made by Magnifying Glasses. With Observations and Inquiries Thereupon*. John Martyn.

Hu, Joyce K., Takashi Kagari, Jonathan M. Clingan, and Mehrdad Matloubian. 2011. "Expression of Chemokine Receptor CXCR3 on T Cells Affects the Balance between Effector and Memory CD8 T cell Generation." *Proceedings of the National Academy of Sciences* 108 (21): E118. https://doi.org/10.1073/pnas.1101881108.

Inwood, Stephen. 2002. *The Man Who Knew Too Much: The Strange and Inventive Life of Robert Hooke, 1635-1703*. Macmillan.

Jacomy, Mathieu, Tommaso Venturini, Sebastien Heymann, and Mathieu Bastian. 2014. "ForceAtlas2, a Continuous Graph Layout Algorithm for Handy Network Visualization Designed for the Gephi Software." *PLOS ONE* 9 (6): e98679. https://doi.org/10.1371/journal.pone.0098679.

Kaech, Susan M., and Weiguo Cui. 2012a. "Transcriptional Control of Effector and Memory CD8 + T Cell Differentiation." *Nature Reviews Immunology* 12 (11): 749–61. https://doi.org/10.1038/nri3307.

———. 2012b. "Transcriptional Control of Effector and Memory CD8 + T Cell Differentiation." *Nature Reviews Immunology* 12 (11): 749–61. https://doi.org/10.1038/nri3307.

Kalman, Dan. 1996. "A Singularly Valuable Decomposition: The SVD of a Matrix." *College Math Journal* 27: 2–23.

Kaplan, Rosandra N., Rebecca D. Riba, Stergios Zacharoulis, Anna H. Bramley, Loïc Vincent, Carla Costa, Daniel D. MacDonald, et al. 2005. "VEGFR1-Positive Haematopoietic Bone Marrow Progenitors Initiate the Pre-Metastatic Niche." *Nature* 438 (7069): 820–27. https://doi.org/10.1038/nature04186.

Kaplanov, Irena, Yaron Carmi, Rachel Kornetsky, Avishai Shemesh, Galina V. Shurin, Michael R. Shurin, Charles A. Dinarello, Elena Voronov, and Ron N. Apte. 2019. "Blocking IL-1β Reverses the Immunosuppression in Mouse Breast Cancer and Synergizes with Anti-PD-1 for Tumor Abrogation." *Proceedings of the National Academy of Sciences of the United States of America* 116 (4): 1361–69. https://doi.org/10.1073/pnas.1812266115.

Kernfeld, Eric M., Ryan M. J. Genga, Kashfia Neherin, Margaret E. Magaletta, Ping Xu, and René Maehr. 2018. "A Single-Cell Transcriptomic Atlas of Thymus Organogenesis Resolves Cell Types and Developmental Maturation." *Immunity* 48 (6): 1258-1270.e6. https://doi.org/10.1016/j.immuni.2018.04.015.

Kharchenko, Peter V., Lev Silberstein, and David T. Scadden. 2014. "Bayesian Approach to Single-Cell Differential Expression Analysis." *Nature Methods* 11 (7): 740–42. https://doi.org/10.1038/nmeth.2967.

Kobourov, Stephen G. 2012. "Spring Embedders and Force Directed Graph Drawing Algorithms." *ArXiv:1201.3011 [Cs]*, January. http://arxiv.org/abs/1201.3011.

Koch, Ute, and Freddy Radtke. 2011. "Mechanisms of T Cell Development and Transformation." *Annual Review of Cell and Developmental Biology* 27: 539–62. https://doi.org/10.1146/annurev-cellbio-092910-154008.

Korotkevich, Gennady, Vladimir Sukhov, and Alexey Sergushichev. 2019. "Fast Gene Set Enrichment Analysis." *BioRxiv*, October, 060012. https://doi.org/10.1101/060012.

Kostopoulos, Ioannis V., Ioannis Ntanasis-Stathopoulos, Maria Gavriatopoulou, Ourania E. Tsitsilonis, and Evangelos Terpos. 2020. "Minimal Residual Disease in Multiple Myeloma: Current Landscape and Future Applications With Immunotherapeutic Approaches." *Frontiers in Oncology* 10. https://doi.org/10.3389/fonc.2020.00860.

Kumar, Shaji K., Vincent Rajkumar, Robert A. Kyle, Mark van Duin, Pieter Sonneveld, María-Victoria Mateos, Francesca Gay, and Kenneth C. Anderson. 2017. "Multiple Myeloma." *Nature Reviews Disease Primers* 3 (1): 1–20. https://doi.org/10.1038/nrdp.2017.46.

Kundaje, Anshul, Wouter Meuleman, Jason Ernst, Misha Bilenky, Angela Yen, Alireza Heravi-Moussavi, Pouya Kheradpour, et al. 2015. "Integrative Analysis of 111 Reference Human Epigenomes." *Nature* 518 (7539): 317–30. https://doi.org/10.1038/nature14248.

La Manno, Gioele, Ruslan Soldatov, Amit Zeisel, Emelie Braun, Hannah Hochgerner, Viktor Petukhov, Katja Lidschreiber, et al. 2018. "RNA Velocity of Single Cells." *Nature* 560 (7719): 494–98. https://doi.org/10.1038/s41586-018-0414-6.

Lähnemann, David, Johannes Köster, Ewa Szczurek, Davis J. McCarthy, Stephanie C. Hicks, Mark D. Robinson, Catalina A. Vallejos, et al. 2020. "Eleven Grand Challenges in Single-Cell Data Science." *Genome Biology* 21 (1): 31. https://doi.org/10.1186/s13059-020-1926-6.

Laurenti, Elisa, and Berthold Göttgens. 2018. "From Haematopoietic Stem Cells to Complex Differentiation Landscapes." *Nature* 553 (7689): 418–26. https://doi.org/10.1038/nature25022.

Lawrence, Michael S., Petar Stojanov, Paz Polak, Gregory V. Kryukov, Kristian Cibulskis, Andrey Sivachenko, Scott L. Carter, et al. 2013. "Mutational Heterogeneity in Cancer and the Search for New Cancer-Associated Genes." *Nature* 499 (7457): 214–18. https://doi.org/10.1038/nature12213.

