Inaugural dissertation

for

obtaining the doctoral degree

of the

Combined Faculty of Mathematics, Engineering and Natural Sciences

of the

Ruprecht - Karls - University

Heidelberg

Presented by

M.Sc. (Tech) Aino-Maija Leppä

Born in: Lieto, Finland

Oral examination: 12th May 2023

Characterization of the molecular heterogeneity in acute myeloid leukemia for effective targeting of disease-driving leukemic stem cells

Referees:

Prof. Dr. Andreas Trumpp

Dr. Michael Milsom

Summary

The human body is made of trillions of cells that are the building blocks of all living things. Hematopoietic cells are a set of such building blocks. They are made in the bone marrow and carry out vital tasks, including fighting off infections, facilitating wound healing and carrying oxygen through the body. In health, hematopoietic cells are crucial for the maintenance of normal blood cell production which is necessary for a healthy life. In Acute Myeloid Leukemia (AML), errors in DNA repair result in the acquisition of genetic variations that drive malignant changes and growth of abnormal myeloid cells. These malignant cells fail to perform their vital functions and further hamper the production of normal blood cells.

AML is driven by a group of cells in the bone marrow with self-renewing capacity that give rise to a diverse progeny of abnormal myeloid cells. AML remains lethal due its complex and plastic cellular nature characterized by a high degree of intra-tumor heterogeneity. With the introduction of single-cell technologies, advancements in characterizing the genetic and non-genetic landscape of AML has improved. However, attempts to connect the different levels of heterogeneity, identify and target the disease-driving leukemic stem cells (LSCs) and assess the resulting functional outcomes are largely still lacking. Moreover, identifying the patients who best benefit from novel targeted therapies compared to standard cytotoxic therapies remains a challenge.

In this thesis, I first investigated the intra-patient heterogeneity of complex karyotype AML using an integrated single-cell multi-omics framework that combines structural variant discovery and nucleosome occupancy profiling (scNOVA) with concurrent immunophenotypic and transcriptomic profiling (CITE-seq). Using this framework, I revealed complex structural variant landscapes in single CK-AML cells, marked by ongoing karyotype evolution with frequent involvement of chromothripsis along with linear and circular breakage-fusion-bridge events mediating genomic remodeling. I further unveiled extensive cell-to-cell karyotype instability, exemplified by instable chromosome intermediates, like complex marker and ring chromosomes. Next, I characterized the intra-patient heterogeneity and revealed the existence of genetically distinct subclones with unique nucleosome occupancy, and transcriptomic and immunophenotypic features. By transplanting these cells into immunocompromised mice, I observed predominantly monoclonal expansion of subclones with high genomic complexity that were enriched for stemness-associated phenotypes, including high 17-gene stemness scores

and expression of stem cell markers such as CD49F and CD90. Furthermore, I showed that these disease-driving LSCs showed resistance to standard chemotherapy *ex vivo* but could be targeted by BH3 mimetics. Finally, in an index patient, I showed that the patient-derived xenograft system recapitulated the subclone-specific evolution also during disease progression in the patient, offering a promising model to study relapse. Together, these data provide unique insights into the ongoing genetic and phenotypic complexity of CK-AML, highlight the clinical relevance of intra-patient heterogeneity in tumor evolution, and offer promising avenues to functionally explore and target the disease-driving LSCs.

Next, I explored the clinical relevance of disease-driving LSCs in AML by taking part in investigating how they can be used to predict response to a newly-approved targeted therapy comprising the BCL-2 inhibitor venetoclax in combination with azacytidine. By integrating transcriptomic, functional and clinical data we aimed to identify predictors of clinical response to this combination therapy. We revealed that while more differentiated monocytic cells had high MCL-1 expression and showed resistance to venetoclax and azacytidine, they consistently lacked disease-initiating potential and thus did not fuel leukemogenesis. In contrast, the cells with consistent LSC potential expressed high levels of BCL-2 and could be efficiently targeted *ex vivo*. We further showed that combining BCL-2, BCL-xL and MCL-1 protein expression ratios in these disease-driving LSCs, could be used to determine the clinical response to venetoclax and azacytidine. This flow cytometry-based "Mediators-of-Apoptosis-Combinatorial-Score" (MAC-Score) predicted initial response with a positive predictive-value of >97% and was associated with increased event-free survival. These data show that the combinatorial levels of BCL-2-family members in the disease-driving LSCs are a key determinate of response to venetoclax and azacytidine and that affordable techniques can be used to reliably predict response to this therapy.

In summary, I investigated different levels of intra-patient heterogeneity in CK-AML patient samples using an integrated single-cell multi-omics framework and explored the resulting functional outcomes. I also took part in predicting clinical response to the newly-approved therapy of venetoclax in combination with azacytidine by establishing a flow cytometry-based response score. Collectively the thesis emphasizes the importance of better identifying and characterizing the disease-driving LSCs to improve our understanding of AML as a dynamic disease entity and to offer effective ways to assess and target the disease-driving LSCs.

Zusammenfassung

Der menschliche Körper besteht aus Billionen von Zellen, die die Bausteine aller Lebewesen sind. Hämatopoetische Zellen sind ein Satz solcher Bausteine. Sie werden im Knochenmark gebildet und erfüllen lebenswichtige Aufgaben wie die Bekämpfung von Infektionen, die Erleichterung der Wundheilung und den Transport von Sauerstoff durch den Körper. Im gesunden Zustand sind blutbildende Zellen entscheidend für die Aufrechterhaltung einer normalen Blutzellenproduktion, die für ein gesundes Leben notwendig ist. Bei der akuten myeloischen Leukämie (AML) führen Fehler in der DNA-Reparatur zum Erwerb von genetischen Variationen, die bösartige Veränderungen und das Wachstum abnormaler myeloischer Zellen fördern. Diese bösartigen Zellen können ihre lebenswichtigen Funktionen nicht erfüllen und behindern zudem die Produktion normaler Blutzellen.

Die AML wird von einer Gruppe von Zellen im Knochenmark mit Selbsterneuerungskapazität angetrieben, die eine vielfältige Nachkommenschaft von abnormen myeloischen Zellen hervorbringen. Aufgrund ihrer komplexen und plastischen zellulären Natur, die durch ein hohes Maß an Heterogenität innerhalb des Tumors gekennzeichnet ist, bleibt die AML tödlich. Mit der Einführung von Einzelzelltechnologien haben sich die Fortschritte bei der Charakterisierung der genetischen und nicht-genetischen Landschaft der AML verbessert. Allerdings fehlt es noch weitgehend an Versuchen, die verschiedenen Ebenen der Heterogenität miteinander zu verbinden, die krankheitsauslösenden leukämischen Stammzellen (LSC) zu identifizieren und gezielt zu behandeln und die daraus resultierenden funktionellen Ergebnisse zu bewerten. Darüber hinaus ist es nach wie vor eine Herausforderung, die Patienten zu identifizieren, die am besten von neuartigen zielgerichteten Therapien im Vergleich zu zytotoxischen Standardtherapien profitieren.

In dieser Arbeit untersuchte ich zunächst die Intra-Tumor-Heterogenität von AML mit komplexem Karyotyp unter Verwendung eines integrierten Einzelzell-Multiomik-Rahmens, der die Entdeckung struktureller Varianten und die Erstellung von Nukleosomenbelegungsprofilen (scNOVA) mit gleichzeitiger immunphänotypischer und transkriptomischer Profilerstellung (CITE-seq) kombiniert. Mithilfe dieses Systems konnte ich komplexe strukturelle Variantenlandschaften in einzelnen CK-AML-Zellen aufdecken, die durch eine fortlaufende Karyotyp-Evolution gekennzeichnet sind, an der häufig Chromothripsen sowie lineare und zirkuläre Bruch- und Fusionsbrückenereignisse beteiligt

sind, die den genomischen Umbau vermitteln. Darüber hinaus konnte ich eine weitreichende Karyotyp-Instabilität von Zelle zu Zelle feststellen, die durch instabile Chromosomen-Intermediate, wie komplexe Marker- und Ringchromosomen, veranschaulicht wird. Anschließend habe ich die Heterogenität innerhalb des Patienten charakterisiert und die Existenz genetisch unterschiedlicher Subklone mit einzigartigen Nukleosomenbesetzungen, transkriptomischen und immunphänotypischen Merkmalen nachgewiesen. Bei der Transplantation dieser Zellen in immungeschwächte Mäuse beobachtete ich eine überwiegend monoklonale Expansion von Subklonen mit hoher genomischer Komplexität, die mit Stammzell-assoziierten Phänotypen angereichert waren, einschließlich hoher 17-Gene-Stammzell-Scores und Expression von Stammzellmarkern wie CD49F und CD90. Darüber hinaus konnte ich zeigen, dass diese krankheitsverursachenden LSC ex vivo resistent gegen eine Standard-Chemotherapie sind, aber durch BH3-Mimetika gezielt bekämpft werden können. Schließlich konnte ich bei einem Indexpatienten zeigen, dass das vom Patienten stammende Xenotransplantatsystem die subklonspezifische Entwicklung auch während des Krankheitsverlaufs im Patienten rekapitulierte und somit ein vielversprechendes Modell zur Untersuchung von Rückfällen darstellt. Zusammengenommen bieten diese Daten einzigartige Einblicke in die fortschreitende genetische und phänotypische Komplexität der CK-AML, unterstreichen die klinische Relevanz der Heterogenität innerhalb eines Patienten in der Tumorevolution und bieten vielversprechende Möglichkeiten, die krankheitsauslösenden LSCs funktionell zu erforschen und zu bekämpfen.

Als Nächstes untersuchte ich die klinische Relevanz von krankheitsfördernden LSCs bei AML und beteiligte mich an der Untersuchung, wie sie zur Vorhersage des Ansprechens auf eine neu zugelassene zielgerichtete Therapie mit dem BCL-2-Inhibitor Venetoclax in Kombination mit Azacytidin genutzt werden können. Durch die Integration von transkriptomischen, funktionellen und klinischen Daten wollten wir Prädiktoren für das klinische Ansprechen auf diese Kombinationstherapie ermitteln. Es zeigte sich, dass stärker differenzierte monozytäre Zellen zwar eine hohe MCL-1-Expression aufwiesen und gegen Venetoclax und Azacytidin resistent waren, ihnen aber durchweg das krankheitsauslösende Potenzial fehlte und sie somit die Leukämogenese nicht förderten. Im Gegensatz dazu exprimierten die Zellen mit konsistentem LSC-Potenzial hohe Mengen an BCL-2 und konnten ex vivo effizient bekämpft werden. Wir konnten außerdem zeigen, dass die Kombination der Proteinexpressionsverhältnisse von BCL-2, BCL-xL und MCL-1 in diesen krankheitsauslösenden LSCs zur Bestimmung des klinischen Ansprechens auf Venetoclax und

Azacytidin verwendet werden kann. Dieser auf der Durchflusszytometrie basierende "Mediators-of-Apoptosis-Combinatorial-Score" (MAC-Score) sagte das anfängliche Ansprechen mit einem positiven Vorhersagewert von >97 % voraus und war mit einem längeren ereignisfreien Überleben verbunden. Diese Daten zeigen, dass die kombinatorischen Spiegel von Mitgliedern der BCL-2-Familie in den krankheitsauslösenden LSCs eine Schlüsseldeterminante für das Ansprechen auf Venetoclax und Azacytidin sind und dass erschwingliche Techniken zur zuverlässigen Vorhersage des Ansprechens auf diese Therapie verwendet werden können.

Zusammenfassend untersuchte ich verschiedene Ebenen der Heterogenität innerhalb von Patientenproben von CK-AML-Patienten unter Verwendung eines integrierten Einzelzell-Multiomics-Systems und untersuchte die daraus resultierenden funktionellen Ergebnisse. Außerdem habe ich an der Vorhersage des klinischen Ansprechens auf die neu zugelassene Therapie mit Venetoclax in Kombination mit Azacytidin mitgewirkt, indem ich einen auf Durchflusszytometrie basierenden Response Score erstellt habe. Insgesamt unterstreicht die Dissertation die Bedeutung einer besseren Identifizierung und Charakterisierung der krankheitsauslösenden LSCs, um unser Verständnis der AML als dynamische Krankheitseinheit zu verbessern und wirksame Methoden zur Bewertung und gezielten Bekämpfung der krankheitsauslösenden LSCs anzubieten.

Table of Contents

List of Abbreviations

ADT – Antibody-Derived Tag

alloHSCT – Allogeneic Hematopoietic Stem Cell Transplantation

AML – Acute Myeloid Leukemia

BFB – Breakage-Fusion-Bridge

BrdU – Bromodeoxyuridine

CITE-seq – Cellular Indexing of Transcriptomes and Epitopes by Sequencing

CK-AML – Complex karyotype AML

CMP- Common myeloid progenitor

CNV – Copy Number Variation

EFS – Event-Free Survival

FAB – French-American-British

FACS – Fluorescent activated cell sorting

GMP – Granulocyte-monocyte progenitor

HMA – Hypomethylating Agents

HSC – Hematopoietic Stem Cell

HSPC – Hematopoietic Stem and Progenitor Cells

IMDM – Iscove's Modified Dulbecco's Medium

LMPP – Lymphoid Primed Multipotent Progenitor

LSC – Leukemic Stem Cell

M-FISH – Multiplex Fluorescence *in situ* Hybridization

MAC-Score – Mediators of Apoptosis Combinatorial-Score

MACS – Magnetic-Activated Cell Sorting

MDS – Myelodysplastic Syndrome

MEP – Megakaryocyte-Erythrocyte Progenitor

MFI – Mean Fluorescent Intensity

MNase – Micrococcal Nuclease

MOMP – Mitochondrial Outer Membrane Permeabilization

Mono-AML – Monocyte-like AML

MPN – Myeloproliferative Neoplasm

NSG – NOD.*Prkdc*^{scid}.*Il2rg*^{null}

PCA – Principal Component Analysis

PDX – Patient-derived Xenograft

Prim-AML – Primitive-like AML

Prog Mk – Megakaryocytic Progenitor

Prog RBC – Red Blood Cell Progenitor

scADT-seq – Single-Cell Antibody-Derived Tag Sequencing

scNOVA – Single-Cell Nucleosome Occupancy and Genetic Variation Analysis

scRNA-seq – Single-Cell RNA Sequencing

scTRIP – Single-Cell Tri-Channel Processing

SD – Stable Disease

SNP – Single-Nucleotide Polymorphism

SNV – Single Nucleotide Variant

Strand-seq – Single-Cell Template Strand-Sequencing

SV – Structural Variant

tSNE – t-Distributed Stochastic Neighbor Embedding

UMAP – Uniform Manifold Approximation and Projection

UMI – Unique Molecular Identifier

WNN – Weighted Nearest Neighbor

List of Figures

List of Tables

Chapter 1 – Introduction

1.1 A brief history of Acute Myeloid Leukemia

Acute myeloid leukemia (AML) is an aggressive hematological malignancy that is characterized by the infiltration of the bone marrow, blood, and other tissues by abnormally differentiated, highly proliferating cells of the hematopoietic system[1-3]. AML stems from hematopoietic stem and progenitor cells (HSPCs) that through a stepwise acquisition of genetic and epigenetic aberrations reprogram and subsequently transform the healthy hematopoietic system (Figure 1.1)[4]. The malignant transformation typically causes uncontrollable proliferation and differentiation block in the HSPCs resulting in the accumulation of dysfunctional leukemic progenitor cells in the bone marrow[3]. To date, AML is the most common adult leukemia with up to 21,000 new cases annually in the USA alone and a 5-year overall survival of 30-35% in younger patients (<60 years) and 10-15% in older patients (≥60 years)[1,5].

The early 19th century presented the first publications describing cases of uncommon or peculiar alterations of the blood with the first accurate description of leukemia likely made by a French physician Velpeau in 1827. He used the descriptive term leukemia (*leukos* referring to white and *haima* referring to blood) to describe the blood of a patient[6,7]. However, the recognition of leukemia as a clinical entity is credited to the nearly simultaneous reports by an English pathologist Bennet and a German pathologist Virchow in 1845[6,8]. Virchow and others attempted to further distinguish different forms of leukemia with Virchow categorizing them into splenic and lymphatic leukemias, closely resembling the current myeloid and lymphoid categorization[6].

In 1868 German pathologist Ernst Neumann established the link between blood and bone marrow by identifying changes in the bone marrow in leukemia patients and concluding that leukemia was a disease of the marrow[6,9]. The description of leukemia was further refined by the development of methods for staining blood cells by Paul Ehrlich who published this technique in 1879[10,11]. These advances in differentiating between types of blood cells simplified the classification of leukemia into myeloid and lymphoid groups[12], abrogating the category of splenic leukemias introduced by Virchow. By staining blood cells, Ehrlich was also likely the first to describe the concept of a stem cell. He reported of a primitive ancestor cell in a semi-transformed state giving rise to cells within a distinct cell lineage[13].

The first clinical description of AML is often attributed to Wilhem Ebstein who used acute leukemia to describe a fast advancing illness in 1889[11,14]. This is considered one of the first distinctions between chronic and acute subtypes of leukemia. This division of leukemia into chronic and acute subtypes together with Ehrlich's classification of myeloid and lymphoid leukemias as well as the description of a new cell type of the myeloid cell lineage, termed myeloblast, by a Swiss hematologist Naegeli formed the discovery of AML as a distinct disease entity[13,15].

1.2 Acute myeloid leukemia – Cancer of the blood and bone marrow

According to the recent updated World Health Organization classification, AML diagnosis is based on disease-defining genetic alterations or by the presence of at least 20% of partially differentiated precursor cells called AML blasts in the bone marrow[16]. The cause of AML is not completely clear but known risk factors include age, chemical exposures to benzene, herbicides, and pesticides as well as exposure to alkylating agents, type II topoisomerases and ionizing radiation[17]. Besides environmental and therapy-related risk factors, non-malignant disorders of the hematopoietic system such as clonal hematopoiesis have been found to commonly precede AML[18,19]. Moreover, genetic disorders such as Fanconi anemia and Shwachman-Diamond syndrome or chronic clonal hematological diseases, including myelodysplastic syndrome (MDS) and myeloproliferative neoplasms (MPNs), often progress to AML[17,18].

With the introduction of second-generation sequencing approaches, the last 10 years have presented a massive explosion of genomic data of different cancer entities. Studies on AML genetics have shown that AML is a genetically heterogeneous disease at diagnosis with at least one subclone present in addition to the major clone[20,21]. Recurrent AML mutations can be categorized into different functional categories based on their putative role in AML pathogenesis. These categories include: transcription-factor fusions (e.g. *RUNX1-RUNX1T1*), tumor suppressor genes (e.g. *TP53*), DNA methylation-related genes (e.g. *DNMT3A*, *TET2*, *IDH1/2*), activated signaling genes (e.g. *FLT3, K/NRAS*), myeloid transcription factor genes (e.g. *RUNX1*), chromatin modifier genes (e.g. *ASXL1*, *KMT2A*), cohesin-complex genes (e.g. *RAD21*, *STAG2*), spliceosome-complex genes, and the gene encoding nucleophosmin (*NPM1*)[20,21]. Moreover, patterns of mutational co-occurrence and exclusivity between most frequent variants suggest biological cooperation between certain mutational events, such as co-occurrence of *NPM1* and *FLT3-ITD* or general mutual exclusivity for mutations affecting the

same functional categories, such as *IDH1/2* and *TET2*[20,21]. More recent single-cell genomic studies have further refined the understanding of clonal architecture and evolutionary histories of AML mutations largely supporting the co-occurrence and exclusivity patterns observed in bulk studies[22].

The increasing insights into the clinical value of genomic abnormalities have promoted the importance of stratifying patients into risk groups based on genetic abnormalities[23,24]. The European LeukemiaNet (ELN) recommends stratification of patients into favorable, intermediate and adverse risk categories based on genetic characteristics at diagnosis to help triage patients for optimal therapies[24]. For instance, *TP53* mutations that commonly occur with complex or monosomal karyotype are associated with adverse risk[24,25]. Both from clinical and molecular perspectives, increasing amounts of evidence even indicate that *TP53*-mutant AML (and MDS) may represent a distinct disease entity[26,27]. In contrast, a set of common translocation resulting into fusion proteins, such as t(8;21)(q22;q22.1)/*RUNX1::RUNX1T1* and t(16;16)(p13.1;q22)/*CBFB::MYH11* are associated with favorable prognosis.

1.2.1 Complex karyotype AML

AML patients that in the absence of the prognostically favorable aberrations t(8;21)(q22;q22), inv(16)(p13q22)/t(6; 16)(p13;q22) carry three or more acquired chromosome aberrations, form a separate category called AML with complex karyotype (CK-AML)[28,29]. CK-AML comprises 10-12% of all AMLs and is the second-largest cytogenetic AML subgroup[25]. CK-AML is defined by a karyotype with ≥3 chromosome abnormalities, and is commonly associated with *TP53* mutations and complex genomic rearrangements including chromothripsis events[28,30]. The survival rates in CK-AML are very poor with only 15% of patients alive one year after diagnosis[25,31,32].

Compared to *de novo* AMLs, CK-AML is twice as common in patients with a previous hematologic malignancy or who have received prior cancer treatment[33,34]. CK-AMLs are not restricted to a specific morphological state but they are enriched for morphologically more immature M0, M1, and erythroid-biased M6 types based on the French-American-British (FAB) classification[34,35]. In line with their morphology, CK-AMLs have a higher expression of the primitive HSPC cell surface marker CD34, and lower expression of differentiation-associated cell surface markers, such as CD64, compared to other AMLs[34].

Genetically CK-AML is a highly heterogeneous subgroup and different karyotype abnormalities have been reported. Unbalanced karyotypes resulting from gain or loss of genetic material are more common than balanced chromosome aberrations[28,34]. The majority of aberrations detected in CK-AML are structural and most frequently lead to loss of part of a chromosome. Indeed, genomic losses predominate in CK-AML and loss of 5q, 7q and deletion of 17p are the three most commonly deleted regions[28,32,34]. In contrast, genomic gains are not as common but when present often involve chromosomes 8, 11 and 21[28,34].

1.2.2 Leukemic Stem Cells

The hematopoietic system is hierarchically organized with self-renewing hematopoietic stem cells (HSCs) located at the top of the hierarchy[36,37]. HSCs give rise to more specialized hematopoietic progenitor cells that differentiate into functional mature blood cells[36,37]. Similar to the healthy system, xenograft models to study leukemia regeneration showed the first direct evidence that AML also retains a hierarchical structure with leukemic stem cells (LSCs) located at the top of the hierarchy and fueling the bulk of the leukemia[38,39]. These studies demonstrated that primitive stem and early progenitor cells contain high self-renewal potential and constitute the cell of origin in malignant transformation of AML (Figure 1.1)[38,39].

In addition to their ability to initiate leukemia when transplanted into immunocompromised mice, LSCs have several characteristic features. These include high self-renewal potential[38-40], ability to give rise to more differentiated cells, cell cycle quiescence[41,42] and chemoresistance[40,43,44]. Moreover, LSCs are clinically relevant as high expression of stemness signatures derived from functional LSCs are a strong predictor of poor outcome in AML patients[45,46]. Notably, cancer stemness features are not always a hard-wired phenotype but may also contain some plasticity (Figure 1.1)[47]. Plasticity can help cancer stem cells escape therapeutic pressure by facilitating different escape strategies. In AML, LSCs have been shown to enter a senescent or dormant-like cell state, take advantage of their metabolic plasticity and upregulate their fatty acid metabolism to survive therapy[48-51]. In addition to plasticity, genetic heterogeneity and epigenetic heterogeneity contribute to LSC diversity, therapy resistance, and relapse (Figure 1.1)[22,52,53].

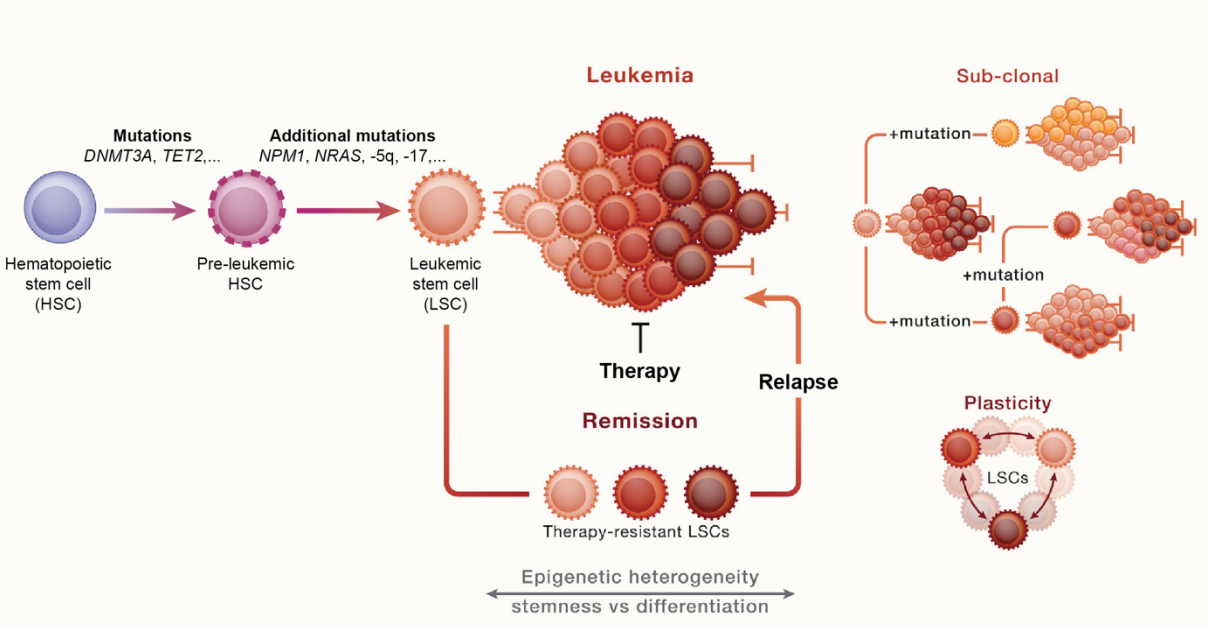


Figure 1.1. Stepwise acquisition of genetic aberrations in AML development and progression. During healthy hematopoiesis HSCs acquire mutations (e.g. DNMT3A and TET2) that give rise the pre-leukemic HSCs with increased clonal growth. With additional genetic aberrations (e.g. NPM1, FLT3-ITD, loss of 5q and 17q), pre-leukemic HSCs transform into LSCs that fuel the bulk of the leukemia. Plasticity and cell-to-cell heterogeneity can help LSCs survive therapy and give rise to relapse. Adapted from Trumpp and Haas [54].

1.2.3 Treatment landscape of current and emerging therapies

The standard of care for AML patients consists of induction chemotherapy followed by appropriate post-remission therapy. Patients eligible for intensive induction chemotherapy receive a combination of high-dose cytarabine with an anthracycline such as daunorubicin or idarubicin[2,23]. The choice of post-remission therapy largely depends on the genetic risk profile of the patient. Generally, patients with favorable genetic risk receive 2 to 4 cycles of intermediate-dose cytarabine while patients with intermediate or adverse risk genetics, if possible, receive an allogeneic hematopoietic transplantation (alloHSCT)[2,23]. In recent years, the use of targeted precision medicine in combination with standard chemotherapy has been shown to improve overall survive in a selected subset of favorable and intermediate risk patients[23]. Such targeted approaches include the use of *FLT3* inhibitors, such as midostaurin and gilterinib, in *FLT3*-mutated AMLs, and antibody-toxin conjugates, such as gemtuzumab ozogamicin in CD33-positive AMLs (Figure 1.2)[23,55].

Treatment alternatives for patients unfit for intensive therapy are often limited to best supportive care, low-intensity therapies, such as low-dose cytarabine or hypomethylating agents (HMAs), or clinical trials with investigational drugs[23]. Over the past few years several

of such investigational drugs have been approved for front-line treatment of patients unfit for intensive therapy (Figure 1.2)[55]. This has paved the way for future research opportunities for patients lacking effective therapeutic options. Such investigational drugs include IDH inhibitors ivosidenib and enasidenib against *IDH1*/*2*-mutated AMLs[56,57] and BH3 mimetic venetoclax together with hypomethylating agent azacytidine or decitabine to target the anti-apoptotic protein BCL-2[58,59].

In addition to targeted inhibitors already approved for treatment of AML, several other agents are currently under (pre-)clinical investigation (Figure 1.2)[55]. Among these are other BH3-mimetics including MCL-1 inhibitors and BCL-xL inhibitors that have entered pre-clinical and phase I/II clinical trials[60]. Among immunotherapies, monoclonal antibodies that promote cell death by blocking important ligand-receptor interactions (e.g. anti-CD70 antibody cusatuzumab) or by Fc-dependent phagocytosis (e.g. anti-CD47 antibody magrolimab) are currently also tested in AML. Bispecific T cell engagers that mediate immune cell recruitment by recognizing a specific T cell epitope, such as CD3, as well as epitopes expressed in AML cells, such as CD33 or CD123, are currently also in clinical development[55]. Other targeted agents that have gained particular interest include: Smoothened inhibitor Glasdegib that targets the hedgehog–glioma-associated oncogene homologue (HH–GLI) signaling pathway; menin inhibitor KO-539 that targets the MLL histone methyltransferase complex; and mutant p53-specific small molecule eprenetapopt that re-activates wild-type p53 function[55].

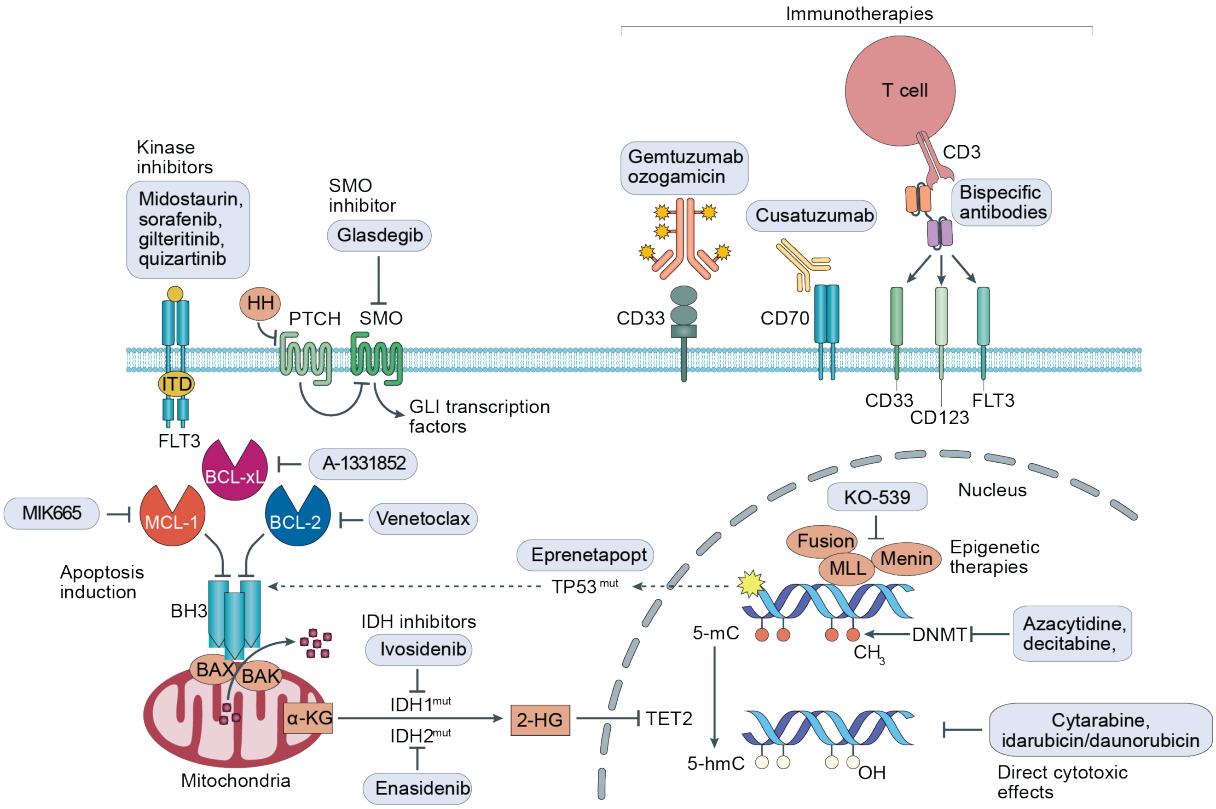


Figure 1.2. Overview of cellular targets in treatment of AML. The development of new therapeutic agents has provided novel treatment modalities in AML, including immunotherapies, epigenetic therapies and targeted therapies directed towards mutations and/or pathway-specific dependencies. These therapeutic agents provide alternatives to standard chemotherapy particularly in patients unfit for regimens with severe toxicities. Abbreviations: Mut, mutated; SMO, Smoothened; DNMT, DNA methyltransferase; GLI, glioma-associated oncogene homologue; 2-HG, 2-hydroxyglutarate; 5-hmC, 5-hydroxymethylcytosine; ITD, internal tandem duplication; α-KG, α-ketoglutarate; 5-mC, 5-methylcytosine. Adapted from Dohner, Wei [55].

1.3 Biology of Human Cancer Genomics

Each cell in the human body receives tens of thousands DNA lesions per day[61]. To maintain genome integrity, human cells have evolved several systems to detect DNA damage, signal its presence and mediate its repair[62]. Yet, each DNA repair pathway carries a chance of error that can cause incorrect repair and lead to formation of structural variants (SVs)[62,63]. SVs are defined as larger genetic variants that are at least 50 base pair in size, and can include inversions, insertions, deletions, duplications, translocations, chromosomal losses and gains, as well as more complex sets of chromosomal rearrangements[63]. SVs are important for the development and progression of cancer as they can amplify, disrupt and fuse cancer-related genes, and thus dysregulate gene expression[63]. Indeed, SVs are ubiquitously detected in cancer but show diversity in their prevalence between different cancer entities[64,65]. Solid cancers, in particular esophagus, ovarian and breast adenocarcinomas and osteosarcomas, harbor a high somatic SV

burden whereas hematological malignancies, including AML, exhibit a relatively low SV burden[64,65].

Several factors determine SV heterogeneity, including genetic and exogeneous determinants[63]. Up to 10% or cancers are considered to have a hereditary predisposition due to germline mutations in cancer susceptible genes[66]. Most well described germline mutations include the *BRCA* mutations that affect the two tumor suppressor genes *BRCA1* and *BRCA2,* commonly mutated in breast and ovarian adenocarcinomas. Mutations in these genes are associated with defective homologous-recombination-based DNA repair, resulting in an increase in SV burden[67]. Similarly, germline mutations in the tumor suppressor *TP53* have also been associated with an increase in SV burden in medulloblastoma through increased prevalence of chromothripsis, the shattering of the chromosome (see below)[30]. Next to germline mutations, environmental sources ranging from viruses to ionizing radiation and chemotherapeutic agents have been linked to SV mutagenesis and represent serious carcinogens[63,68].

1.3.1 Technologies to characterize structural variants in cancer

The development of technologies to assess SVs has been crucial for the current understanding of genomic rearrangements and their role in cancer. Traditional cytogenetic techniques, including G-banding and fluorescence *in situ* hybridization, have existed for many decades and still remain an important diagnostic tool for detecting large chromosomal abnormalities especially in hematological malignancies (Figure 1.3A)[63,69]. The development of microarray-based methods has improved the resolution of SV detection and enabled the study of submicroscopic SVs, including copy-number changes arising from deletions and duplications (Figure 1.3A)[63]. Microarray-based studies have also fostered the study of more complex genomic concepts, such as chromothripsis[30].

Short-read-based sequencing methods have followed probe-based microarrays, markedly improving the study of single nucleotide variants (SNVs) and small inversion and deletions less than 50 bases in size. However, study of SVs using whole-exome or panel-sequencing is suboptimal due to the sparse coverage of the genome, making the detection of SVs smaller than a few megabases in size difficult[63]. In contrast, short-read-based whole-genome sequencing has been extensively used to study a wide spectrum of SVs by combining paired-end mapping, read-depth analysis, and split-read detection (Figure 1.3B)[64,65,70-72].

More recently, long-fragment-based methods have gained popularity due to their ability to overcome the limitations of short reads and improve the detection of large-scale SVs (Figure 1.3C)[63,73]. Long-read sequencing technologies from Oxford Nanopore Technologies and Pacific Biosciences have already shown the applicability of long-read sequencing for population-based SV studies and enabled the assembly of the first complete human genome[74-76]. Next to long-read sequencing approaches, optical genome mapping has been introduced as a non-sequencing-based alternative for assessment of large-scale SVs[77]. It uses high molecular-weight DNA for enzymatic labeling of the genome at specific sequence motifs and compares the resulting genomic landmark maps to a reference. Optical genome mapping has been used for a variety of applications from the study of SVs in primary patient samples with MDS to reconstructing complex genomic rearrangements in cancer cell lines[78,79]. In addition to the bulk analytical approaches described here, single-cell sequencing approaches, including single-cell strand-based methods, have received increasing attention for the study of SVs, and are discussed in more detail in Section 1.4 (Figure 1.3D).