Ledergor, Guy, Assaf Weiner, Mor Zada, Shuang-Yin Wang, Yael C. Cohen, Moshe E. Gatt, Nimrod Snir, et al. 2018. "Single Cell Dissection of Plasma Cell Heterogeneity in Symptomatic and Asymptomatic Myeloma." *Nature Medicine* 24 (12): 1867–76. https://doi.org/10.1038/s41591-018-0269-2.

Leone, Patrizia, Simona Berardi, Maria Antonia Frassanito, Roberto Ria, Valli De Re, Sebastiano Cicco, Stefano Battaglia, et al. 2015. "Dendritic Cells Accumulate in the Bone Marrow of Myeloma Patients Where

They Protect Tumor Plasma Cells from CD8+ T cell Killing." *Blood* 126 (12): 1443–51. https://doi.org/10.1182/blood-2015-01-623975.

Leun, Anne M. van der, Daniela S. Thommen, and Ton N. Schumacher. 2020. "CD8 + T Cell States in Human Cancer: Insights from Single-Cell Analysis." *Nature Reviews Cancer* 20 (4): 218–32. https://doi.org/10.1038/s41568-019-0235-4.

Li, Chaofan, Bibo Zhu, Young Min Son, Zheng Wang, Li Jiang, Min Xiang, Zhenqing Ye, et al. 2019. "The Transcription Factor Bhlhe40 Programs Mitochondrial Regulation of Resident CD8+ T Cell Fitness and Functionality." *Immunity* 51 (3): 491-507.e7. https://doi.org/10.1016/j.immuni.2019.08.013.

Li, Hongfei, Catherine A. Calder, and Noel Cressie. 2007. "Beyond Moran's I: Testing for Spatial Dependence Based on the Spatial Autoregressive Model." *Geographical Analysis* 39 (4): 357–75. https://doi.org/10.1111/j.1538-4632.2007.00708.x.

Liberzon, Arthur, Chet Birger, Helga Thorvaldsdóttir, Mahmoud Ghandi, Jill P. Mesirov, and Pablo Tamayo. 2015. "The Molecular Signatures Database (MSigDB) Hallmark Gene Set Collection." *Cell Systems* 1 (6): 417–25. https://doi.org/10.1016/j.cels.2015.12.004.

Linker, Stephanie M., Lara Urban, Stephen J. Clark, Mariya Chhatriwala, Shradha Amatya, Davis J. McCarthy, Ingo Ebersberger, et al. 2019. "Combined Single-Cell Profiling of Expression and DNA Methylation Reveals Splicing Regulation and Heterogeneity." *Genome Biology* 20 (1): 30. https://doi.org/10.1186/s13059-019-1644-0.

Liu, Ting, Lingyun Zhang, Donghyun Joo, and Shao-Cong Sun. 2017. "NF-ΚB Signaling in Inflammation." *Signal Transduction and Targeted Therapy* 2 (July): 17023. https://doi.org/10.1038/sigtrans.2017.23.

Liu, Xindong, Yun Wang, Huiping Lu, Jing Li, Xiaowei Yan, Minglu Xiao, Jing Hao, et al. 2019. "Genome-Wide Analysis Identifies NR4A1 as a Key Mediator of T Cell Dysfunction." *Nature* 567 (7749): 525–29. https://doi.org/10.1038/s41586-019-0979-8.

Lopez, Romain, Jeffrey Regier, Michael B. Cole, Michael I. Jordan, and Nir Yosef. 2018. "Deep Generative Modeling for Single-Cell Transcriptomics." *Nature Methods* 15 (12): 1053–58. https://doi.org/10.1038/s41592-018-0229-2.

Love, Michael I., Wolfgang Huber, and Simon Anders. 2014. "Moderated Estimation of Fold Change and Dispersion for RNA-Seq Data with DESeq2." *Genome Biology* 15 (12): 550. https://doi.org/10.1186/s13059-014-0550-8.

Luecken, M. D., M. Büttner, K. Chaichoompu, A. Danese, M. Interlandi, M. F. Mueller, D. C. Strobl, et al. 2020. "Benchmarking Atlas-Level Data

Integration in Single-Cell Genomics." *BioRxiv*, May, 2020.05.22.111161. https://doi.org/10.1101/2020.05.22.111161.

Lun, Aaron T. L., Samantha Riesenfeld, Tallulah Andrews, The Phuong Dao, Tomas Gomes, John C. Marioni, and participants in the 1st Human Cell Atlas Jamboree. 2019. "EmptyDrops: Distinguishing Cells from Empty Droplets in Droplet-Based Single-Cell RNA Sequencing Data." *Genome Biology* 20 (1): 63. https://doi.org/10.1186/s13059-019-1662-y.

Lun, Aaron T.L., Davis J. McCarthy, and John C. Marioni. 2016. "A Step-by-Step Workflow for Low-Level Analysis of Single-Cell RNA-Seq Data with Bioconductor." *F1000Research* 5 (October): 2122. https://doi.org/10.12688/f1000research.9501.2.

Lundmark, Anna, Haleh Davanian, Tove Båge, Gunnar Johannsen, Catalin Koro, Joakim Lundeberg, and Tülay Yucel-Lindberg. 2015. "Transcriptome Analysis Reveals Mucin 4 to Be Highly Associated with Periodontitis and Identifies Pleckstrin as a Link to Systemic Diseases." *Scientific Reports* 5 (1): 1–13. https://doi.org/10.1038/srep18475.

Lust, J. A., and K. A. Donovan. 1999. "The Role of Interleukin-1 Beta in the Pathogenesis of Multiple Myeloma." *Hematology/Oncology Clinics of North America* 13 (6): 1117–25. https://doi.org/10.1016/s0889-8588(05)70115-5.

Maaten, Laurens van der, and Geoffrey Hinton. 2008. "Visualizing Data Using T-SNE." *Journal of Machine Learning Research* 9 (Nov): 2579–2605.

Macosko, Evan Z., Anindita Basu, Rahul Satija, James Nemesh, Karthik Shekhar, Melissa Goldman, Itay Tirosh, et al. 2015. "Highly Parallel Genome-Wide Expression Profiling of Individual Cells Using Nanoliter Droplets." *Cell* 161 (5): 1202–14. https://doi.org/10.1016/j.cell.2015.05.002.