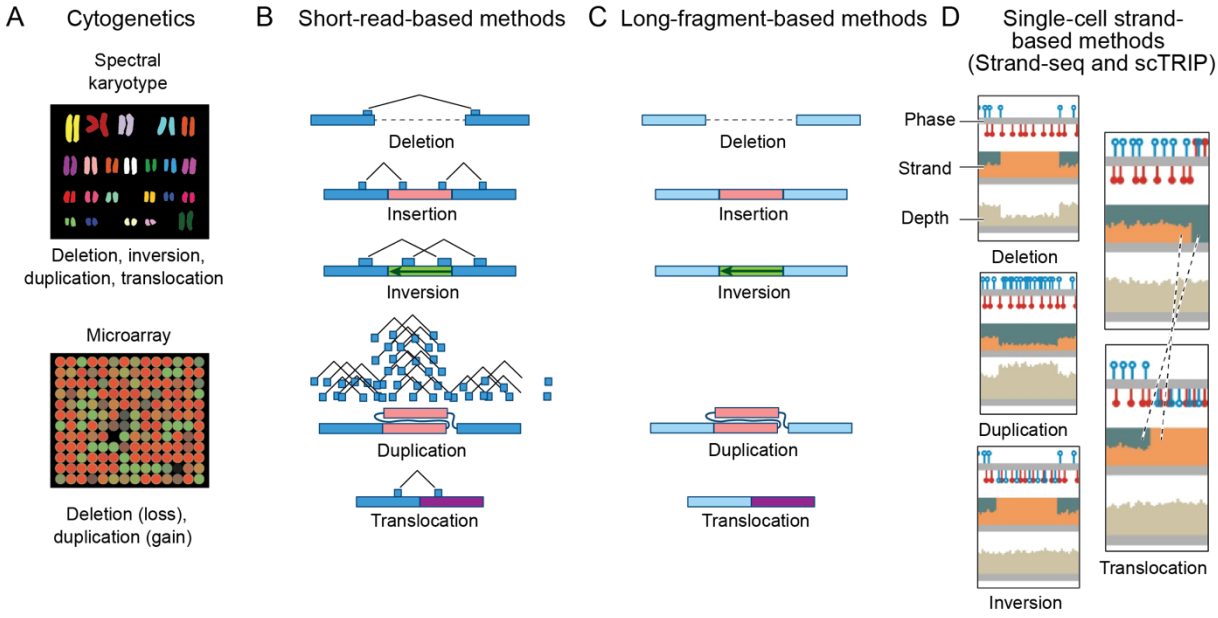


Figure 1.3. Overview of technologies for SV discovery. A Cytogenetics include methods developed prior to sequencing approaches that rely on DNA staining (G-banding) or on fluorescent probes that target chromosome abnormalities and genomic loci (FISH technologies (e.g. spectral karyotype) and microarrays). **B** Short-read-based methods, in particular whole genome short-read sequencing, enable the detection of SVs using computational pipelines that combine paired-end mapping, read-depth analysis, and split-read detection. **C** Long-fragment-based methods include long-read sequencing and optical genome mapping and provide improved detection of large-scale SVs due to reads or maps that span the SV. **D** Single-cell strand-based methods enable SV calling at the single-cell level by integrating three distinct data layers to build single-cell SV landscapes. These data layers include read-orientation, read-depth, and haplotype-phase (see Section 1.4). Adapted from Cosenza, Rodriguez-Martin [63].

1.3.2 Simple to complex genomic rearrangements

The detection and repair of DNA breaks are of crucial importance for the maintenance of genomic integrity. Somatic SVs arise when such breaks are not properly repaired by the multiple repair pathways. The unique characteristics of each repair pathway serve as causal mechanisms for different SV patterns[63]. Large pan-cancer analyses have described a spectrum of SV patterns that share complementary causal mechanism and breakpoint features that together can be used to classify somatic SVs into different categories[64,65,80]. This classification results in different SV classes that range from simple events, such as duplications and deletions, to complex, localized rearrangements that arise from cascading events such as breakage-fusion-bridge (BFB) cycles or from one-off burst-like events termed chromoanagenesis (*chromo* referring to chromosome and *anagenesis* to regeneration)[63]. One form of chromoanagenesis includes chromothripsis where one or several chromosomes shatter into pieces[30,81,82]. The resulting fragments are stitched back together in a random order and orientation, which results in highly localized clustered DNA rearrangements with oscillating copy-number levels often between two copy states[30,81,82].

Simple SV events are often initiated by double-strand breaks that can be repaired by four major DNA repair pathways[63]. These include nonhomologous end joining (NHEJ), homology-directed repair (HDR), break-induced replication (BIR) and error-prone DNA repair, such as microhomology-mediated end joining (MMEJ)[63]. Errors in these repair pathways can lead to the formation of simple SVs ranging from a few bases to kilobases in size with 1-3 breakpoints (Figure 1.4A)[63]. In contrast, single-strand breaks caused by replication stress are repaired by break-induced replication mechanisms, leading to SV clusters of intermediate complexity and less than 1Mb in size. These include, among others, tandem duplications and templated insertions (Figure 1.4B)[63].

Several triggers leading to more complex sets of rearrangements exist, with telomere erosion, terminal deletion and chromosome mis-segregation representing a few[63]. Telomere erosion can lead to telomere crisis that causes a wide array of genomic aberrations, including dicentric-chromosome-linked SVs (Figure 1.4C)[63,83]. A variety of dicentric-chromosome-linked SVs have been reported, including isochromosomes and ring chromosomes, that are triggered by telomere erosion followed by the fusion of chromosome ends or chromatids. These rearrangements can further be remodeled by BFB cycles, giving rise to SV events of increasing complexity and size ranging from megabases to whole chromosomes, affecting potentially

multiple chromosomes[63]. By comparison, chromosome mis-segregation, caused by lagging chromosomes that are left behind during the anaphase in mitosis, generates aneuploidy and micronuclei[63]. Both aneuploidy and micronuclei are further susceptible to chromosome shattering, aka chromothripsis (Figure 1.4D). Chromothripsis can lead to the formation of characteristic copy-number oscillations as well as circular, extrachromosomal double minute chromosomes that can further be remodeled by additional rearrangemens[63,84,85]. Such complex chromothripsis-related rearrangements and dicentric-chromosome-linked SVs are highly interconnected; BFB cycles can lead to chromothripsis and *vice versa*[85,86]. Furthermore, extrachromosomal double minute chromosomes can integrate into homogeneously staining regions in a reversible fashion, making these SV types deeply intertwined (Figure 1.4C-D)[85,87].

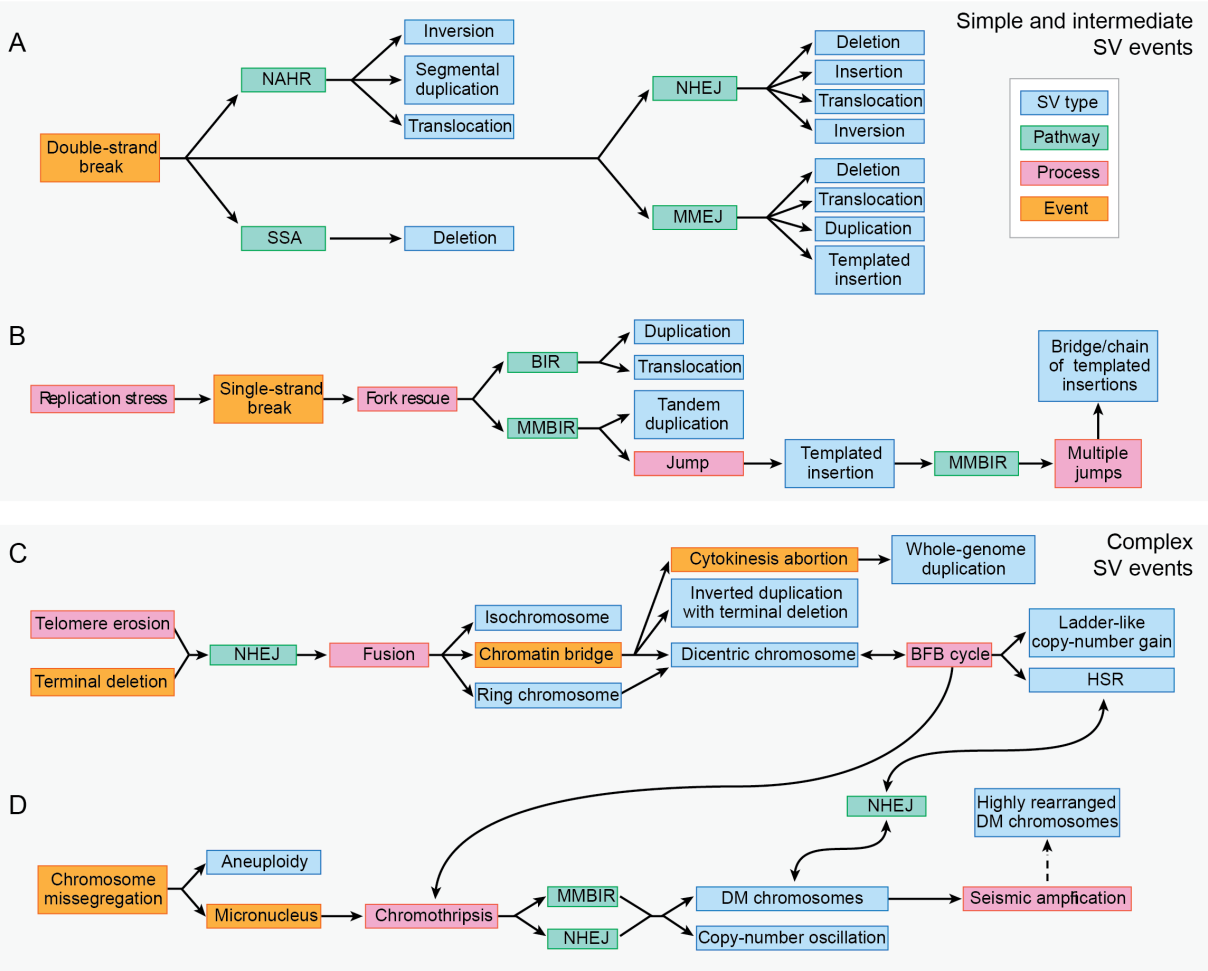


Figure 1.4. SV patterns and the underlying rearrangement trajectories. A-B Simple and intermediate SV events are caused by **A** double- or **B** single-strand breaks that are repaired by different repair pathways. Each repair pathway has its unique characteristics that determine the features of the arising SVs. **C-D** Complex SV events often arise from more severe triggers. These include **C** telomere erosion and deletion as well as **D** chromosome mis-segregation that trigger cascading or one-off burst-like events resulting in SVs with increasing complexity. Complex SV types and processes are deeply intertwined making their rearrangement trajectories highly convoluted. Abbreviations: BFB, breakage–fusion–bridge; BIR, break-induced replication; DM, double minute;

HSR, homogeneously staining region; MMBIR, microhomology-mediated break-induced replication; MMEJ, microhomology-mediated end joining; NAHR, nonallelic homologous recombination; NHEJ, nonhomologous end joining; SSA, single-strand annealing; SV, structural variant. Adapted from Cosenza, Rodriguez-Martin [63].

1.4 Single-Cell Sequencing Technologies

1.4.1 Single-cell template strand-sequencing (Strand-seq)

Different single-cell technologies have emerged as powerful tools to study genetic changes, providing the basis to study the extent and nature of chromosomal rearrangements in single cells[88,89]. Strand-seq is a novel haplotype-resolved single-cell sequencing technology that enables the construction of directional genomic single-cell libraries (Figure 1.5)[90]. Strand-seq relies on the selective removal of the nascent non-template strand to preserve the identity of the two haplotypes within one cell. To achieve this, cells of interest are cultured in the presence of Bromodeoxyuridine (BrdU), a thymidine analog, for a round of one cell division, during which the nascent non-template strands are labeled[90]. The BrdU-labeled non-template strands are selectively removed during genomic library construction, and only the original DNA template strand is amplified to produce a directional single-cell library (Figure 1.5)[90]. During the genomic library construction, the genomic DNA is fragmented using micrococcal nuclease (MNase) to generate mononucleosomal fragments. The MNase digestion produces a specific DNA fragmentation pattern that can be utilized for nucleosome occupancy profiling (see 1.4.2).

Several studies have used the Strand-seq technology to study human genomic variation for different applications, including structural variation analyses[90-94], sister-chromatid exchanges[91,95], and whole-chromosome haplotyping[96,97].

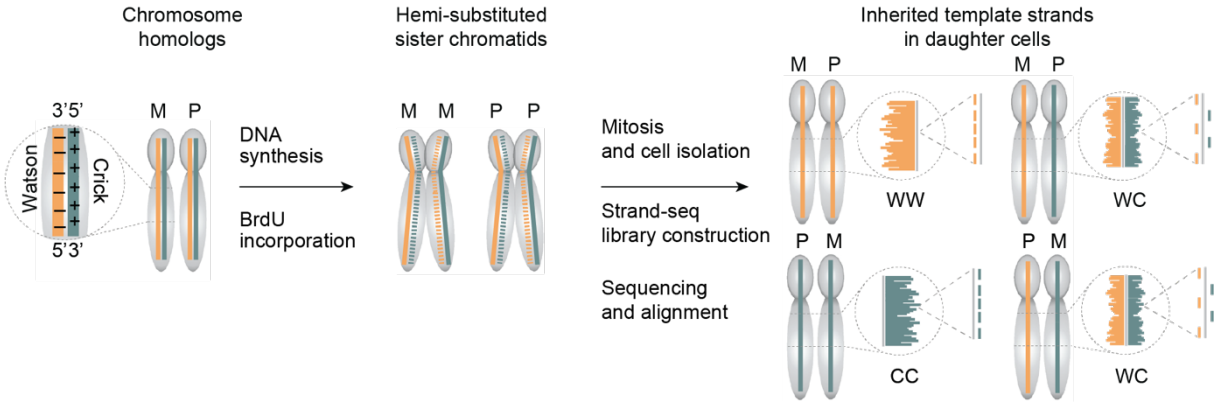


Figure 1.5. Overview of Strand-seq protocol. Cells of interest are cultured in the presence of BrdU for a round of one cell division. The BrdU-labeled non-template strands are selectively removed during genomic library construction, and only the original DNA template strand is amplified to produce a directional single-cell library.

Abbreviations: BrdU, Bromodeoxyuridine; W, Watson; C, Crick; M, maternal; P, paternal. Adapted from Sanders, Falconer [90].

1.4.2 Single-cell Nucleosome Occupancy and Genetic Variation Analysis (scNOVA)

scNOVA is a computational framework built on top of Strand-seq that couples SV discovery with nucleosome occupancy analysis[98]. By leveraging data from Strand-seq, scNOVA further expands the potential of strand-specific sequencing in two orthogonal ways. First, scNOVA performs SV discovery by single-cell tri-channel processing (scTRIP)[93]. scTRIP uses a joint calling framework to integrate three distinct Strand-seq data layers to build single-cell SV landscapes with a detection resolution of 200kb[93]. These data layers include read-orientation, read-depth, and haplotype-phase[93]. Second, scNOVA uses the specific DNA fragmentation pattern introduced during MNase digestion of Strand-seq libraries to directly measure nucleosome occupancy and indirectly infer *cis*-regulatory element accessibility[99-102]. By coupling the SV discovery with nucleosome occupancy analysis, state-of-the-art multi-modal single-cell profiling can be performed (Figure 1.6).

By integrating these different data layers, haplotype-resolved SVs can be characterized together with the arising epigenetic changes in the same cell. SV discovery by scTRIP has been used to identify disease-relevant SV classes in single cells, including simple rearrangements such as deletions and duplications as well as complex genomic rearrangements, including BFB events[93]. Moreover, scTRIP has been applied to detect ongoing mutational processes in cancer acting in individual cells, providing detail to the role of genomic instability in cancer evolution[93]. As scNOVA allows the direct assessment of SVs with their functional consequences in heterogeneous cell populations, it has also in a recent study been used to deconvolve the effects of complex genomic rearrangements in cancer[98]. This study showed at single-cell level that somatic SVs can result in epigenetic dysregulation of oncogenic transcription factors[98].

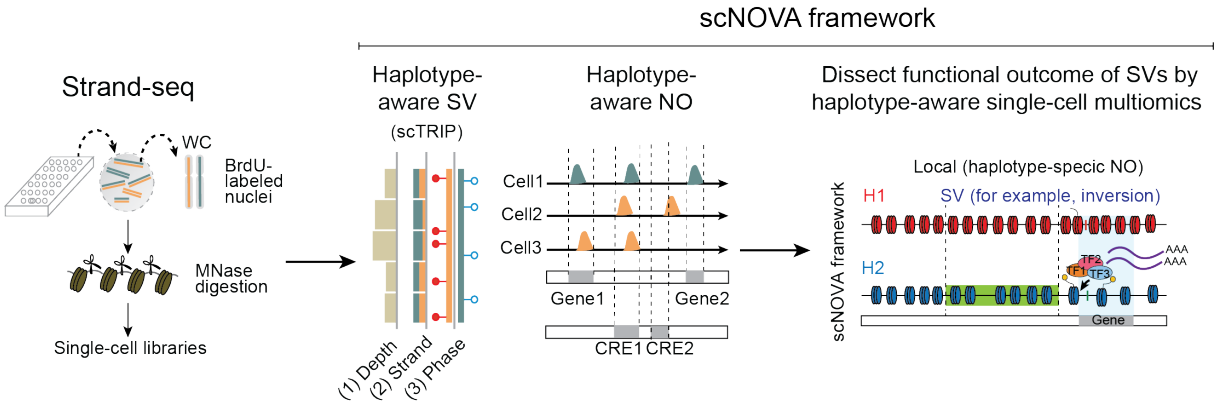


Figure 1.6. Overview of the scNOVA framework. scNOVA leverages data from Strand-seq in two ways: 1) by performing haplotype-aware SV calling using scTRIP and 2) by measuring nucleosome occupancy and indirectly inferring cis-regulatory element accessibility. Together this enables dissecting the functional outcomes of SVs in a haplotype-aware manner. Abbreviations: BrdU, Bromodeoxyuridine; W, Watson; C, Crick; SV, structural variant; scTRIP, single-cell tri-channel processing; CRE, *cis*-regulatory element; NO, nucleosome occupancy. Adapted from Jeong, Grimes [98].

1.4.3 Cellular Indexing of Transcriptomes and Epitopes by sequencing (CITE-seq)

Single-cell RNA sequencing approaches have been instrumental for deconstructing heterogeneous cell populations and inferring gene expression relationships[103,104]. In addition to measuring genome-wide gene expression in single cells, more detailed characterization and discovery of cellular phenotypes and cell states can be achieved by simultaneous indexing of cellular epitopes[105]. With the use of oligonucleotide-labeled antibodies, CITE-seq enables the integration of cell surface protein and transcriptome measurements into an efficient, single-cell readout[105]. The antibody conjugated oligonucleotides are captured by oligo-dT primers and the workflow can be integrated with existing single-cell sequencing approaches, including the commercially available single-cell platform 10x Genomics[105].

Multimodal analyses can be used to explore how multiple cellular modalities affect cellular state and function[106]. CITE-seq technology has been utilized in multiple studies, including detailed characterization and generation of a multimodal "atlas" of human mononuclear cells[105,107], analysis of regulatory programs in cancer[108] and protein velocity studies[109]. Recently, the CITE-seq method has further been expanded to include additional features, including CRIPSR-Cas9 perturbations[110,111], providing even richer phenotypic readouts and ways to study healthy systems as well as cancer.

1.5 Apoptosis

The word apoptosis is derived from the ancient Greek word *apoptosis,* meaning "falling off'. It is the process of genetic elimination of cells, also known as the programmed cell death[112]. Apoptosis occurs during early development and aging, and as part of normal tissue homeostasis to remove unwanted, worn-out, or damaged cells[112,113]. Apoptosis is a complex molecular process involving a cascade of events involving two main pathways, namely the extrinsic and the intrinsic pathways.