Martin, Matthew D., and Vladimir P. Badovinac. 2018. "Defining Memory CD8 T Cell." *Frontiers in Immunology* 9. https://doi.org/10.3389/fimmu.2018.02692.

Mazzarello, Paolo. 1999. "A Unifying Concept: The History of Cell Theory." *Nature Cell Biology* 1 (1): E13–15. https://doi.org/10.1038/8964.

McInnes, Leland, John Healy, and James Melville. 2020. "UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction." *ArXiv:1802.03426 [Cs, Stat]*, September. http://arxiv.org/abs/1802.03426.

Medyouf, Hind. 2017. "The Microenvironment in Human Myeloid Malignancies: Emerging Concepts and Therapeutic Implications." *Blood* 129 (12): 1617–26. https://doi.org/10.1182/blood-2016-11-696070.

"Micrographia by Robert Hooke, 1665." n.d. The British Library. The British Library. Accessed November 8, 2020. https://www.bl.uk/collection-items/micrographia-by-robert-hooke-1665.

Miller, J. F. 1961. "Immunological Function of the Thymus." *Lancet (London, England)* 2 (7205): 748–49. https://doi.org/10.1016/s0140-6736(61)90693-6.

Mognol, Giuliana P., Roberto Spreafico, Victor Wong, James P. Scott-Browne, Susan Togher, Alexander Hoffmann, Patrick G. Hogan, Anjana Rao, and Sara Trifari. 2017. "Exhaustion-Associated Regulatory Regions in CD8+ Tumor-Infiltrating T Cells." *Proceedings of the National Academy of Sciences of the United States of America* 114 (13): E2776–85. https://doi.org/10.1073/pnas.1620498114.

Moretta, A., E. Marcenaro, S. Parolini, G. Ferlazzo, and L. Moretta. 2008. "NK Cells at the Interface between Innate and Adaptive Immunity." *Cell Death & Differentiation* 15 (2): 226–33. https://doi.org/10.1038/sj.cdd.4402170.

Morgan, Gareth J., Brian A. Walker, and Faith E. Davies. 2012. "The Genetic Architecture of Multiple Myeloma." *Nature Reviews. Cancer* 12 (5): 335–48. https://doi.org/10.1038/nrc3257.

Mosser, David M., and Justin P. Edwards. 2008. "Exploring the Full Spectrum of Macrophage Activation." *Nature Reviews. Immunology* 8 (12): 958–69. https://doi.org/10.1038/nri2448.

Mu, Shidai, Lisha Ai, Fengjuan Fan, Chunyan Sun, and Yu Hu. 2018. "Prognostic Role of Neutrophil–Lymphocyte Ratio in Multiple Myeloma: A Dose–Response Meta-Analysis." *OncoTargets and Therapy* 11 (January): 499–507. https://doi.org/10.2147/OTT.S153146.

Natürliche Schöpfungsgeschichte : Gemeinverständliche Wissenschaftliche Vorträge Über Die Entwickelungslehre Im Allgemeinen Und Diejenige von Darwin, Goethe Und Lamarck Im Besonderen, Über Die Anwendung Derselben Auf Den Ursprung Des Menschen Und Andere Damit Zusammenhängende Grundfragen Der Naturwissenschaft / von Ernst Haeckel. 1868. http://www.e-rara.ch/zut/7388510.

Nicolet, Benoît P., Aurélie Guislain, Floris P. J. van Alphen, Raquel Gomez-Eerland, Ton N. M. Schumacher, Maartje van den Biggelaar, and Monika C. Wolkers. 2020. "CD29 Identifies IFN-γ–Producing Human CD8+ T Cells with an Increased Cytotoxic Potential." *Proceedings of the National Academy of Sciences* 117 (12): 6686–96. https://doi.org/10.1073/pnas.1913940117.

Ntranos, Vasilis, Lynn Yi, Páll Melsted, and Lior Pachter. 2019. "A Discriminative Learning Approach to Differential Expression Analysis

for Single-Cell RNA-Seq." *Nature Methods* 16 (2): 163–66. https://doi.org/10.1038/s41592-018-0303-9.

Oliveira, Sofia de, Emily E. Rosowski, and Anna Huttenlocher. 2016. "Neutrophil Migration in Infection and Wound Repair: Going Forward in Reverse." *Nature Reviews Immunology* 16 (6): 378–91. https://doi.org/10.1038/nri.2016.49.

Onec, Birgul, Harika Okutan, Murat Albayrak, Esra Saribacak Can, Vedat Aslan, Basak Unver Koluman, Ozge Soyer Kosemehmetoglu, Aynur Albayrak, and Durdu Mehmet Kos. 2017. "The Predictive Role of the Neutrophil/Lymphocyte Ratio in Survival with Multiple Myeloma: A Single Center Experience." *Journal of Clinical Laboratory Analysis* 31 (2): e22032. https://doi.org/10.1002/jcla.22032.

Ouyang, Wenjun, and Anne O'Garra. 2019. "IL-10 Family Cytokines IL-10 and IL-22: From Basic Science to Clinical Translation." *Immunity* 50 (4): 871–91. https://doi.org/10.1016/j.immuni.2019.03.020.

Palumbo, Antonio, Asher Chanan-Khan, Katja Weisel, Ajay K. Nooka, Tamas Masszi, Meral Beksac, Ivan Spicka, et al. 2016. "Daratumumab, Bortezomib, and Dexamethasone for Multiple Myeloma." *New England Journal of Medicine* 375 (8): 754–66. https://doi.org/10.1056/NEJMoa1606038.

Paul, Franziska, Ya'ara Arkin, Amir Giladi, Diego Adhemar Jaitin, Ephraim Kenigsberg, Hadas Keren-Shaul, Deborah Winter, et al. 2015. "Transcriptional Heterogeneity and Lineage Commitment in Myeloid Progenitors." *Cell* 163 (7): 1663–77. https://doi.org/10.1016/j.cell.2015.11.013.

Pawlyn, Charlotte, and Gareth J. Morgan. 2017. "Evolutionary Biology of High-Risk Multiple Myeloma." *Nature Reviews. Cancer* 17 (9): 543–56. https://doi.org/10.1038/nrc.2017.63.