The extrinsic signaling pathways or the death receptor pathways initiate apoptosis via an apoptotic signaling molecule, such as Fas-ligand (FasL), that binds to its death receptor[114,115]. Death receptors involve members of the transmembrane tumor necrosis factor receptor gene superfamily that transmit death signals from the cell surface to the intracellular signaling pathways[114,116]. The most well-characterized death receptor-ligand pairs, include FasR/FasL and TNF-α/TNFR1[114,115]. The tumor necrosis factor receptors contain a death domain in the intracellular part of the transmembrane protein that is crucial for transmitting the death signal[114,115]. The death domain forms the death-inducing signaling complex that initiates the caspase cascade events and subsequently leads to degradation of cellular material in preparation for phagocytic cell clearance[113,115,117].

In contrast, the intrinsic signaling pathway or the mitochondrial pathway involves non-receptor-mediated stimuli, such as DNA damage or oxidative stress, that act directly on targets within the cell and are initiated by mitochondrial events[115]. The B-cell lymphoma 2 (BCL-2) protein family controls the activation of the intrinsic pathway. BCL-2 family consists of members that either promote or inhibit apoptosis and govern the mitochondrial membrane permeability and thus regulate the release of cytochrome c from the mitochondria (Figure 1.7)[113,118]. To induce apoptosis, pro-apoptotic multidomain proteins, such as BAX, and BAK, must be activated by the BH3-only proteins, such as BID, BAD, BIM, and PUMA[113,119]. 'Direct activators' directly activate BAX or BAK whereas 'indirect activators' or 'sensitizers' sequester the anti-apoptotic proteins, including BCL-2, MCL-1, and BCL-xL[119]. Similar to BH3-only proteins, p53 can also activate BAX directly demonstrating the direct apoptogenic role of p53 at the mitochondria[120,121]. Upon activation, BAX and BAK undergo conformational changes leading to homodimer formation and sequential higher-order oligomerization at the outer mitochondrial membrane[115,118]. This oligomerization ultimately results in the formation of pores that alter the mitochondrial permeability causing mitochondrial outer membrane permeabilization (MOMP),

also considered the "point of no return"[115,118]. MOMP results in the release of numerous proteins from the mitochondria to the cytosol, including cytochrome c that acts as a cofactor for the apoptotic protease-activating factor-1 (APAF1)[115,119]. Together these proteins trigger the formation of the apoptosome that activates the initiator and executioner caspases that further prepare the cell for phagocytosis (Figure 1.7)[115,119].

To maintain an equilibrium between cell death and cell survival in normal tissue, apoptosis is kept at a crucial balance. When this equilibrium is lost and apoptosis prevented, cells can undergo uncontrolled cell division[113]. Cancer is an example where regulation of apoptosis has failed resulting in the over-proliferation of cells. Transformation of a healthy cell to a malignant cancerous cell is accompanied by DNA damage and aberrant growth signals that stimulate the intrinsic signaling pathway[113,115]. The resistance to apoptosis can be acquired by an increased expression of anti-apoptotic proteins, including BCL-2, MCL-1, and BCL-xL. Indeed, many cancers overexpress and show dependency to at least one anti-apoptotic protein[122-124]. Promoting apoptosis via BH3-mimetics that target anti-apoptotic proteins has become a promising avenue of drug development with several inhibitors showing good efficacy in several cancer entities, including AML[58,59] (see 1.2.3).

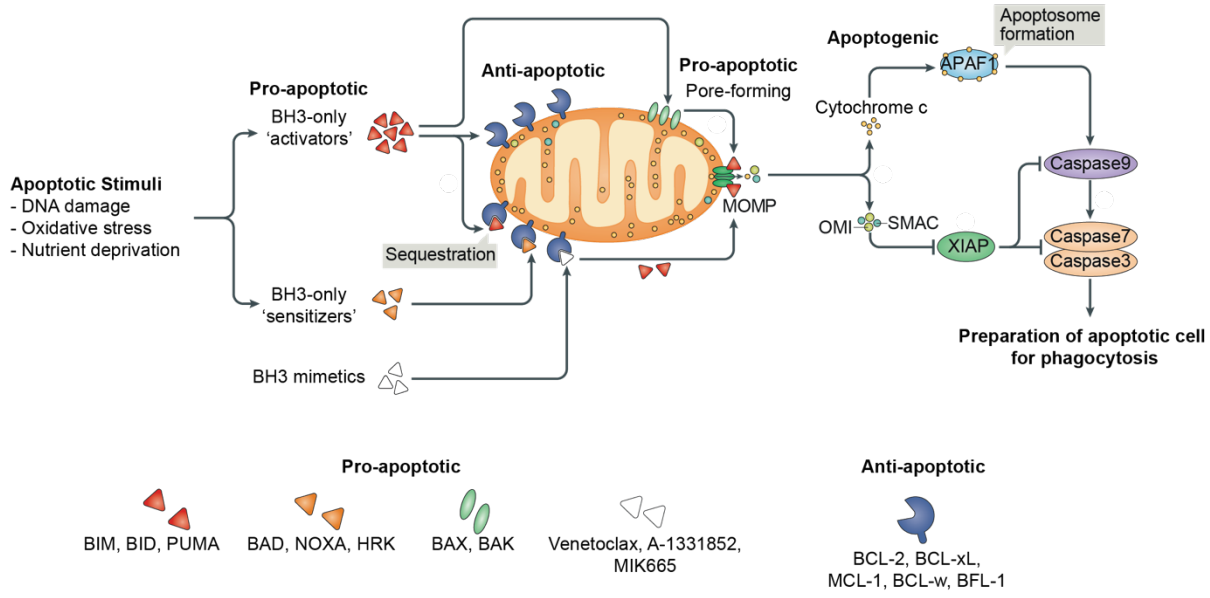


Figure 1.7. The intrinsic or mitochondrial apoptosis pathway. Cellular stress or damage signals function as apoptotic stimuli to initiate the intrinsic apoptosis pathway. Pro-apoptotic BH3-family proteins or BH3 mimetics promote apoptosis by directly activating the pore-forming pro-apoptotic proteins (e.g. BAX and BAK) or by sequestering the anti-apoptotic proteins (e.g. BCL-2, MCL-1 and BCL-xL). Upon activation, the pore-forming BH3-family proteins undergo oligomerization which results in the formation of pores at the mitochondrial outer membrane. Pore formation causes MOMP which results in release of cytochrome c from mitochondria.

Cytochrome c binds APAF1 to form the apoptosome. This activates the initiator and executioner caspases that prepare the cell for phagocytosis. Abbreviations: MOMP, Mitochondrial outer membrane permeabilization; APAF1, apoptotic protease-activating factor-1. Adapted from Singh, Letai [113].

Chapter 2 – Motivation and Outline of the Thesis

At the start of my PhD, the field of cancer research had started to benefit from new treatments and technologies that offered exciting prospects for cancer care. The emergence of different single-cell technologies and computational frameworks enabled the assessment of different levels of intra- as well as inter-patient heterogeneity. Meanwhile, countless new therapeutic modalities had also emerged and were on the horizon. These ranged from immunotherapies to targeting mutation-specific dependencies, as described earlier (see 1.2.3).

However, the implementation of single-cell technologies for clinically relevant and translatable data was lacking. This included the connection between cells' genotype and phenotype and the resulting functional properties, as well as the assessment of changes during disease progression. Simultaneously, implementation of new therapeutics into routine clinical care was and remains slow. Predictive biomarkers that rely on genomic sequencing have enabled most of the recent advances in terms of biomarker-guided therapeutic decision-making but accurate biomarkers relying on globally accessible and affordable screening are largely lacking. These are some of the gaps that I wanted to address in this thesis.

In **Chapter 3**, I describe an integrated single-cell multi-omics framework that combines the analysis of SVs and nucleosome occupancy profiling (scNOVA) with concurrent immunophenotypic and transcriptomic profiling (CITE-seq). I describe how the integrated single-cell analytical framework can be applied on primary CK-AML patient samples to gain valuable information of the complex genomic processes that drive ongoing genomic heterogeneity. Moreover, I describe how following clonal dynamics longitudinally in patient-derived xenografts (PDXs) is of translational relevance and offers promising avenues for identifying and targeting the disease-driving LSCs.

In **Chapter 4**, I introduce a novel strategy to predict treatment response to the BCL-2 inhibitor venetoclax in combination with hypomethylating agent azacytidine. I describe how disease-driving LSCs are the primary target of venetoclax-based therapy and how their elimination determine disease outcome. Furthermore, I show that combining the protein expression of BCL-2 family members into a combinatorial score in the disease-driving LSCs by a flow cytometry-based assay enables the prediction of clinical response to venetoclax-based therapy

Overall, this thesis describes the power of using different approaches to study the disease-driving cells in AML. It shows that single-cell technologies can be utilized to deconvolve different levels of complexity and produce clinically relevant data in complex cancer entities, such as CK-AML. Moreover, the thesis offers clinically relevant alternatives to single-cell technologies for predicting clinical response, highlighting that accessible and affordable techniques can provide easily translatable data. Together the thesis emphasizes the importance of better identifying and characterizing the disease-driving LSCs to improve our understanding of AML as a dynamic disease.

Chapter 3 – Dissection of intra-patient heterogeneity in CK-AML patients with ongoing karyotypic instability identifies targetable relapse-driving leukemic stem cell clones

3.1 Introduction and Motivation I

Genetic tumor heterogeneity and tumor evolution propagated by genomic instability has been well recorded in different cancer entities using longitudinal sequencing of cancer exomes and genomes[125]. These studies have pioneered the concept of distinct evolutionary patterns that can change with shifts in selective pressure[126]. As discussed in **Section 1.3.2**, some tumors exhibit evidence of punctuated evolution that posits rapid bursts of adaptive evolution that results in complex sets of rearrangements[30,126-128]. Whole-genome doubling, chromosomal chromoplexy and chromothripsis represent examples of punctuated evolution in which a single catastrophic event can drive tumor evolution with 'hopeful monsters' representing the most extreme cases[81,126,129,130].

Large and complex chromosomal rearrangements driven by genomic instability are associated with a variety of aggressive cancers. They are usually characterized by metastasis, resistance to therapy, and poor outcome[30,81,131-133]. As discussed in **Section 1.4**, different single-cell technologies have emerged as powerful tools to study such rearrangements[88,89,93]. These technologies provide the basis to study the extent and nature of chromosomal rearrangements in single cells. While these studies have shaped our understanding of SVs in cancer, the connection between cells' genotype and phenotype and their functional properties remain largely unexplored.

To address this, I used a genetically complex AML subgroup, CK-AML, to dissect intra-patient heterogeneity at the genetic, epigenetic, transcriptional, and cell surface proteome levels. As discussed in **Section 1.2.1**, CK-AML comprises 10–12% of all AML patients and is the second-largest cytogenetic subgroup in AML. It is commonly associated with *TP53* mutations, complex genomic rearrangements including chromothripsis[30] and poor overall survival rates[25,31,32]. The adverse prognosis is due to an aggressive, fast-progressing disease, with high rates of refractoriness and relapse following standard induction therapy[28]. The molecular and cellular mechanisms causing this behavior remain mostly unclear, but the frequent loss of p53 function and the high subclonal heterogeneity is considered to contribute to the typical therapeutic failure and dismal outcome[31,134]. Despite major clinical need, CK-AML remains insufficiently studied,

in part due to the limited technical capabilities in analyzing cellular heterogeneity of this level of complexity.

In this chapter, I describe an integrated single-cell multi-omics approach that combines the analysis of SVs and nucleosome occupancy profiling (scNOVA) with concurrent immunophenotypic and transcriptomic profiling (CITE-seq). I describe how the integrated single-cell analytical framework can be applied to primary CK-AML patient samples to gain valuable information of the complex genomic processes that drive ongoing genomic heterogeneity. Moreover, I describe how following clonal dynamics longitudinally is of translational relevance and offers promising avenues for precision medicine approaches.

The results detailed in this section were part of a close collaboration with Dr. Jan Korbel's group, in particular with Dr. Karen Grimes and Dr. Hyobin Jeong. The generation and analysis of the scNOVA data was done as a joint effort and is also stated as such in the text. The contents of this chapter are based on a manuscript that is currently in preparation.

"Dissection of intra-patient heterogeneity in CK-AML patients with ongoing karyotypic instability identifies targetable relapse-driving leukemic stem cell clones"
Aino-Maija Leppä[*], Karen Grimes[*], Hyobin Jeong[*], Tobias Boch, Alexander Waclawiczek, Darja Karpova, Florian Grünschläger, Anna Jauch, Anna Dolnik, Vera Thiel, Bernardo Rodriguez-Martin, Lars Bullinger, Alwin Krämer, Ashley D. Sanders[#], Jan O. Korbel[#], Andreas Trumpp[#]

[*,#]Contributed equally

3.2 Results I: Exploring genetic and non-genetic changes driving tumor evolution in CK-AML at single-cell resolution

3.2.1 Multiomic analysis of genetic and non-genetic landscapes of single CK-AML cells

CK-AML is marked by complex genomic rearrangements, but the disposition and extent of the rearrangements in single cells remains elusive and hampers the assessment of cellular heterogeneity. To gain insight into the evolution of genomic rearrangements and the resulting phenotypical heterogeneity, I together with Dr. Karen Grimes and Dr. Hyobin Jeong established a single-cell framework to study SVs together with non-genetic properties in CK-AML. We integrated scNOVA[98] and CITE-seq[107] frameworks (see 1.4) for parallel construction of single-cell SV landscapes along with epigenetic, transcriptomic and cell surface protein maps of CK-AML, and termed it scNOVA-CITE (Figure 3.1). I together with Dr. Karen Grimes generated Strand-seq[90] libraries from bone marrow or peripheral blood cells from four primary CK-AML patient samples, and applied the scNOVA framework on the libraries (Appendix Table 1). The Strand-seq libraries contained a median of 195,225 mapped nonduplicate read pairs[96], which amounts to ~0.0094x coverage per cell and cumulative coverage of ~0.73x per patient (Appendix Figure 1).

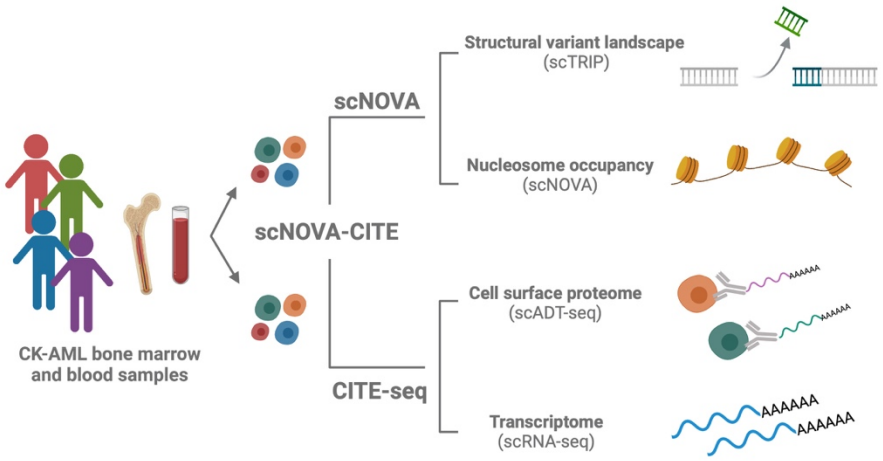


Figure 3.1 Overview of scNOVA-CITE workflow. Schematic of the single-cell multi-omics framework scNOVA-CITE, applied on four primary CK-AML patient samples collected at diagnosis. scNOVA was used to assess structural variant landscapes and nucleosome occupancy. CITE-seq was applied to assess transcriptome and cell surface proteome.

First using scNOVA, Dr. Karen Grimes and I dissected the SV landscapes of the CK-AML patient samples. All samples carried multiple somatic SVs, including losses and gains of terminal chromosome regions, whole-chromosome aneuploidies, and complex sets of genetic

rearrangements (Figure 3.2A-C). Six to ten chromosomes harbored at least one large genomic rearrangement (> 2Mb) in each patient which was evident in the majority of the cells (Figure 3.2A). Only chromosome 10 did not show detectable aberrations in at least one patient when SVs present only in single cells were excluded. Samples CK295 and CK397 mostly harbored clonal SVs present in almost all cells, with very few cells carrying additional or diverging SVs. They represented a low intra-patient heterogeneity group. In contrast, CK282 and CK349 showed vast levels of intra-patient heterogeneity and harbored several competing subclones with distinct SV profiles, representing a high intra-patient heterogeneity group. Notably, several of the rearranged chromosomes harbored sets of complex rearrangements, irrespective of the level of intra-patient heterogeneity. In CK282, the copy-number profiles of chromosomes 12 and 17 oscillated between three states and displayed islands of gain and loss of material (Figure 3.2B, Appendix Figure 2A). This was suggestive of chromothripsis, a mutational process with a one-off catastrophic event[30,81,82]. This mutational process was also present in a subset of cells in CK295 affecting chromosome 20 whereas CK349 showed ongoing chromosome instability at chromosome 13 resulting in chromothripsis-like rearrangements in a subset of cells (Figure 3.2C, Appendix Figure 2B). Using optical genome mapping and multiplex fluorescence *in situ* hybridization (M-FISH) analysis, I was able to show that segments from the complex rearrangements joined and formed large contiguous DNA structures involving multiple chromosomes giving rise to complex derivative and marker chromosomes (Figure 3.2D, Appendix Figure 2C-D).

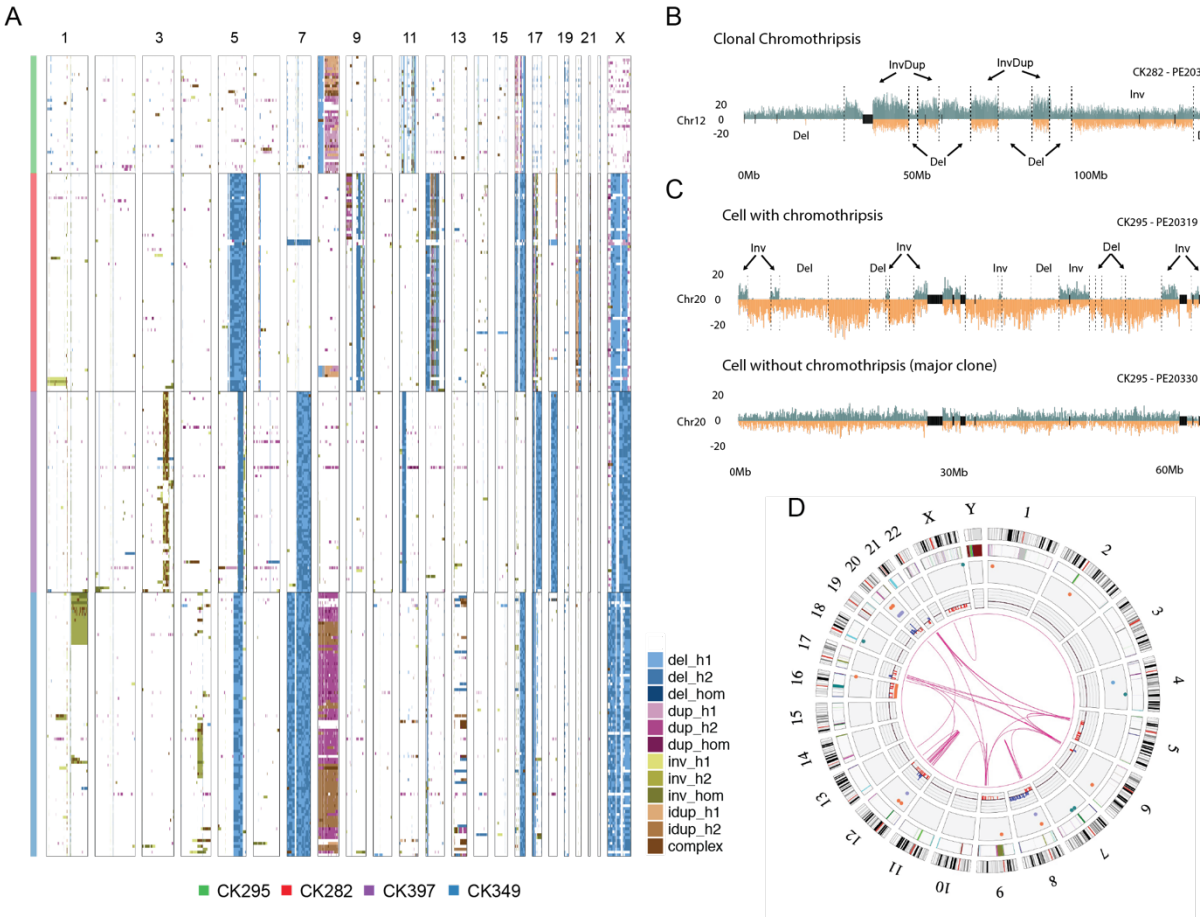


Figure 3.2 Genetic single-cell landscape of primary CK-AML patient samples. A Heatmap of SVs from 278 single cells arranged using Ward's method for hierarchical clustering of SV genotype likelihoods in four patient samples. **B** Strand-specific read depth of a representative single-cell from CK282 showing clustered deletions, inverted duplications and inversions along a single homolog at chromosome 12, resulting from clonal chromothripsis. Reads denoting somatic SVs, discovered using scTRIP, mapped to the Watson (orange) or Crick (green) strand. Grey: single-cell IDs. **C** Strand-specific read depth of an example CK295 cell showing clustered deletions and inversions along a single homolog at chromosome 20, resulting from chromothripsis (top), compared to a cell without chromothripsis (bottom). **D** Circos plot illustrating complex rearrangements and translocations involving multiple chromosomes, assessed from optical genome maps from the patient-derived-xenograft of CK282. Chromosomes (outside of the circular plot) and chromosomal rearrangements are shown as arcs connecting the two relevant genomic regions in the middle. Abbreviations: Chr, Chromosome; Del, Deletion; Dup, Duplication; Inv, Inversion; idup, Inverted duplication; hom, Homozygous; h, Haplotype; CNV, Copy number variation. Experiments performed jointly by Dr. Karen Grimes and myself.

Next, I constructed independent immunophenotypic and transcriptomic maps of the same CK-AML patient samples using CITE-seq. I simultaneously generated 10x Genomics 3′ single-cell RNA sequencing[135] (scRNA-seq) and antibody-derived tag sequencing[105] (scADT-seq) libraries from 14,941 AML cells (Methods Table 6.1, Appendix Figure 3). To mitigate patient-driven batch effects, I projected cells from each patient into a shared embedding using

Harmony[136], and visualized the data using uniform manifold approximation and projection (UMAP) (Figure 3.3A-B). I identified clusters expressing markers for myeloid differentiation (*LYZ*, *S100A8,* CD11b and CD64), mostly encompassing cells from CK295, and several clusters enriched for more primitive markers (*HOPX*, *FAM30A*, CD34 and GPR56), mostly encompassing cells from the other three samples (Figure 3.3C-D). These findings were recapitulated by single-cell nucleosome occupancy profiling, performed by Dr. Hyobin Jeong, where CK295 primarily consisted of epigenetically more mature granulocyte-monocyte progenitor (GMP)-like cells while the other samples comprised to a greater extent of more primitive cell states along the myeloid differentiation trajectory (Figure 3.3E). In both assays sample-specific differences were most prominent with cell clusters largely driven by individual patients, suggesting a strong phenotypic heterogeneity across the patients.

Collectively, these data highlight that CK-AML is a genetically heterogeneous disease subentity characterized by multiple SVs and ongoing accumulation of punctuated genomic complexity that is associated with largely primitive phenotypic landscapes.

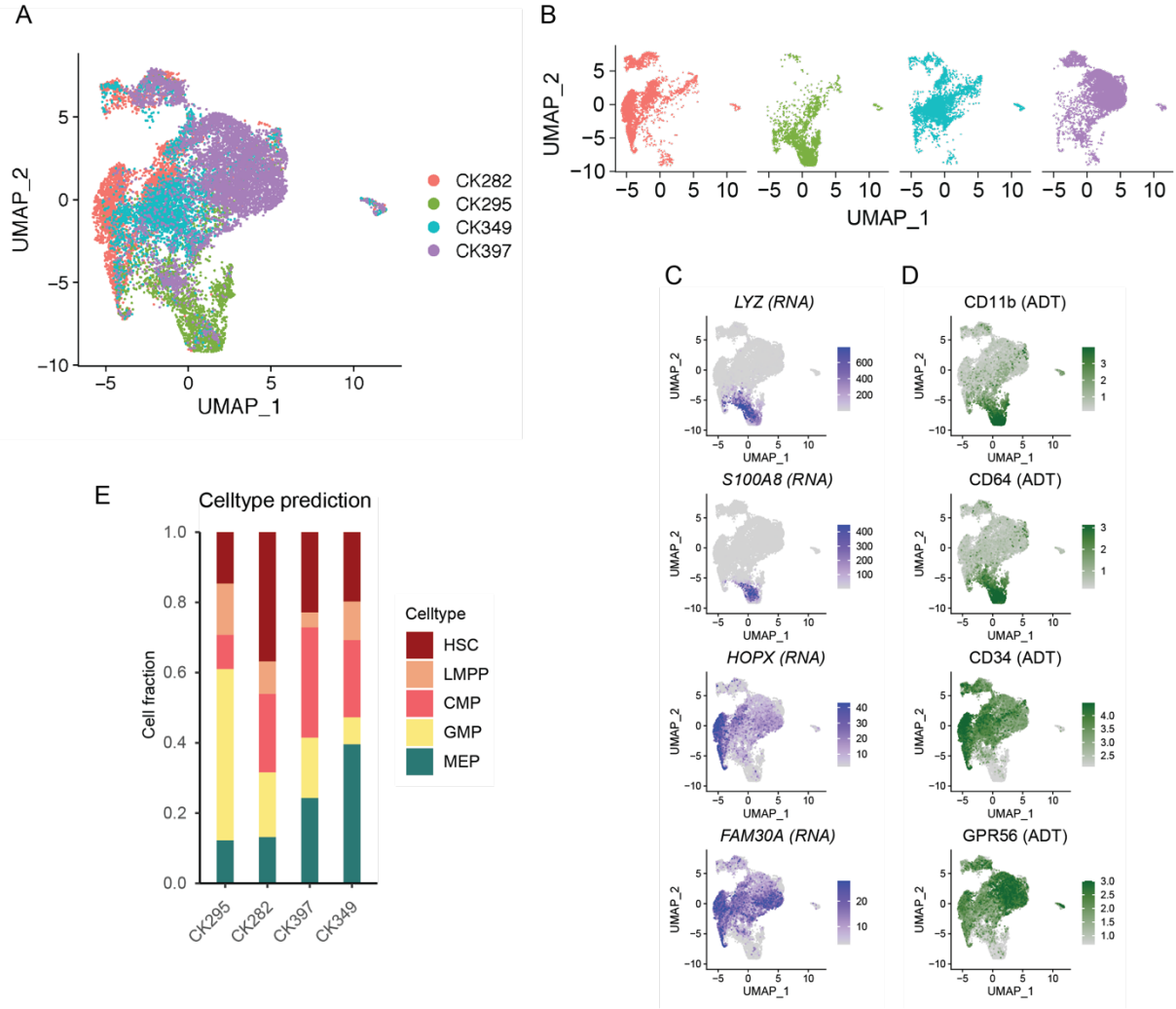


Figure 3.3 Phenotypic single-cell landscape of primary CK-AML patient samples. A-B UMAP plots of the CITE-seq data form the four CK-AML patient samples projected into a shared embedding using Harmony with cells colored by patient. In **A**, all patient samples are shown together and in **B**, separated by patient. **C** Gene expression of myeloid differentiation markers (*LYZ*, *S100A8*) and primitive markers (*HOPX*, *FAM30A*) colored as normalized and variance-stabilized counts. **D** Cell surface marker expression of myeloid differentiation markers (CD11b, CD64) and primitive markers (CD34, GPR56) colored as centered log ratio-transformed counts. **E** Stacked bar plots showing the fraction of indicated hematopoietic stem and progenitor cell-like states in the four CK-AMLs based on single-cell nucleosome occupancy-profiling. Abbreviations: HSC, Hematopoietic stem cell; LMPP, Lymphoid primed multipotent progenitor; CMP, Common myeloid progenitor; GMP, Granulocyte-monocyte progenitor; MEP, Megakaryocyte-erythroid progenitor; ADT, Antibody-derived tag. Analysis of nucleosome occupancy data in E performed by Dr. Hyobin Jeong.

3.2.2 Haplotype aware dissection of functional outcomes in patient with 3q-rearranged CK-AML

Due to the large genomic and phenotypic heterogeneity observed across the patients, my next aim was to investigate different layers of intra-patient heterogeneity by focusing on each leukemic sample individually. First, focusing on the diagnosis sample CK397, I together with

Dr. Karen Grimes prepared good quality Strand-seq and CITE-seq libraries from 70 and 5545 cells, respectively. We observed monosomies of chromosomes 17 and 18 along with larger deletions affecting almost the entire q arm of chromosomes 5 and 7 as well as the majority of chromosome X. Chromosome 5 also showed heterozygous inversions on both sides of the observed deletion and chromosome X an amplification at the p arm (Figure 3.2). Furthermore, we detected five focal deletions three of which affected known tumor suppressor genes or genes reported to play a role in AML tumorigenesis, including *FOXO3*, *DNMT3A*, and *TM2D1*[137-139]. Last, we observed a complex set of rearrangements involving several inversions and deletions at the q arm of chromosome 3. As scNOVA enables haplotype-aware readout, we were able to resolve that both haplotypes were involved in these complex rearrangements at chromosome 3 (Figure 3.4A). The majority of cells supported the karyotype of the major clone, with only a few individual cells showing karyotypic differences.

Several DNA segments affected by the SVs did not segregate with the respective chromosomes they originated from, indicating inter-chromosomal SV formation. We searched for co-segregation footprints[93] and identified an unbalanced translocation involving 7q and 21q, consistent with the t(7;21)(q11.2;q11.2) derivative chromosome detected in clinical karyograms (Appendix Table 1-Appendix Table 2). Detailed analysis of the complex 3q arm revealed that fragments from one 3q haplotype were involved in inter-chromosomal SV formation with 5q while the other 3q haplotype contained a complex intra-chromosomal multi-inversion rearrangement (Figure 3.4A). High-resolution construction of the 3q arm by optical genome mapping further supported these findings and revealed an even more complicated network of rearrangements. Multiple segments joined across the 3q arm and generated, *inter alia*, the recurrent oncogenic *RPN1*/*MECOM* fusion (Figure 3.4A)[140-142]. To further support the genomic data, I generated bulk RNA-seq data from the same sample and was able to identify fusion transcripts between *MECOM* (located at 169.4 Mb) and the inter-genic region of *RPN1* (located at 128.6 Mb) (Figure 3.4B).

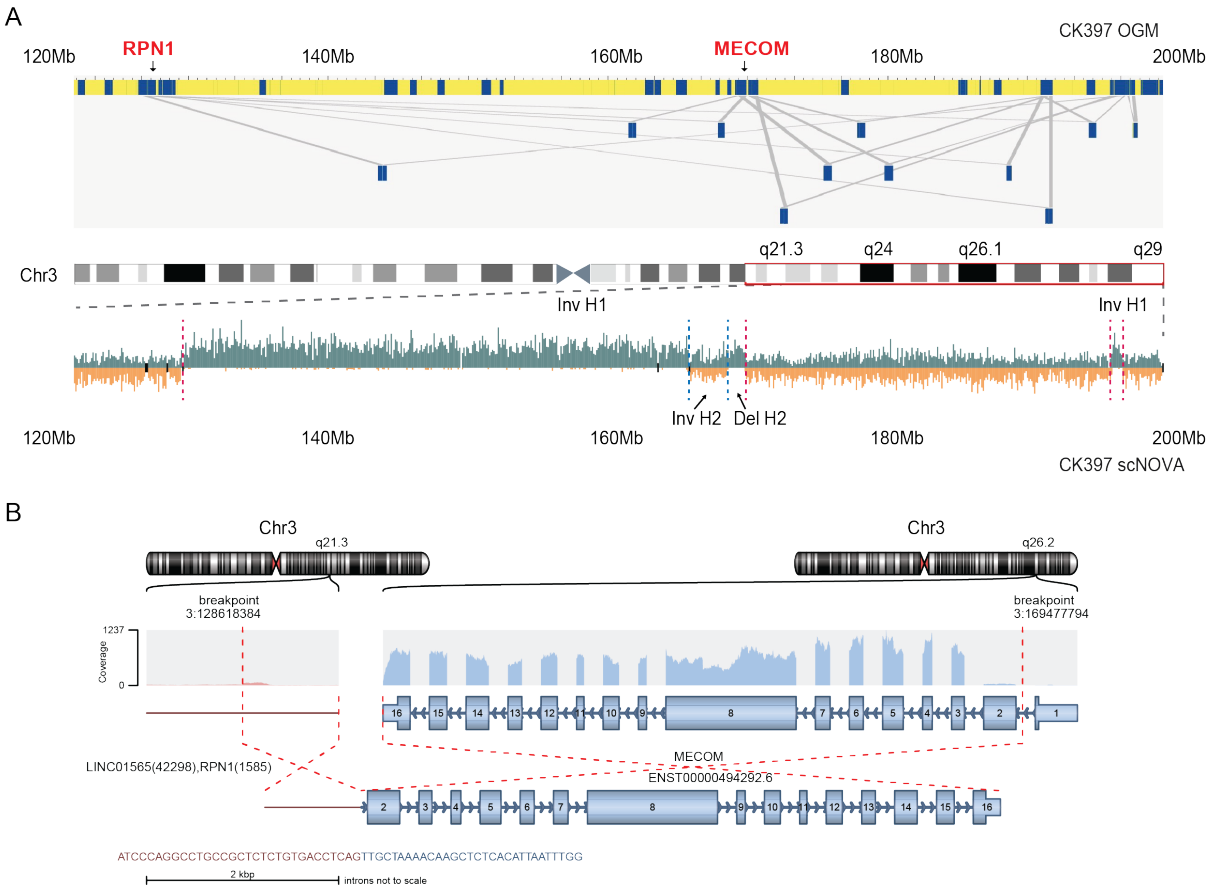


Figure 3.4 Multi-inversion rearrangement at 3q. A Chromosome view of 3q with mapping of segments by optical genome mapping (top) and scNOVA (bottom) showing multiple inversions spanning parts of the q arm. In optical genome mapping, *de novo* genome maps (blue with grey lines) are aligned to the reference genome (yellow). In scNOVA, reads indicate somatic SVs, discovered using scTRIP, mapped to the Watson (orange) or Crick (green) strand. **B** Fusion transcripts between *MECOM* and the intergenic region of *RPN1* confirming the intra-chromosomal rearrangement at 3q. Shown are the fusion partners, their orientation, the retained exons in the fusion transcript and the expression of the exons. Abbreviations: Chr, Chromosome; Del, Deletion; Inv, Inversion; OGM, Optical genome mapping; H, Haplotype.

Previous studies in AML have reported a range of 3q rearrangements that result in *MECOM* overexpression and are driven by different enhancer hijacking events, such as *GATA2* enhancer repositioning[140-142]. I reasoned that if the *MECOM* dysregulation observed in CK397 was similarly driven by the multi-inversion rearrangements generating the oncogenic *RPN1*/*MECOM* fusion, then *MECOM* overexpression should be restricted to the single haplotype affected by the oncogenic fusion. Nucleosome occupancy analysis performed by Dr. Hyobin Jeong supported this hypothesis. It revealed haplotype-specific epigenetic dysregulation at the *MECOM* gene body with the haplotype harboring the *RPN1*/*MECOM* fusion showing lower nucleosome occupancy (Figure 3.5A). Using bulk and single-cell RNA-seq, I detected high gene expression of the *MECOM* oncogene uniquely in CK397. This further highlighted the strong epigenetic and transcriptomic dysregulation of *MECOM*, very likely

driven by the complex multi-inversion rearrangement (Figure 3.5B-C). As *MECOM* expression has been shown to promote expansion of HSPCs and to induce subsequent megakaryopoiesis[143,144], I further hypothesized that the haplotype-specific nucleosome occupancy observed at *MECOM* gene body would show cell type-specific biases. Dr. Hyobin Jeong made use of the MNase-seq reference dataset from healthy sorted CD34+ bone marrow cells to assign each CK-AML cell to its corresponding normal hematopoietic differentiation state (Grimes et al., in preparation), and assessed the nucleosome occupancy of the assigned cell types. Intriguingly, scNOVA showed haplotype-specific nucleosome occupancy imbalance uniquely in the HSC-like and megakaryocyte-erythroid progenitor (MEP)-like cells in CK397 (Figure 3.5D). This finding offers a molecular rational for the megakaryocytic phenotypes often observed in patients with 3q-rearranged AML[145]. I confirmed the cell type-specific dysregulation of *MECOM* also on a transcriptional level where HSC-, lymphoid-primed multipotent progenitor (LMPP)-like cells and megakaryocytic progenitor (Prog Mk)-like cells showed highest expression of *MECOM* (Figure 3.5E). Taken together, these data underscore the ability of scNOVA-CITE to characterize haplotype-specific cancer-related changes, contributing to aberrant and cell type-specific epigenetic and transcriptomic oncogene dysregulation.