Peperzak, Victor, Elise A. M. Veraar, Anna M. Keller, Yanling Xiao, and Jannie Borst. 2010. "The Pim Kinase Pathway Contributes to Survival Signaling in Primed CD8+ T Cells upon CD27 Costimulation." *The Journal of Immunology* 185 (11): 6670–78. https://doi.org/10.4049/jimmunol.1000159.

Phan, Anthony T., and Ananda W. Goldrath. 2015. "Hypoxia-Inducible Factors Regulate T Cell Metabolism and Function." *Molecular Immunology* 68 (2, Part C): 527–35. https://doi.org/10.1016/j.molimm.2015.08.004.

Pierson, Emma, and Christopher Yau. 2015. "ZIFA: Dimensionality Reduction for Zero-Inflated Single-Cell Gene Expression Analysis." *Genome Biology* 16 (1): 241. https://doi.org/10.1186/s13059-015-0805-z.

Pittari, Gianfranco, Luca Vago, Moreno Festuccia, Chiara Bonini, Deena Mudawi, Luisa Giaccone, and Benedetto Bruno. 2017. "Restoring

Natural Killer Cell Immunity against Multiple Myeloma in the Era of New Drugs." *Frontiers in Immunology* 8 (November). https://doi.org/10.3389/fimmu.2017.01444.

Pump, Wiebke C., Thomas Kraemer, Trevor Huyton, Gia-Gia T. Hò, Rainer Blasczyk, and Christina Bade-Doeding. 2019. "Between Innate and Adaptive Immune Responses: NKG2A, NKG2C, and CD8+ T Cell Recognition of HLA-E Restricted Self-Peptides Acquired in the Absence of HLA-Ia." *International Journal of Molecular Sciences* 20 (6). https://doi.org/10.3390/ijms20061454.

Purnick, Priscilla E. M., and Ron Weiss. 2009. "The Second Wave of Synthetic Biology: From Modules to Systems." *Nature Reviews Molecular Cell Biology* 10 (6): 410–22. https://doi.org/10.1038/nrm2698.

Qiu, Xiaojie, Andrew Hill, Jonathan Packer, Dejun Lin, Yi-An Ma, and Cole Trapnell. 2017. "Single-Cell MRNA Quantification and Differential Analysis with Census." *Nature Methods* 14 (3): 309–15. https://doi.org/10.1038/nmeth.4150.

Qiu, Xiaojie, Qi Mao, Ying Tang, Li Wang, Raghav Chawla, Hannah A. Pliner, and Cole Trapnell. 2017. "Reversed Graph Embedding Resolves Complex Single-Cell Trajectories." *Nature Methods*. https://doi.org/10.1038/nmeth.4402.

Quach, H., D. Ritchie, A. K. Stewart, P. Neeson, S. Harrison, M. J. Smyth, and H. M. Prince. 2010. "Mechanism of Action of Immunomodulatory Drugs (IMiDS) in Multiple Myeloma." *Leukemia* 24 (1): 22–32. https://doi.org/10.1038/leu.2009.236.

Rajkumar, S. Vincent. 2020. "Multiple Myeloma: 2020 Update on Diagnosis, Risk-Stratification and Management." *American Journal of Hematology* 95 (5): 548–67. https://doi.org/10.1002/ajh.25791.

Rajkumar, S. Vincent, Meletios A. Dimopoulos, Antonio Palumbo, Joan Blade, Giampaolo Merlini, María-Victoria Mateos, Shaji Kumar, et al. 2014. "International Myeloma Working Group Updated Criteria for the Diagnosis of Multiple Myeloma." *The Lancet. Oncology* 15 (12): e538-548. https://doi.org/10.1016/S1470-2045(14)70442-5.

Rajkumar, S. Vincent, and Shaji Kumar. 2016. "Multiple Myeloma: Diagnosis and Treatment." *Mayo Clinic Proceedings* 91 (1): 101–19. https://doi.org/10.1016/j.mayocp.2015.11.007.

Rajkumar, S. Vincent, Ruben A. Mesa, Rafael Fonseca, Georgene Schroeder, Matthew F. Plevak, Angela Dispenzieri, Martha Q. Lacy, et al. 2002. "Bone Marrow Angiogenesis in 400 Patients with Monoclonal Gammopathy of Undetermined Significance, Multiple Myeloma, and Primary Amyloidosis." *Clinical Cancer Research: An Official Journal of the American Association for Cancer Research* 8 (7): 2210–16.

Research, Center for Drug Evaluation and. 2019. "FDA Alerts Healthcare Professionals and Oncology Clinical Investigators about Two Clinical Trials on Hold Evaluating KEYTRUDA® (Pembrolizumab) in Patients with Multiple Myeloma." *FDA*, February. https://www.fda.gov/drugs/drug-safety-and-availability/fda-alerts-healthcare-professionals-and-oncology-clinical-investigators-about-two-clinical-trials.

Rodriques, Samuel G., Robert R. Stickels, Aleksandrina Goeva, Carly A. Martin, Evan Murray, Charles R. Vanderburg, Joshua Welch, Linlin M. Chen, Fei Chen, and Evan Z. Macosko. 2019. "Slide-Seq: A Scalable Technology for Measuring Genome-Wide Expression at High Spatial Resolution." *Science* 363 (6434): 1463–67. https://doi.org/10.1126/science.aaw1219.

Rölle, Alexander, Marten Meyer, Silvia Calderazzo, Dirk Jäger, and Frank Momburg. 2018. "Distinct HLA-E Peptide Complexes Modify Antibody-Driven Effector Functions of Adaptive NK Cells." *Cell Reports* 24 (8): 1967-1976.e4. https://doi.org/10.1016/j.celrep.2018.07.069.

Romano, A., N. L. Parrinello, V. Simeon, F. Puglisi, P. La Cava, C. Bellofiore, C. Giallongo, et al. 2020. "High-Density Neutrophils in MGUS and Multiple Myeloma Are Dysfunctional and Immune-Suppressive Due to Increased STAT3 Downstream Signaling." *Scientific Reports* 10 (1): 1–18. https://doi.org/10.1038/s41598-020-58859-x.

Rosales, Carlos. 2018. "Neutrophil: A Cell with Many Roles in Inflammation or Several Cell Types?" *Frontiers in Physiology* 9 (February). https://doi.org/10.3389/fphys.2018.00113.