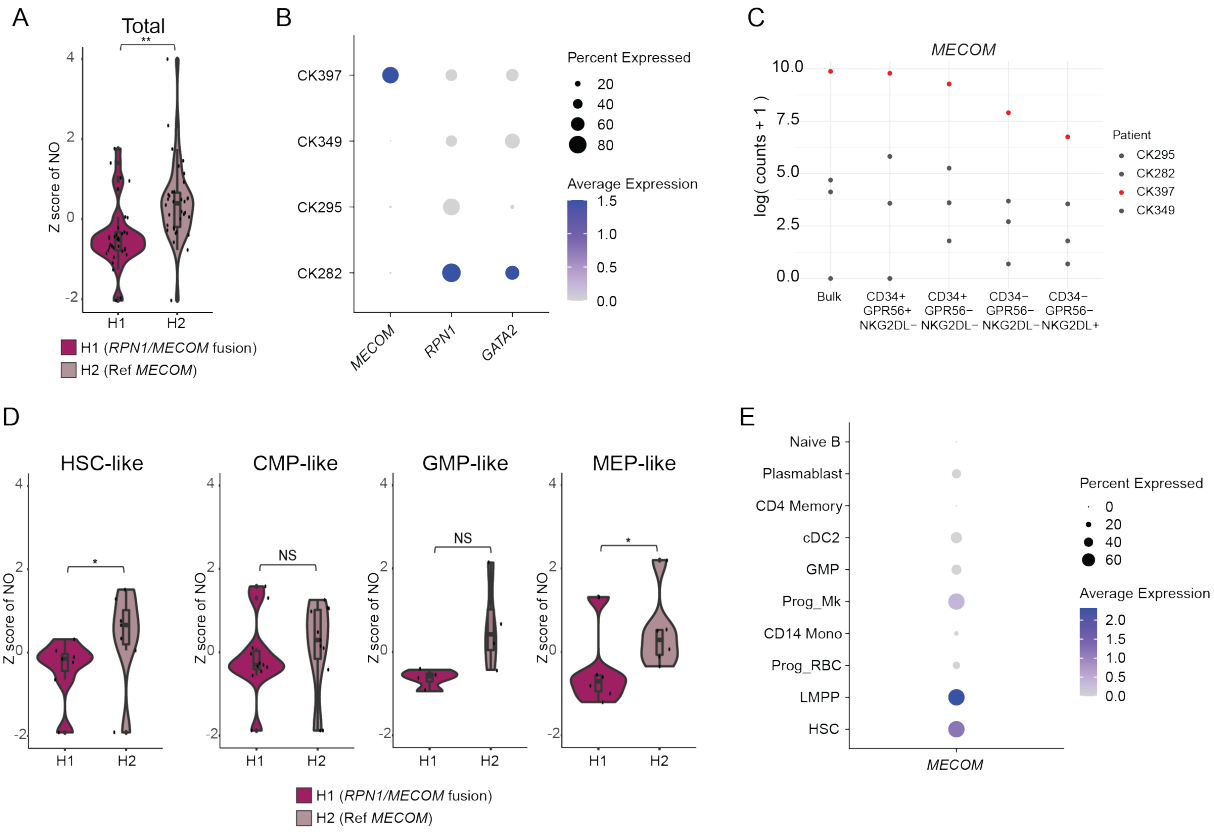


Figure 3.5 Nucleosome occupancy and gene expression at *MECOM* in 3q-rearranged AML. A Violin plot showing haplotype-specific nucleosome occupancy at the *MECOM* gene body (10% FDR) for CK397. H1 contains the multi-inversion rearrangement resulting in *RPN1*/*MECOM* fusion whereas H2 is normal at *MECOM* locus. **B** Dot plot of the single cell gene expression of selected genes known to be affected by 3q rearrangements shown in the four CK-AMLs. Dot size indicates the percentage of cells of a sample expressing the given gene with average expression shown as normalized and variance-stabilized counts. **C** Scatter plot of *MECOM* expression in bulk AML cells and FACS-sorted populations shown in the four CK-AMLs. Expression in CK397 is colored in red. **D** Violin plot showing haplotype-specific nucleosome occupancy at the *MECOM* gene body (10% FDR) for CK397 in cells of the indicated hematopoietic cell-like states based on nucleosome occupancy. **E** Dot plot of *MECOM* gene expression in the indicated hematopoietic cell-like states based on gene expression in single cells. For gene-body nucleosome occupancy measurements both haplotypes were converted into $\log_2$-scale and compared using Wilcoxon ranksum test followed by Benjamini-Hochberg multiple correction. Abbreviations: H, Haplotype; NO, Nucleosome occupancy; HSC, Hematopoietic stem cell; CMP, Common myeloid progenitor; GMP, Granulocyte-monocyte progenitor; MEP, Megakaryocyte-erythrocyte progenitor; LMPP, Lymphoid primed multipotent progenitor; Ref, Reference. Analysis of nucleosome occupancy data performed Dr. Hyobin Jeong.

3.2.3 Ongoing genomic instability drives tumor evolution

To further investigate different layers of intra-patient heterogeneity, I turned to two diagnosis CK-AML samples with high karyotypic heterogeneity: CK282 and CK349. Dr. Karen Grimes and I sequenced 76 and 91 cells from CK282 and CK349, respectively, using Strand-seq. Both samples contained multiple subclones with highly rearranged genomes (Figure 3.6A-B). A set of clonal SVs were present in (almost) all cells affecting ten chromosomes in CK282 and eight chromosomes in CK349 (Figure 3.6A-B). Both samples showed signs of active mutational processes when SVs identified in individual cells across the subclones was used as a proxy for ongoing genomic instability. In CK282, chromosome 20 displayed a classical BFB event[93] consisting of an inverted duplication, and a terminal deletion in 63% of the cells (48/76 cells). The length of the terminal deletion at 20q differed between cells, with a total of 10 different breakpoints present in the sequenced cells (Figure 3.6C-D). In contrast, CK349 harbored deleted segments of different lengths in all cells at chromosome 17. We identified ten different breakpoints at 17q and most of the SVs with the distinct breakpoints were shared only between few cells (Appendix Figure 4).

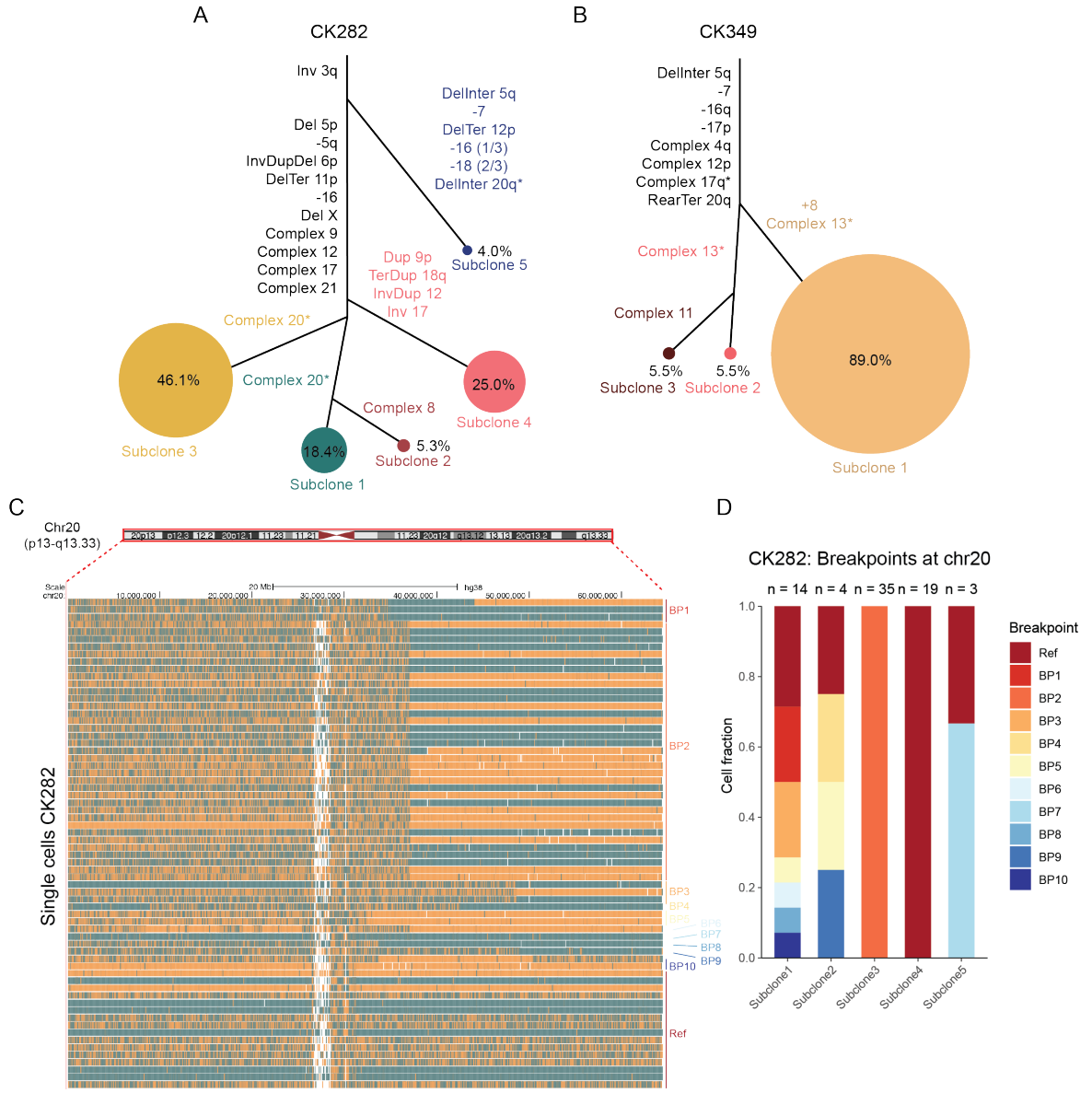


Figure 3.6 Ongoing genomic instability in CK-AML. A-B Manually curated clonal trees showing the hierarchy of somatic SV subclones discovered using scTRIP in **A** CK282 and **B** CK349. Each colored circle represents a subclone, representing genetically-similar cells, with cumulative SVs. These can be traced with solid lines towards the root. The size of the circle is proportional to the clonal population, and the percentage within/next to each circle shows the percentage of each clone among the total leukemic cells (*, differing breakpoints for the complex SV at chromosome 20, see C and D, and chromosome 13, see Appendix Figure 5). **C** Signs of active mutational processes at chromosome 20 in CK282 displayed by varying breakpoints of the terminal deletion at 20q. Reads mapped to the Watson (orange) or Crick (green) strand. **D** Stacked bar plots showing the fraction of indicated breakpoints at chromosome 20 in the different subclones. Number of cells assigned to each subclone is shown above the bars. Abbreviation: Chr, Chromosome; Del, Deletion; Inv, Inversion; Dup, Duplication; Ter, Terminal; Inter, Interstitial; BP, Breakpoint; Ref, Reference. Experiments performed jointly by Dr. Karen Grimes and myself.

In CK282, Dr. Karen Grimes and I further assigned the cells to five genetic subclones based on their SVs (Figure 3.6A). Subclones 1 to 3 were genetically very similar (69.7% cells; 53/76

cells). They shared the clonal 'backbone' of SVs described above and differed only in a set of SVs detected on the unstable chromosome 20 and on chromosome 8 that was rearranged in subclone 2. Subclone 2 (5.3% cells; 4/76 cells) had acquired two inverted duplications, three deletions and one larger inverted duplication at chromosome 8, resulting in five breakpoints spanning the whole chromosome (Figure 3.7A-B). This pattern was suggestive of another BFB event[93] with additional islands of loss in heterozygosity. Subclone 4 (25.0% cells; 19/76 cells) displayed several unique SVs on four different chromosomes in addition to the 'backbone' of SVs shared with subclones 1 to 3. These unique SVs included three duplications and one inversion (Figure 3.7A). Moreover, subclone 4 lacked rearrangements on chromosome 20 detected in the other subclones (Figure 3.7A). In contrast, subclone 5 (3.95% cells; 3/76 cells) differed from the other subclones markedly. Cells in subclone 5 harbored a distinct set of simple SVs and lacked almost entirely complex rearrangements characteristic of subclones 1 to 4 (Figure 3.7A). Finally, we also identified an isolate cell with its own SVs, hinting to even more complex intra-patient heterogeneity than described here. Taken together, these data demonstrate the parallel existence of complex and simpler subclones in a CK-AML with a highly rearranged genome.

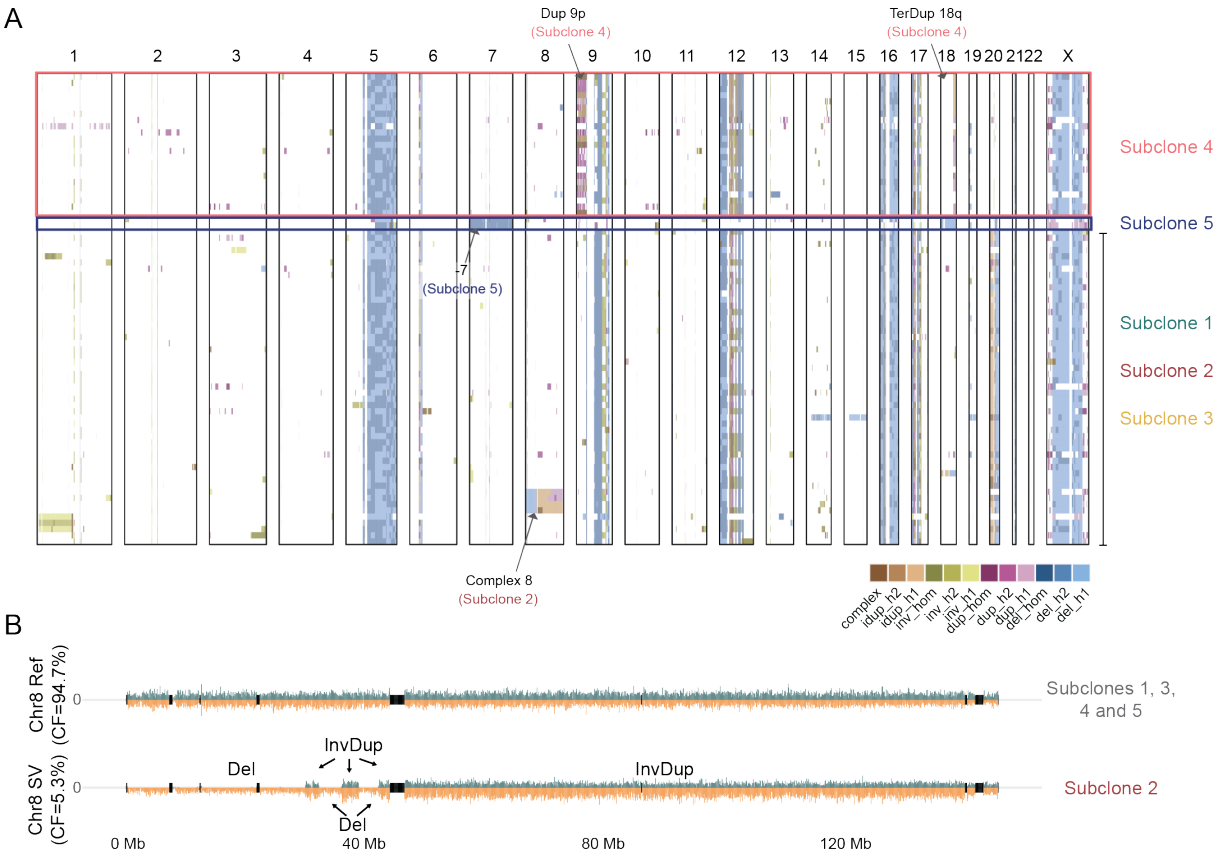


Figure 3.7 Genetic intra-patient heterogeneity in CK282. A Heatmap of SVs from 76 single cells arranged using Ward's method for hierarchical clustering of SV genotype likelihoods in CK282. Assigned subclones are indicated on the right and examples of subclone-specific SVs are shown in the heatmap. **B** Strand-specific read depth of two representative single-cells from CK282 showing a complex genetic rearrangement comprising of two inverted duplications, three deletions and one larger duplication, spanning the whole chromosome 8. Reads denoting somatic SVs, discovered using scTRIP, mapped to the Watson (orange) or Crick (green) strand. Grey and red: subclones affected. Abbreviations: Chr, Chromosome; idup, Inverted duplication; Del, Deletion; Dup, Duplication; Inv, Inversion; Ter, Terminal; CF, Cell fraction; h, Haplotype; hom, Homologous. Experiments performed jointly by Dr. Karen Grimes and myself.

In CK349, Dr. Karen Grimes and I assigned the cells to three subclones based on their SVs. Subclone 1 (89% cells; 81/91 cells) represented the biggest clone and harbored a chromosome 8 trisomy (Figure 3.7B). Chromosome 13 showed signs of active mutational processes in a subset of cells, resulting in somatic gains and losses of terminal chromosome regions (gain: 3/81 cells; loss: 4/81 cells), whole-chromosome aneuploidies (gain 3/81 cells; loss 4/81 cells) and complex rearrangements involving serially acquired SVs (2/81 cells) (Appendix Figure 5). In subclone 2 (5.5% cells; 5/91 cells) and 3 (5.5% cells; 5/91 cells) we identified a distinct set of rearrangements affecting chromosome 13 that differed from the SVs described above (Appendix Figure 5). These consisted of two duplications (10/10 cells) juxtaposed next to a nonamplified (7/10 cells) or deleted region (3/10 cells), spanning the whole chromosome 13 (Appendix Figure 5). Additionally, subclone 3 had acquired a set of SVs at chromosome 11 that displayed wave-like copy-number profiles (Figure 3.8A). The set of SVs at chromosome 11 comprised of multiple rearrangements covering numerous genomic segments that were amplified at distinct levels and interrupted by nonamplified or deleted regions. Closer analysis of chromosome 11 suggested a step-wise acquisition of the amplifications with cell-specific copy-number statuses (Figure 3.8A). I identified cells showing intermediate levels of amplification mostly localized to the q arm as well as cells with vastly higher levels and more spread amplifications spanning both p and q arms (Figure 3.8A). These genomic rearrangement patterns differed from chromothripsis that comprises multiple rearrangements but a limited number of oscillating copy-number states[30,81,82]. They also differed from other common amplifications that mostly harbor few rearrangements with one or few copy-number states[93]. The pattern rather resembled a phenomenon reported in solid cancers termed 'seismic amplification'[85]. It is a multistep mutational process starting with chromothripsis, followed by formation of circular extrachromosomal DNA that subsequently undergo repetitive rounds of circular recombination[85].

I identified two common deletions at the q arm that were shared between all cells with the complex rearrangement: (1) a deletion located between two amplified inverted duplication segments; and (2) a terminal deletion adjacent to an amplified inverted duplication segment (Figure 3.8A). These indicate a fold-back mechanism at 11q as an initiator of an intermediate inverted duplication resulting from a loss of telomeric region followed by circularization. Due to several shared breakpoints, I speculate that the telomeric loss was a result of chromothripsis, but cannot exclude BFB as a potential initial hit. Importantly, M-FISH analysis of the PDX revealed the chromosome 11 amplifications to result in a large ring chromosome containing several copies of segments from 11p and 11q (Figure 3.8B), supporting the model of seismic amplifications. This, to my knowledge, represents the first report and detailed characterization of a seismic amplification event in single cells and further showcases ongoing genomic instability in CK-AML.