Rosean, Timothy R., Van S. Tompkins, Guido Tricot, Carol J. Holman, Alicia K. Olivier, Fenghuang Zhan, and Siegfried Janz. 2014. "Preclinical Validation of Interleukin 6 as a Therapeutic Target in Multiple Myeloma." *Immunologic Research* 59 (0): 188–202. https://doi.org/10.1007/s12026-014-8528-x.

Roychoudhuri, Rahul, David Clever, Peng Li, Yoshiyuki Wakabayashi, Kylie M. Quinn, Christopher A. Klebanoff, Yun Ji, et al. 2016. "BACH2 Regulates CD8[+] T Cell Differentiation by Controlling Access of AP-1 Factors to Enhancers." *Nature Immunology* 17 (7): 851–60. https://doi.org/10.1038/ni.3441.

Sabatino, Joseph J., Anne-Katrin Pröbstel, and Scott S. Zamvil. 2019. "B Cells in Autoimmune and Neurodegenerative Central Nervous System Diseases." *Nature Reviews Neuroscience* 20 (12): 728–45. https://doi.org/10.1038/s41583-019-0233-2.

Saelens, Wouter, Robrecht Cannoodt, Helena Todorov, and Yvan Saeys. 2019. "A Comparison of Single-Cell Trajectory Inference Methods."

Nature Biotechnology 37 (5): 547–54. https://doi.org/10.1038/s41587-019-0071-9.

Satpathy, Ansuman T., Jeffrey M. Granja, Kathryn E. Yost, Yanyan Qi, Francesca Meschi, Geoffrey P. McDermott, Brett N. Olsen, et al. 2019. "Massively Parallel Single-Cell Chromatin Landscapes of Human Immune Cell Development and Intratumoral T Cell Exhaustion." *Nature Biotechnology* 37 (8): 925–36. https://doi.org/10.1038/s41587-019-0206-z.

Schölkopf, Bernhard. 2019. "Causality for Machine Learning." *ArXiv:1911.10500 [Cs, Stat]*, December. http://arxiv.org/abs/1911.10500.

Scott, Charlotte L., and Kyla D. Omilusik. 2019. "ZEBs: Novel Players in Immune Cell Development and Function." *Trends in Immunology* 40 (5): 431–46. https://doi.org/10.1016/j.it.2019.03.001.

Seo, Hyungseok, Joyce Chen, Edahí González-Avalos, Daniela Samaniego-Castruita, Arundhoti Das, Yueqiang H. Wang, Isaac F. López-Moyado, et al. 2019. "TOX and TOX2 Transcription Factors Cooperate with NR4A Transcription Factors to Impose CD8+ T Cell Exhaustion." *Proceedings of the National Academy of Sciences of the United States of America* 116 (25): 12410–15. https://doi.org/10.1073/pnas.1905675116.

Silver, David, Julian Schrittwieser, Karen Simonyan, Ioannis Antonoglou, Aja Huang, Arthur Guez, Thomas Hubert, et al. 2017. "Mastering the Game of Go without Human Knowledge." *Nature* 550 (7676): 354–59. https://doi.org/10.1038/nature24270.

Sinclair, Linda V., Julia Rolf, Elizabeth Emslie, Yun-Bo Shi, Peter M. Taylor, and Doreen A. Cantrell. 2013. "Control of Amino-Acid Transport by Antigen Receptors Coordinates the Metabolic Reprogramming Essential for T Cell Differentiation." *Nature Immunology* 14 (5): 500–508. https://doi.org/10.1038/ni.2556.

Soneson, Charlotte, and Mark D. Robinson. 2018. "Bias, Robustness and Scalability in Single-Cell Differential Expression Analysis." *Nature Methods* 15 (4): 255–61. https://doi.org/10.1038/nmeth.4612.

Spangrude, G. J., S. Heimfeld, and I. L. Weissman. 1988. "Purification and Characterization of Mouse Hematopoietic Stem Cells." *Science (New York, N.Y.)* 241 (4861): 58–62. https://doi.org/10.1126/science.2898810.

Spendlove, Ian, and Ruhcha Sutavani. 2010. "The Role of CD97 in Regulating Adaptive T cell Responses." In *Adhesion-GPCRs: Structure to Function*, edited by Simon Yona and Martin Stacey, 138–48. Advances in Experimental Medicine and Biology. Boston, MA: Springer US. https://doi.org/10.1007/978-1-4419-7913-1_12.

Stegle, Oliver, Sarah A. Teichmann, and John C. Marioni. 2015. "Computational and Analytical Challenges in Single-Cell Transcriptomics." *Nature Reviews Genetics* 16 (3): 133–45. https://doi.org/10.1038/nrg3833.

Stemberger, Christian, Katharina M. Huster, Martina Koffler, Florian Anderl, Matthias Schiemann, Hermann Wagner, and Dirk H. Busch. 2007. "A Single Naive CD8+ T Cell Precursor Can Develop into Diverse Effector and Memory Subsets." *Immunity* 27 (6): 985–97. https://doi.org/10.1016/j.immuni.2007.10.012.

Stoeckius, Marlon, Christoph Hafemeister, William Stephenson, Brian Houck-Loomis, Pratip K. Chattopadhyay, Harold Swerdlow, Rahul Satija, and Peter Smibert. 2017. "Simultaneous Epitope and Transcriptome Measurement in Single Cells." *Nature Methods* 14 (9): 865–68. https://doi.org/10.1038/nmeth.4380.

Stuart, Tim, Andrew Butler, Paul Hoffman, Christoph Hafemeister, Efthymia Papalexi, William M. Mauck, Yuhan Hao, Marlon Stoeckius, Peter Smibert, and Rahul Satija. 2019. "Comprehensive Integration of Single-Cell Data." *Cell* 177 (7): 1888-1902.e21. https://doi.org/10.1016/j.cell.2019.05.031.

Stuart, Tim, and Rahul Satija. 2019. "Integrative Single-Cell Analysis." *Nature Reviews Genetics* 20 (5): 257–72. https://doi.org/10.1038/s41576-019-0093-7.

Svensson, Valentine. 2020. "Droplet ScRNA-Seq Is Not Zero-Inflated." *Nature Biotechnology* 38 (2): 147–50. https://doi.org/10.1038/s41587-019-0379-5.

Svensson, Valentine, Roser Vento-Tormo, and Sarah A. Teichmann. 2018. "Exponential Scaling of Single-Cell RNA-Seq in the Past Decade." *Nature Protocols* 13 (4): 599–604. https://doi.org/10.1038/nprot.2017.149.

Tang, Fuchou, Catalin Barbacioru, Yangzhou Wang, Ellen Nordman, Clarence Lee, Nanlan Xu, Xiaohui Wang, et al. 2009. "MRNA-Seq Whole-Transcriptome Analysis of a Single Cell." *Nature Methods* 6 (5): 377–82. https://doi.org/10.1038/nmeth.1315.

Taube, Janis M., Jérôme Galon, Lynette M. Sholl, Scott J. Rodig, Tricia R. Cottrell, Nicolas A. Giraldo, Alexander S. Baras, et al. 2018. "Implications of the Tumor Immune Microenvironment for Staging and Therapeutics." *Modern Pathology* 31 (2): 214–34. https://doi.org/10.1038/modpathol.2017.156.

Thorsson, Vésteinn, David L. Gibbs, Scott D. Brown, Denise Wolf, Dante S. Bortone, Tai-Hsien Ou Yang, Eduard Porta-Pardo, et al. 2018. "The Immune Landscape of Cancer." *Immunity* 48 (4): 812-830.e14. https://doi.org/10.1016/j.immuni.2018.03.023.

Tirado-Gonzalez, I., E. Czlonka, A. Nevmerzhitskaya, D. Soetopo, E. Bergonzani, A. Mahmoud, A. Contreras, et al. 2018. "CRISPR/Cas9-Edited NSG Mice as PDX Models of Human Leukemia to Address the Role of Niche-Derived SPARC." *Leukemia* 32 (4): 1048–51. https://doi.org/10.1038/leu.2017.346.

Tirosh, Itay, Benjamin Izar, Sanjay M. Prakadan, Marc H. Wadsworth, Daniel Treacy, John J. Trombetta, Asaf Rotem, et al. 2016. "Dissecting the Multicellular Ecosystem of Metastatic Melanoma by Single-Cell RNA-Seq." *Science (New York, N.Y.)* 352 (6282): 189–96. https://doi.org/10.1126/science.aad0501.

Townes, F. William, Stephanie C. Hicks, Martin J. Aryee, and Rafael A. Irizarry. 2019. "Feature Selection and Dimension Reduction for Single-Cell RNA-Seq Based on a Multinomial Model." *Genome Biology* 20 (1): 295. https://doi.org/10.1186/s13059-019-1861-6.

Traag, V. A., L. Waltman, and N. J. van Eck. 2019. "From Louvain to Leiden: Guaranteeing Well-Connected Communities." *Scientific Reports* 9 (1): 5233. https://doi.org/10.1038/s41598-019-41695-z.

Vacca, A., and D. Ribatti. 2006. "Bone Marrow Angiogenesis in Multiple Myeloma." *Leukemia* 20 (2): 193–99. https://doi.org/10.1038/sj.leu.2404067.

Velten, Lars, Simon F. Haas, Simon Raffel, Sandra Blaszkiewicz, Saiful Islam, Bianca P. Hennig, Christoph Hirche, et al. 2017. "Human Haematopoietic Stem Cell Lineage Commitment Is a Continuous Process." *Nature Cell Biology* 19 (4): 271–81. https://doi.org/10.1038/ncb3493.

Vento-Tormo, Roser, Mirjana Efremova, Rachel A. Botting, Margherita Y. Turco, Miquel Vento-Tormo, Kerstin B. Meyer, Jong-Eun Park, et al. 2018. "Single-Cell Reconstruction of the Early Maternal–Fetal Interface in Humans." *Nature* 563 (7731): 347–53. https://doi.org/10.1038/s41586-018-0698-6.

Visekruna, Alexander, Anton Volkov, and Ulrich Steinhoff. 2012. "A Key Role for NF-ĸB Transcription Factor c-Rel in T-Lymphocyte-Differentiation and Effector Functions." *Clinical and Developmental Immunology* 2012. https://doi.org/10.1155/2012/239368.

Vivier, Eric, David H. Raulet, Alessandro Moretta, Michael A. Caligiuri, Laurence Zitvogel, Lewis L. Lanier, Wayne M. Yokoyama, and Sophie Ugolini. 2011. "Innate or Adaptive Immunity? The Example of Natural Killer Cells." *Science (New York, N.Y.)* 331 (6013): 44–49. https://doi.org/10.1126/science.1198687.

Vo, Manh-Cuong, Sung-Hoon Jung, Tan-Huy Chu, Hyun-Ju Lee, Thangaraj Jaya Lakshmi, Hye-Seong Park, Hyeoung-Joon Kim, Joon Haeng Rhee, and Je-Jung Lee. 2018. "Lenalidomide and Programmed Death-1

Blockade Synergistically Enhances the Effects of Dendritic Cell Vaccination in a Model of Murine Myeloma." *Frontiers in Immunology* 9. https://doi.org/10.3389/fimmu.2018.01370.

Wagner, Daniel E., and Allon M. Klein. 2020. "Lineage Tracing Meets Single-Cell Omics: Opportunities and Challenges." *Nature Reviews Genetics* 21 (7): 410–27. https://doi.org/10.1038/s41576-020-0223-2.

Wagner, Robert P. 1999. "Rudolph Virchow and the Genetic Basis of Somatic Ecology." *Genetics* 151 (3): 917–20.

Waizenegger, J. S., I. Ben-Batalla, N. Weinhold, T. Meissner, M. Wroblewski, M. Janning, K. Riecken, et al. 2015. "Role of Growth Arrest-Specific Gene 6-Mer Axis in Multiple Myeloma." *Leukemia* 29 (3): 696–704. https://doi.org/10.1038/leu.2014.236.

Waltman, Ludo, and Nees Jan van Eck. 2013. "A Smart Local Moving Algorithm for Large-Scale Modularity-Based Community Detection." *The European Physical Journal B* 86 (11): 471. https://doi.org/10.1140/epjb/e2013-40829-0.