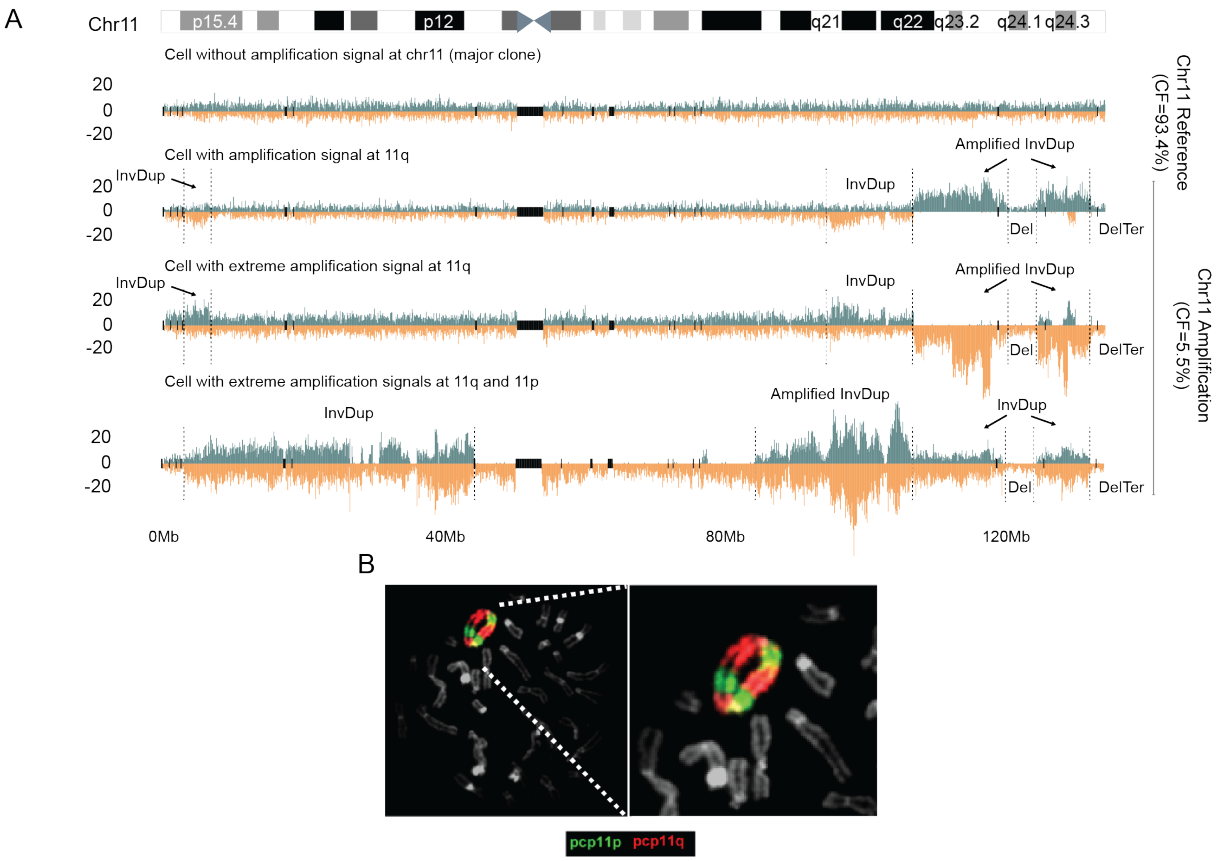


Figure 3.8 Seismic amplification in CK-AML. A Depiction of four representative single cells from CK349 with different amplification statuses at chromosome 11, based on no amplification (upper, cell without seismic amplification), amplifications restricted mostly to q arm (middle) and amplifications spanning p and q arms (bottom). Reads indicate somatic SVs, discovered using scTRIP, mapped to the Watson (orange) or Crick (green) strand. Affected cell fractions are shown on the right. **B** Two-color fluorescence in situ hybridization of ring

chromosome 11 from PDX of CK349 using 11p (green) and 11q (red) partial chromosome painting probes. Abbreviations: Chr, Chromosome; Del, Deletion; Dup, Duplication; Inv, Inversion; Ter, Terminal; CF, Cell fraction; pcp, Partial chromosome paint. Experiments performed jointly by Dr. Karen Grimes and myself.

3.2.4 Genetic heterogeneity drives phenotypic diversity

As CK282 and CK349 showed high levels of intra-patient genetic heterogeneity, I next investigated the extent to which the genetic heterogeneity was associated with different layers of phenotypic heterogeneity. Together with Dr. Hyobin Jeong, I first used scNOVA to explore the consequences of the SVs on the cells' regulome by measuring the genome-wide nucleosome occupancy alongside the SVs in the same cell. This revealed complex relationships among the subclones including differences in cell type compositions and partially overlapping nucleosome occupancy signatures (Figure 3.9A-D, Appendix Figure 6A-D). A total of 525 and 351 genes were epigenetically dysregulated between the subclones in CK282 and CK349, respectively. In CK282 subclones 2 and 5 harbored the most dysregulated genes and in CK349 subclone 3 (Appendix Figure 6A-B). These subclones also differed the most in their cell type composition compared to the other subclones. For instance, subclone 5 in CK282 showed a high abundance of MEP-like cells, suggesting a genetically driven bias towards the erythroid/megakaryocytic lineage, whereas subclone 2 in CK282 and subclone 3 in CK349 showed high abundances of more primitive HSC- and CMP-like cells, indicating a strong differentiation block (Figure 3.9A-B). Pathway analysis based on Hallmark[146] and KEGG[147] gene sets revealed, for instance, that PI3K-AKT-mTOR signaling and cytoskeletal organization-related pathways were dysregulated particularly in subclone 2 in CK282 (Figure 3.9C, Appendix Figure 6C). Cytoskeletal organization-related pathway was also active in subclone 3 in CK349, whereas cell-cycle-related pathways (e.g. E2F Targets and G2-M Checkpoint) were more active in subclone 1 (Figure 3.9D, Appendix Figure 6D).

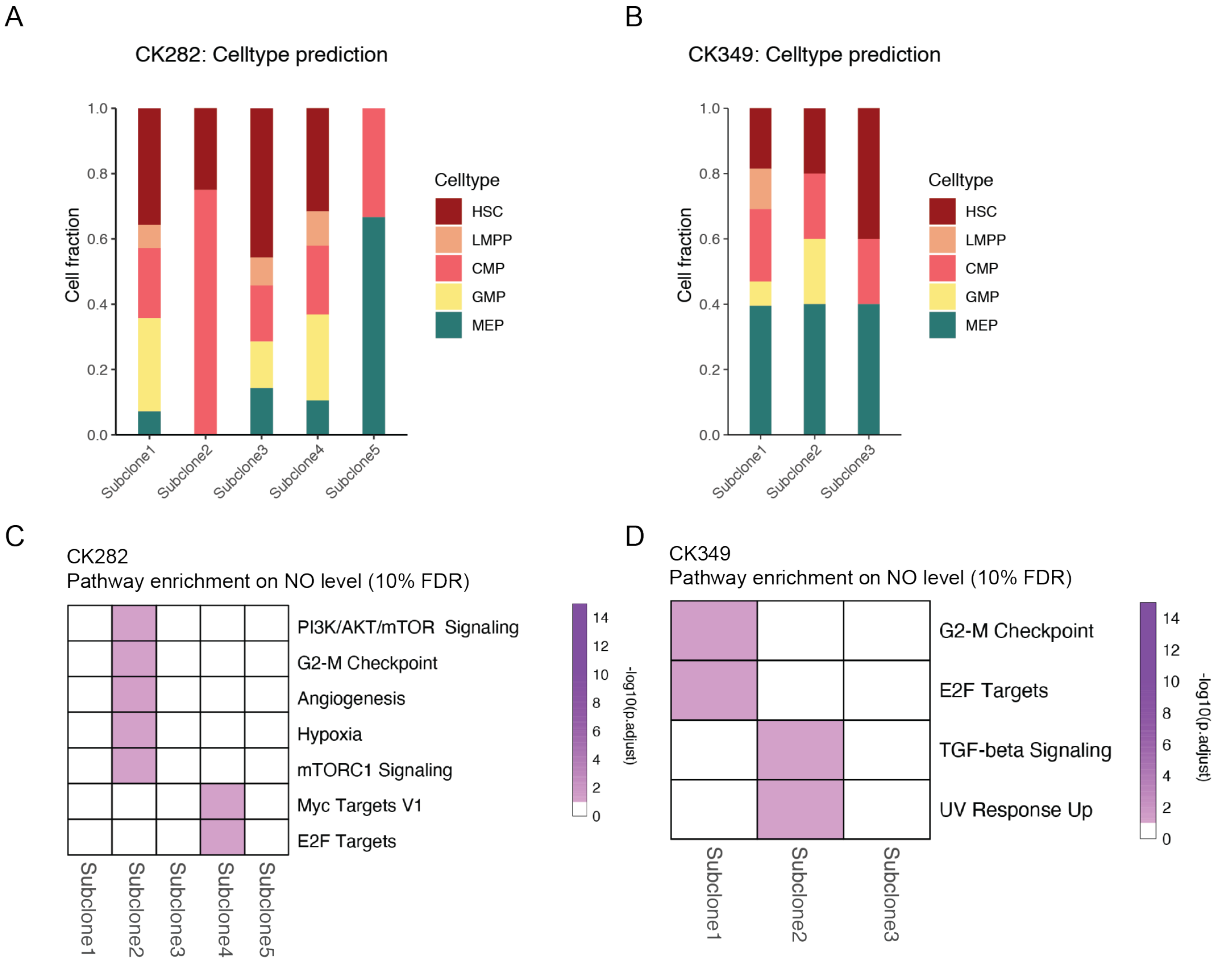


Figure 3.9 Subclone heterogeneity based on nucleosome occupancy. A-B Stacked bar plots showing the fraction of indicated hematopoietic stem and progenitor cell-like states in the different subclones in **A** CK282 and **B** CK349. Celltypes were annotated using a MNase-seq reference dataset from healthy sorted CD34+ bone marrow cells. **C-D** Heatmap showing enriched pathway modules (FDR < 0.1) in the different subclones in **C** CK282 and **D** CK349 based on single-cell nucleosome occupancy-profiling. Enrichment analysis was performed against the H (Hallmark gene sets) collection from MSigDB. Abbreviations: HSC, Hematopoietic stem cell; CMP, Common myeloid progenitor; GMP, Granulocyte-monocyte progenitor; MEP, Megakaryocyte-erythrocyte progenitor; LMPP, Lymphoid primed multipotent progenitor; FDR, False discovery rate; NO, Nucleosome occupancy. Analysis of nucleosome occupancy data performed jointly by Dr. Hyobin Jeong and myself.

To further study the phenotypic heterogeneity on transcriptional level, I sought to integrate scNOVA data with CITE-seq data from the same patients. I took advantage of the high-resolution SV breakpoint coordinates from the scNOVA framework and used them as input for CONICSmat[148], a computational tool that allows targeted somatic copy-number alteration identification from scRNA-seq data (Figure 3.10A, see Methods 6.1.8.1). I used the different copy-number alteration probabilities to assign each cell from the CITE-seq data to the corresponding subclone defined by scNOVA and visualized the combination of RNA and

protein expression using weighted nearest neighbor (WNN)-based UMAP (Figure 3.10B-C). Similar to the nucleosome occupancy-based analysis, I identified complex relationships between the subclones. In both CK282 and CK349, integration of the data showed clustering of the cells largely based on genetic subclones, suggesting that the phenotype of the cell is largely shaped by its genetic background (Figure 3.10B-C). Moreover, I detected subclone-specific upregulation of several genes associated with chemoresistance and adverse outcome particularly in subclones 2 and 3 in CK282 and in subclone 3 in CK349. Among these genes, I found, for instance, *DLK1*[149] and *HOPX*[150] to be upregulated in subclone 2 in CK282, and *HM13*[151] and *PDLIM1*[152] in subclone 3 (Figure 3.10D and Appendix Table 3). In CK349, the leukemic oncogene *KMT2A*[153] was particularly highly expressed in subclone 3 (Figure 3.10E, and Appendix Table 4). This subclone-specific transcriptomic heterogeneity was also recapitulated on the cell surface proteome level. For instance, in CK282, CD56 was highly expressed in subclone 1 whereas subclone 2 showed high expression of CD26 (Appendix Figure 7A). CD56 expression has been associated with adverse prognosis in AML while CD26 is an LSC marker in chronic myeloid leukemia[154-157]. In CK349, subclone 3 expressed high levels of CD49F and CD25, markers also linked with self-renewal and poor survival in AML (Appendix Figure 7B)[158,159].

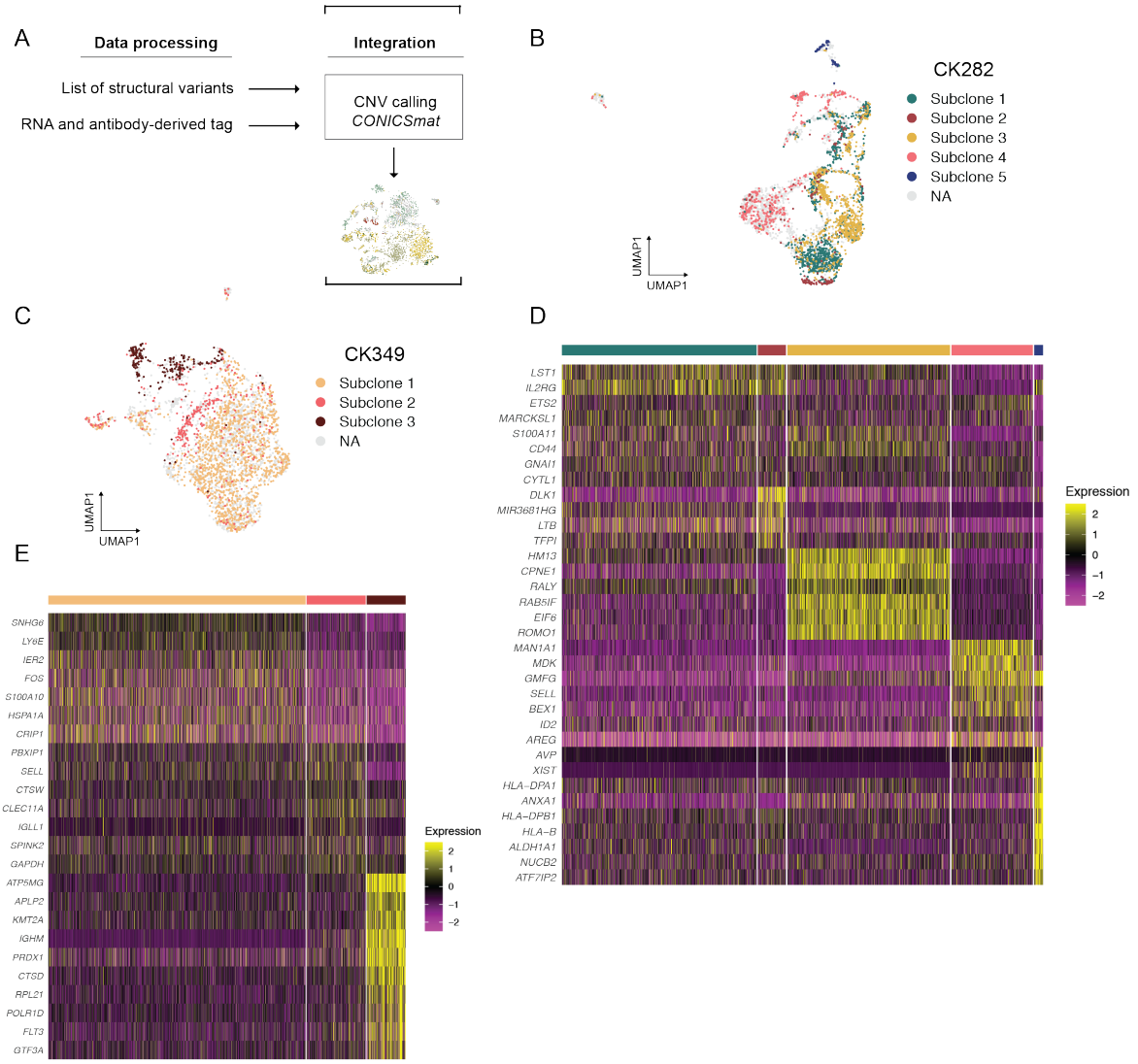


Figure 3.10 Transcriptomic heterogeneity in CK-AML. A Schematic of the data integration framework for scNOVA-CITE. Single-cell SV information from scNOVA and single-cell gene expression from CITE-seq was used as input for CONICSmat[148], a computational tool for somatic copy-number alteration discovery from scRNA-seq data. **B-C** Weighted nearest neighbor-based UMAP plot of CITE-seq data of leukemic cells from **B** CK282 and **C** CK349. Cells are colored based on the subclones identified using scTRIP. Annotation of each cell was based on somatic copy-number alteration discovery using CONICSmat. Cell that could not be confidently assigned are shown in gray as NAs. **D-E** Heatmap showing top differentially expressed (DE) genes for each subclone in **D** CK282 and **E** CK349. FDR < 0.05 and log-fold-change > 0.25 was considered significant. DE genes were calculated from normalized and variance stabilized counts using findMarkers function from scran with scaled expression values shown. Abbreviations: CNV, Copy number variation.

To characterize the processes underlying differential gene expression between the subclones, I performed pathway analysis on the upregulated genes. Supporting the nucleosome occupancy-based pathway analysis, mTORC1 signaling and cytoskeletal organization-related pathways showed dysregulation in subclone 2 in CK282 (Figure 3.11A-C). Other pathways with

increased activity in subclone 2 included the oxidative phosphorylation-related gene set (Figure 3.11A and D). These findings are in line with PI3K-Akt-mTOR pathway contributing to the regulation of cellular metabolism in some hematological malignancies[160-162]. Other subclones also showed activation of these pathways, indicating transcriptional similarities between subclones. In contrast to CK282, pathway analysis based on nucleosome occupancy compared to transcriptomics shared few similarities in CK349. Based on gene expression, mTORC1 and cytoskeletal organization-related pathways showed enrichment in subclone 3 in CK349 together with cell proliferation-related pathways (e.g. MYC targets) (Figure 3.11E-G). The other subclones in CK349 showed activation of multiple inflammatory response-related pathways (Figure 3.11E and H). As dysregulated pathway analysis between nucleosome occupancy and transcriptomics largely differed in CK349, other factors aside from epigenetic changes likely influenced gene expression.

Collectively, these results show that the scNOVA-CITE integrated framework can capture phenotypic intra-patient heterogeneity of genetically unique leukemic subclones. Moreover, it can identify subclones with more adverse characteristics and assess the associated pathway dysregulation.