Wang, Chien-Neng. 2018. "The Role of Myristoylated, Alanine-Rich C-Kinase Substrate (MARCKS) in a Murine Model of Chronic Airway Inflammation." *The Journal of Immunology* 200 (1 Supplement): 44.8-44.8.

Weinreb, Caleb, Alejo Rodriguez-Fraticelli, Fernando D. Camargo, and Allon M. Klein. 2020. "Lineage Tracing on Transcriptional Landscapes Links State to Fate during Differentiation." *Science*, January. https://doi.org/10.1126/science.aaw3381.

Weinreb, Caleb, Samuel Wolock, and Allon M. Klein. 2018. "SPRING: A Kinetic Interface for Visualizing High Dimensional Single-Cell Expression Data." *Bioinformatics (Oxford, England)* 34 (7): 1246–48. https://doi.org/10.1093/bioinformatics/btx792.

Wherry, E. John, and Makoto Kurachi. 2015. "Molecular and Cellular Insights into T Cell Exhaustion." *Nature Reviews. Immunology* 15 (8): 486–99. https://doi.org/10.1038/nri3862.

Wolf, F. Alexander, Fiona K. Hamey, Mireya Plass, Jordi Solana, Joakim S. Dahlin, Berthold Göttgens, Nikolaus Rajewsky, Lukas Simon, and Fabian J. Theis. 2019. "PAGA: Graph Abstraction Reconciles Clustering with Trajectory Inference through a Topology Preserving Map of Single Cells." *Genome Biology* 20 (1): 59. https://doi.org/10.1186/s13059-019-1663-x.

Xu, Chen, and Zhengchang Su. 2015. "Identification of Cell Types from Single-Cell Transcriptomes Using a Novel Clustering Method." *Bioinformatics* 31 (12): 1974–80. https://doi.org/10.1093/bioinformatics/btv088.

Yamauchi, Takayoshi, and Toshiro Moroishi. 2019. "Hippo Pathway in Mammalian Adaptive Immune System." *Cells* 8 (5): 398. https://doi.org/10.3390/cells8050398.

Yan, Siyang, Niels Vandewalle, Nathan De Beule, Sylvia Faict, Ken Maes, Elke De Bruyne, Eline Menu, Karin Vanderkerken, and Kim De Veirman. 2019. "AXL Receptor Tyrosine Kinase as a Therapeutic Target in Hematological Malignancies: Focus on Multiple Myeloma." *Cancers* 11 (11). https://doi.org/10.3390/cancers11111727.

Zaman, Mohammad Mahabub-Uz, Kazuya Masuda, Kishan Kumar Nyati, Praveen Kumar Dubey, Barry Ripley, Kai Wang, Jaya Prakash Chalise, Mitsuru Higa, Hamza Hanieh, and Tadamitsu Kishimoto. 2016. "Arid5a Exacerbates IFN-γ–Mediated Septic Shock by Stabilizing T-Bet MRNA." *Proceedings of the National Academy of Sciences* 113 (41): 11543–48. https://doi.org/10.1073/pnas.1613307113.

Zavidij, Oksana, Nicholas J. Haradhvala, Tarek H. Mouhieddine, Romanos Sklavenitis-Pistofidis, Songjie Cai, Mairead Reidy, Mahshid Rahmat, et al. 2020. "Single-Cell RNA Sequencing Reveals Compromised Immune Microenvironment in Precursor Stages of Multiple Myeloma." *Nature Cancer* 1 (5): 493–506. https://doi.org/10.1038/s43018-020-0053-3.

Zebley, Caitlin C., Stephen Gottschalk, and Ben Youngblood. 2020. "Rewriting History: Epigenetic Reprogramming of CD8+ T Cell Differentiation to Enhance Immunotherapy." *Trends in Immunology* 41 (8): 665–75. https://doi.org/10.1016/j.it.2020.06.008.

Zeng, Yang, Chen Liu, Yandong Gong, Zhijie Bai, Siyuan Hou, Jian He, Zhilei Bian, et al. 2019. "Single-Cell RNA Sequencing Resolves Spatiotemporal Development of Pre-Thymic Lymphoid Progenitors and Thymus Organogenesis in Human Embryos." *Immunity* 51 (5): 930-948.e6. https://doi.org/10.1016/j.immuni.2019.09.008.

Zhang, Qiming, and Zemin Zhang. 2020. "Stepwise Immune Alterations in Multiple Myeloma Progression." *Nature Cancer* 1 (5): 477–79. https://doi.org/10.1038/s43018-020-0063-1.

Zheng, Danping, Timur Liwinski, and Eran Elinav. 2020. "Interaction between Microbiota and Immunity in Health and Disease." *Cell Research* 30 (6): 492–506. https://doi.org/10.1038/s41422-020-0332-7.

Zheng, Grace X. Y., Jessica M. Terry, Phillip Belgrader, Paul Ryvkin, Zachary W. Bent, Ryan Wilson, Solongo B. Ziraldo, et al. 2017. "Massively Parallel Digital Transcriptional Profiling of Single Cells." *Nature Communications* 8 (1): 1–12. https://doi.org/10.1038/ncomms14049.

Zheng, Yuhuan, Qiang Wang, Tianshu Li, Jianfei Qian, Yong Lu, Yi Li, Enguang Bi, et al. 2016. "Role of Myeloma-Derived MIF in Myeloma Cell Adhesion to Bone Marrow and Chemotherapy Response." *JNCI*

Journal of the National Cancer Institute 108 (11). https://doi.org/10.1093/jnci/djw131.

Zhou, Wenqing, Lingyan Cao, Jacob Jeffries, Xiaoguang Zhu, Christopher J. Staiger, and Qing Deng. 2018. "Neutrophil-Specific Knockout Demonstrates a Role for Mitochondria in Regulating Neutrophil Motility in Zebrafish." *Disease Models & Mechanisms* 11 (3). https://doi.org/10.1242/dmm.033027.

Ziegler, Steven F., Fred Ramsdell, and Mark R. Alderson. 1994. "The Activation Antigen CD69." *STEM CELLS* 12 (5): 456–65. https://doi.org/10.1002/stem.5530120502.

pproa

8 Abbreviations

205

9 Talks, poster presentations, abstracts and publications

9.1.1 Conferences talks

Single Cell Atlas of Bone Marrow Microenvironment in Multiple Myeloma Long-Term Survivors

<u>Abdelrahman Mahmoud</u>, Raphael Lutz, Mohamed H.S. Awwad, Charles Imbusch, Tobias Boch, Niels Weinhold, Marc S. Raab, Carsten Müller- Tidow, Brian Durie, Simon Haas, Hartmut Goldschmidt, Benedikt Brors, and Michael Hundemer

(26th October 2019, Keynote speaker in EG-CompBio conference)

Single Cell Atlas of Bone Marrow Microenvironment in Multiple Myeloma Long-Term Survivors

<u>Abdelrahman Mahmoud</u>, Raphael Lutz, Mohamed H.S. Awwad, Charles Imbusch, Tobias Boch, Niels Weinhold, Marc S. Raab, Carsten Müller- Tidow, Brian Durie, Simon Haas, Hartmut Goldschmidt, Benedikt Brors, and Michael Hundemer

(15th September 2019, 6th Next-Generation Sequencing Symposium)

9.1.2 Video talks

Single Cell Project Summary: Multiple Myeloma Long-Term Survivors.

Abdelrahman Mahmoud,

(26th July 2020, Link: https://youtu.be/gC2pV_ezjiU)

Single Cell Data Analysis - Overview

Abdelrahman Mahmoud

(13th September 2018, Link: https://vimeo.com/289672930)

9.1.3 Poster presentations

Deciphering the Immune Evolution Landscape of Multiple Myeloma Long-Term Survivors Using Single Cell Genomics

Abdelrahman Mahmoud*, Raphael Lutz*, Mohamed H.S. Awwad, Charles Imbusch, Tobias Boch, Niels Weinhold, Marc S. Raab, Carsten Müller- Tidow, Brian Durie, Simon Haas, Hartmut Goldschmidt, Benedikt Brors, and Michael Hundemer

(16th November 2020, DKFZ 2020 PhD Poster presentations)

9.1.4 Abstracts

The Bone Marrow Microenvironment of Multiple Myeloma Long-Term Survivors at Single Cell Resolution

Raphael Lutz*, **Abdelrahman Mahmoud*,** Mohamed H.S. Awwad, Charles Imbusch, Tobias Boch, Niels Weinhold, Marc S. Raab, Carsten Müller- Tidow, Brian Durie, Simon Haas, Hartmut Goldschmidt, Benedikt Brors, and Michael Hundemer

Deconvolution of Hematopoietic Commitment Decisions By Genome-Wide Analysis of Progressive DNA Methylation Changes

Sina Staeble, MSc, Stephen Kraemer, MSc, Jens Langstein, MSc, Ruzhica Bogeska, PhD, Mark Hartmann, PhD, Maximilian Schoenung, MSc, Melinda Czeh, PhD, Julia Knoch, Natasha Anstee, PhD, Simon Haas, PhD, **Abdelrahman Mahmoud**, Julius Graesel, MSc, Daniel Huebschmann, MD PhD, Lars Feuerbach, PhD, Weichenhan Dieter, PhD, Benedikt Brors, PhD, Karsten Rippe, PhD, Jan-Philipp Mallm, PhD, Frank Rosenbauer, PhD, Christoph Plass, PhD, Matthias Schlesner, PhD, Michael D. Milsom, PhD, Daniel B. Lipka

9.1.5 Manuscripts

Deciphering the Immune Evolution Landscape of Multiple Myeloma Long-Term Survivors Using Single Cell Genomics

<u>Abdelrahman Mahmoud</u>*, Raphael Lutz*, Mohamed H.S. Awwad, Charles Imbusch, Tobias Boch, Niels Weinhold, Marc S. Raab, Carsten Müller- Tidow, Brian Durie, Simon Haas, Hartmut Goldschmidt, Benedikt Brors, and Michael Hundemer

(in preparation)

Loss of the LSD1 protein but not its enzymatic activity promotes leukemia in mice through mislocalization of NPM1 to the cytoplasm

Jonas Samuel Jutzi, **<u>Abdelrahman Mahmoud</u>,** Judith Mueller, Lars Feuerbach, Monika Gothwal, Benedikt Brors, Roland Schuele and Heike Luise Pahl

(in preparation)

Selective Elimination of Immunosuppressive T cells in Patients with Multiple Myeloma

Mohamed H.S. Awwad, **<u>Abdelrahman Mahmoud</u>,** Heiko Bruns, Hakim Echchannaoui, Katharina Kriegsmann, Marc S. Raab, Uta Bärtsch, Markus Munder, Anna Jauch, Katja Weisel6, Hans Jürgen Salwender, Volker Eckstein, Mathias Hänel, Roland Fenk, Jan Dürig, Benedikt Brors, Carsten Müller-Tidow, Hartmut Goldschmidt, Michael Hundemer

(Accepted in Leukemia journal)

Induction of autoreactive regulatory T cells through promiscuous gene expression by bone marrow-resident antigen presenting cells

Chih-Yeh Chen, Felix Klug, Siao-Han Wong , Franziska Durst , Sheena Pinto, Tomoyoshi Yamano, Dania Riege, Michael Delacher , Maria Dinkelacker , Charles D. Imbusch , **Abdelrahman Mahmoud ,** Roman Kurilov , Miograd Guzvic , Claudia Gebhard , Guido Wabnitz , Valentina Volpin, Ayse Nur Menevse, Yvonne Samstag, Pärt Peterson , Michael Rehli , Slava Stamova, Maria Xydia , Christoph A. Klein , Mark S. Anderson , Christian Schmid , Markus Feuerer , Benedikt Brors, Ludger Klein, Bruno Kyewski , Philipp Beckhove

(under revision)

AKT-dependent NOTCH3 activation drives tumor progression in a model of mesenchymal colorectal cancer

Varga, Julia, Adele Nicolas, Valentina Petrocelli, Marina Pesic, **Abdelrahman Mahmoud**, Birgitta E. Michels, Emre Etlioglu, et al.

(2020, Journal of Experimental Medicine journal)

CRISPR/Cas9-edited NSG mice as PDX models of human leukemia to address the role of niche-derived SPARC

Tirado-Gonzalez, I., E. Czlonka, A. Nevmerzhitskaya, D. Soetopo, E. Bergonzani, **A. Mahmoud**, A. Contreras, et al.

(2018, Leukemia journal)

10 Acknowledgments