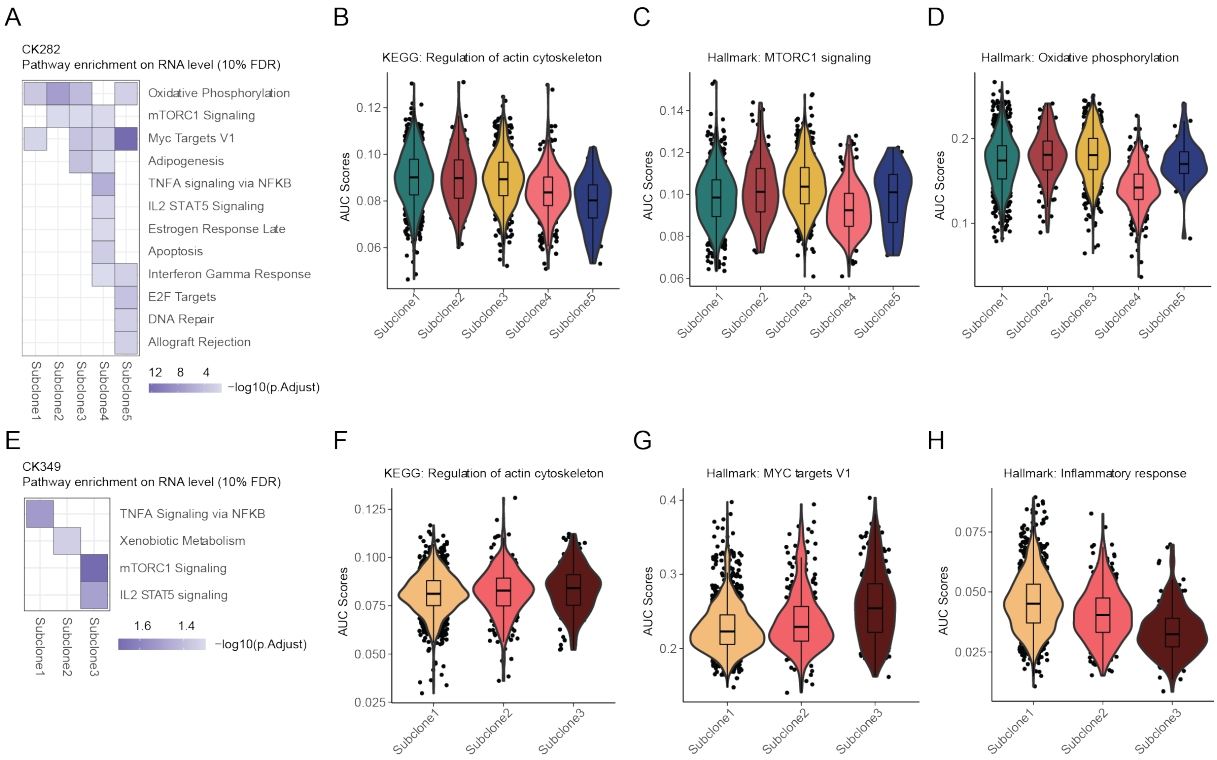


Figure 3.11 Subclone-specific transcriptional pathway dysregulation. A Heatmap showing enriched pathway modules (FDR < 0.1) in the different subclones in CK282 based on gene expression. Enrichment analysis was performed against the H (Hallmark gene sets) collection from MSigDB. **B-D** Violin plots of the Area Under the

Curve (AUC) score for activity of the indicated pathways in the different subclones in CK282, calculated using AUCell[163]. Expression levels of the individual genes in the score were calculated from normalized and variance-stabilized counts. **E** Heatmap showing enriched pathway modules (FDR < 0.1) in the different subclones in CK349 based on gene expression. Enrichment analysis was performed against the H (Hallmark gene sets) collection from MSigDB. **F-H** Violin plots of the Area Under the Curve (AUC) score for activity of the indicated pathways in the different subclones in CK349, calculated using AUCell. Abbreviations: FDR, False discovery rate.

3.2.5 Genetically complex subclones drive phenotypic and functional stem cell activity

I reasoned that the phenotypic diversity presented by different subclones might not only reflect genetic heterogeneity but also differences in disease propagating capacity. To explore this hypothesis, I first focused on patient CK397 that, unlike CK282 and CK349, showed lower subclonal heterogeneity (Figure 3.2A). The major clone of CK397 harbored a complex deletion at 7q spanning the entire q arm with only a 1Mb segment remaining at 7q21.11 (98.6% cells, 69/70 cells). Next to the major clone, I identified a singleton cell that resembled the major clone but lacked the large 7q deletion and had instead acquired a duplication at 2q (1.4% cells, 1/70 cells) (Figure 3.11A).

I used the scNOVA-CITE data integration framework to identify leukemic cells from the CITE-seq data with and without the 7q SV. When projected onto the WNN-based t-distributed stochastic neighbor embedding (tSNE), cells lacking the 7q deletion clustered largely separate from the other leukemic cells (Figure 3.11B). Next, I assessed the corresponding normal hematopoietic differentiation state of the leukemic cells and noticed a significant enrichment for primitive HSC-like and LMPP-like cells in the cells harboring the 7q rearrangement (Fisher's Exact test, $p < 0.01$). In contrast, cells lacking the deletion had a high abundance of red blood cell progenitor (Prog RBC)-like cells (Figure 3.11C). To further assess subclonal stemness, I used transcriptional stemness scores that have previously been applied to determine prognosis in AML[45]. Similar to the primitive differentiation state, LSC-associated transcription was also increased in the clone harboring the 7q deletion (Figure 3.11D). Encouraged by this finding, I next focused on the two CK-AML samples with higher subclonal heterogeneity. In CK349, subclone 3 had the highest LSC-associated transcription, whereas in CK282 three subclones (1 to 3) showed similarly increased enrichment of LSC-associated transcription compared to the other subclones (Figure 3.11E-F). Similar to CK397, the subclones with the highest LSC-associated transcription in both CK282 and CK349 were also most enriched for more primitive cell states compared to the other subclones (Figure 3.9A-B). Collectively, based

on the transcriptomic profiles one or multiple subclones may be enriched in stemness potential in CK-AML. These subclones also show enrichment for more primitive cell states.

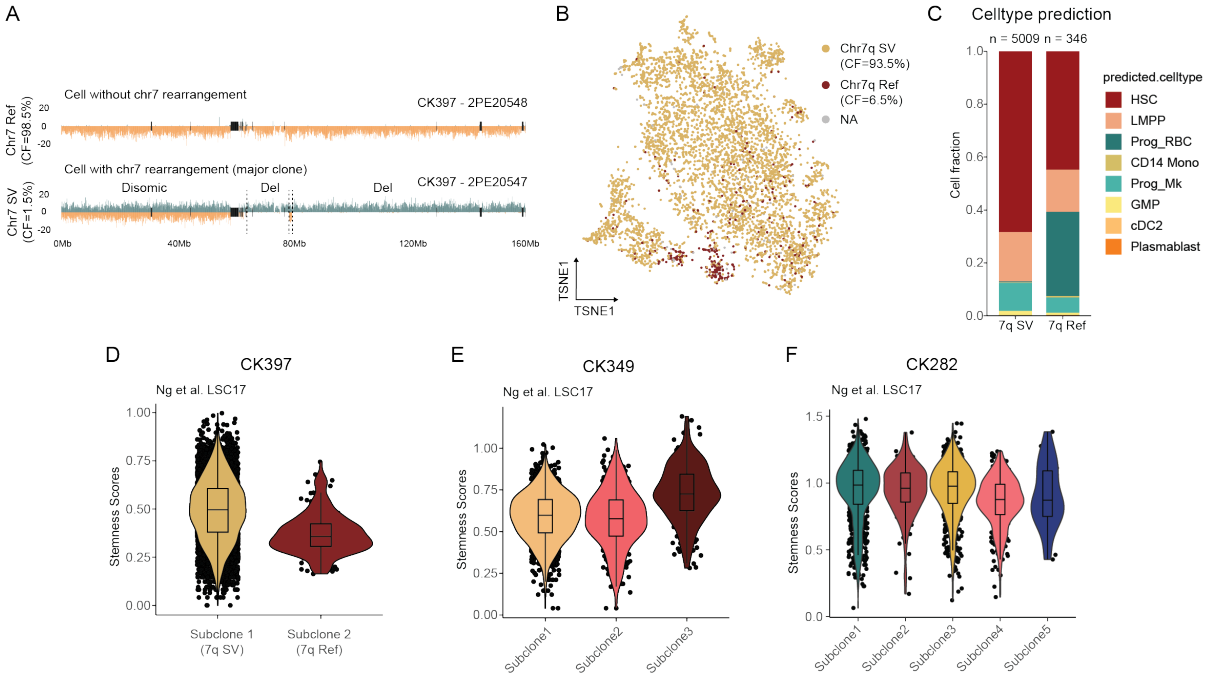


Figure 3.12 Subclone-specific stemness potential. A Chromosome 7 in two representative single cells from CK397 portraying the two subclones identified in CK397. Top plot shows a cell from subclone with a normal chromosome 7 (reference), represented by a single-cell. Bottom plot shows a cell from the major clone with large 7q deletion spanning almost the entire arm, represented by 69 cells. Reads denoting somatic SVs, discovered using scTRIP, mapped to the Watson (orange) or Crick (green) strand. Grey: single-cell IDs. **B** Weighted nearest neighbor-based tSNE plot of CITE-seq data of leukemic cells from CK397. Cells are colored based on the subclones identified using scTRIP. **C** Stacked bar plots showing the fraction of indicated hematopoietic cell-like states out of all cells in the different subclones in CK397. Number of cells assigned to each subclone is indicated on top of the bars. Celltype prediction was done by projecting single leukemic cells into a healthy reference bone marrow embedding and determining the similarity to the reference cell types. **D-F** Violin plots of the LSC17 score[45] from single-cell gene expression data in different subclones from **D** CK397, **E** CK349 and **F** CK282. Expression levels of the individual genes in the score were calculated from normalized and variance-stabilized counts. Wilcoxon test was used to compare the groups Abbreviations: chr, Chromosome; CF, Cell fraction; Ref, Reference; Del, Deletion.

To further confirm and study the clonal diversity and architecture of AML leukemia initiating cells, as defined by their ability to initiate (or regenerate) AML in immunocompromised mice, I transplanted cells from the primary CK-AML patient samples into sublethally irradiated NOD.*Prkdc*^{scid}.*Il2rg*^{null} (NSG) mice. I characterized the PDXs using scNOVA and/or optical genome mapping and bulk RNA-seq (Figure 3.13A). In CK397, the PDX harbored the complex 7q deletion with a fractional copy number of 1, indicating that all cells harbored the deletion

(Figure 3.13B). In the original diagnosis patient sample the cells with the 7q deletion represented the major clone and had higher LSC-associated transcription compared to the cells lacking the deletion. This exemplified preferential engraftment of the major clone. In contrast, both CK282 and CK349 showed preferential engraftment of minor subclones (Figure 3.13C-D).

In CK282, I identified five broader genetic subclones at diagnosis with subclones 1 to 3 showing a more adverse phenotype and higher LSC-associated transcription (Figure 3.12F). In the PDX, only individual cells harbored additional rearrangements besides the clonal SVs present already at diagnosis. Intriguingly, chromosome 20 that showed ongoing genomic instability at diagnosis, remained unstable in the engrafted cells (Appendix Figure 8A-B). The set of complex genetic rearrangements at chromosome 8 present in 5.3% of the diagnosis patient sample (4/75 cells) and uniquely in subclone 2, was now present in 97.8% of the engrafted cells (45/46 cells) (Figure 3.13D-E). This revealed a significant expansion of a minor subclone upon transplantation. Moreover, only individual cells harbored additional rearrangements. The set of complex genetic rearrangements at chromosome 8 presented as isochromosome 8q (Figure 3.13F), a secondary change often associated with secondary AMLs and solid tumors, leading to gain of *MYC* and *RAD21*[164-167]. To further validate the preferential engraftment of the subclone harboring the complex genetic rearrangements at chromosome 8, I analyzed two additional PDXs from the same diagnosis patient sample and observed the same clone expanding (Appendix Figure 9A). These findings demonstrate the robustness of the subclone with the complex chromosome 8 to expand and initiate leukemia over the other subclones.

In CK349, the PDX harbored the clonal SVs already present in the diagnosis patient sample with few additional rearrangements, and mostly only in individual cells. Next to these clonal SVs, the engrafted cells harbored a set of rearrangements spanning the whole chromosome 13, that were present in subclones 2 and 3 in the primary patient sample (Figure 3.13C). Notably, subclone 3 had the highest LSC-associated transcription out of all three subclones present at diagnosis (Figure 3.12E). The complex ‘seismic amplification’ at chromosome 11, detected in 5.5% of the cells at diagnosis (5/91 cells) and uniquely in subclone 3, was now present in 97.5% of the engrafted cells (45/46 cells) (Figure 3.13C). Moreover, the set of rearrangements had expanded, spanning the whole length of chromosome 11 (discussed later, see Figure 3.18). This suggests that cells from subclone 3 preferentially engrafted and expanded in the PDX while undergoing further amplifications. Similar to CK282, a second PDX from CK349 showed

engraftment of the same minor subclone, highlighting the clone's strong ability to initiate leukemia in mice (Appendix Figure 9B). Collectively, these data demonstrate that punctuated evolution in the form of genomic amplification events can act as a catalyst in the generation of evolutionary advantageous progeny with gross genome-wide karyotype alterations.

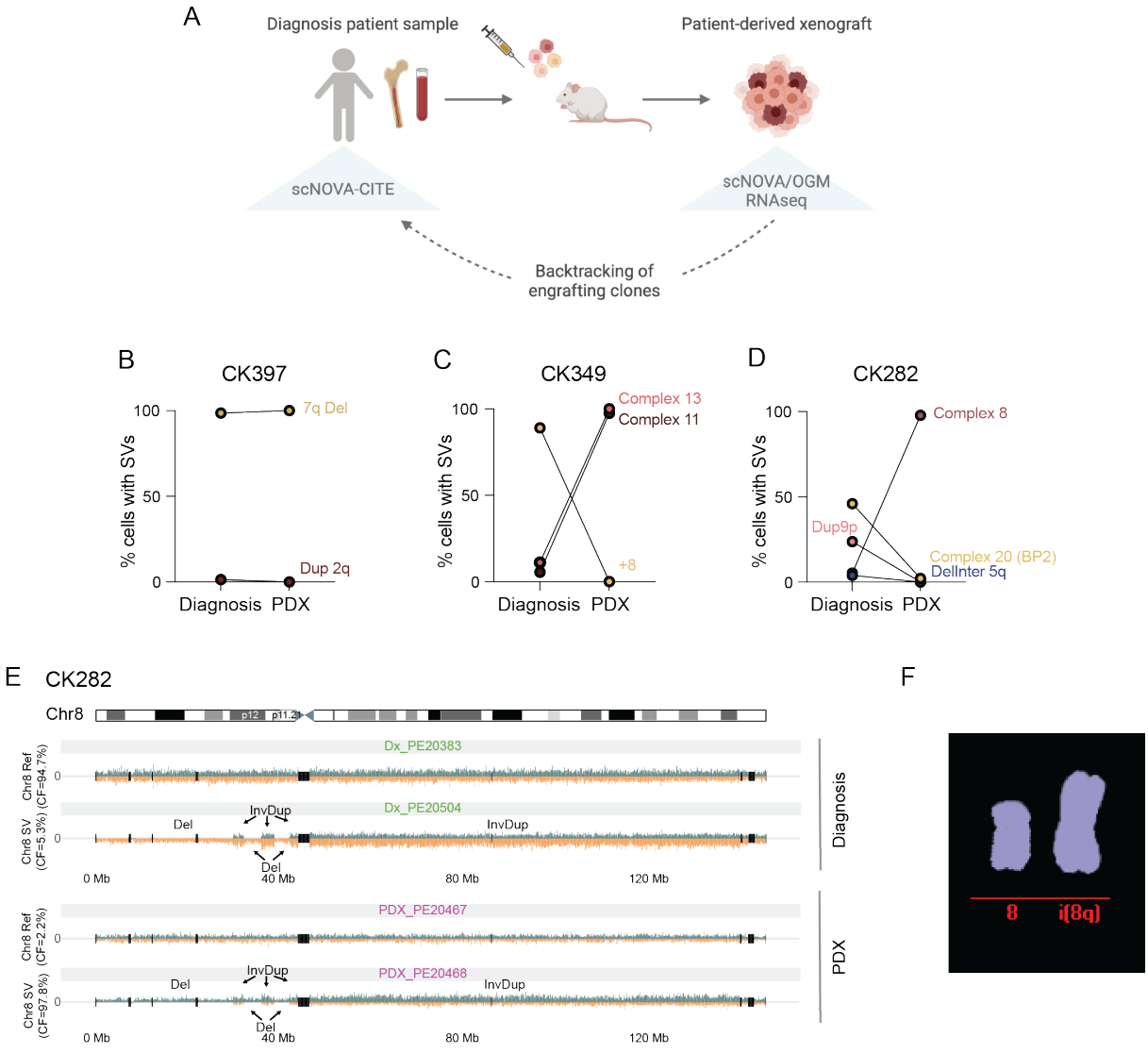


Figure 3.13 Subclone-specific leukemia re-initiation. A Schematic of the transplantation assay to study clonal evolution. Primary patient samples were transplanted into immunocompromised mice. Engrafting cells were assessed via scNOVA and/or optical genome mapping and bulk RNA-seq, and tracked back to cells at diagnosis. Engraftment driving cells at diagnosis were analyzed using scNOVA and CITE-seq. **B-D** Percentage of cells with subclone-defining SVs at diagnosis and in the engrafted cells in **B** CK397, **C** CK349 and **D** CK282. Lines connect the same SV at different time points (cells at diagnosis vs cells in PDX). **E** Complex genetic rearrangement comprising of two inverted duplications (InvDups), three deletions (Dels) and one larger inverted duplication (InvDup), spanning the whole chromosome 8q arm in two representative single cells from CK282 at diagnosis (top) and in two representative single cells in PDX (bottom). The cell fraction of the SV at the different time points is shown on the left. Reads denoting somatic SVs, discovered using scTRIP, mapped to the Watson (orange) or

Crick (green) strand. Green: single-cell IDs from diagnosis. Pink: single-cell IDs from PDX. **F** M-FISH of PDX cells from CK282 showing isochromosome 8q. Abbreviations: OGM, Optical genome mapping; Del, Deletion; Dup, Duplication; Inter, Interstitial; Ref, Reference; i, Isochromosome.

3.2.6 Dissecting the phenotype of the engraftment driving subclones

To study the non-genetic characteristics of the engraftment driving subclones, I tracked the engrafted cells in the PDX back to the subclones in the original patient samples (Figure 3.14A-B). Based on cell surface protein expression revealed by CITE-seq, both samples showed enrichment for different stem cell markers in the engraftment driving subclones. In CK282, I observed significant enrichment for a $CD34^{+}CD38^{-}CD90^{high}CD45RA^{-}$ cell surface phenotype in the engraftment driving subclone, a cell surface marker expression commonly associated with HSCs (Figure 3.14C). In AML, the expression of CD90 has been shown to be significantly higher in AML subtypes characterized by poor prognosis, such as secondary AMLs[168]. Moreover, it has been postulated that CD90-expressing HSPCs may be the target of the initial genetic lesions in AML or MDS that express CD90[168,169], supporting the preferential leukemia re-initiating potential by the $CD90^{high}$ cells in CK282. In CK349, the engraftment driving subclone expressed high levels of the stem cell marker CD49F together with CD45RA (Figure 3.14D)[158]. Together, these data suggest that the engraftment driving LSCs are phenotypically more immature. This was supported by the nucleosome occupancy-based cell type analysis performed by Dr. Hyobin Jeong that revealed enrichment of more primitive HSC-like nucleosome occupancy profiles in the engrafted cells compared to the cells at diagnosis (Figure 3.14E-F).

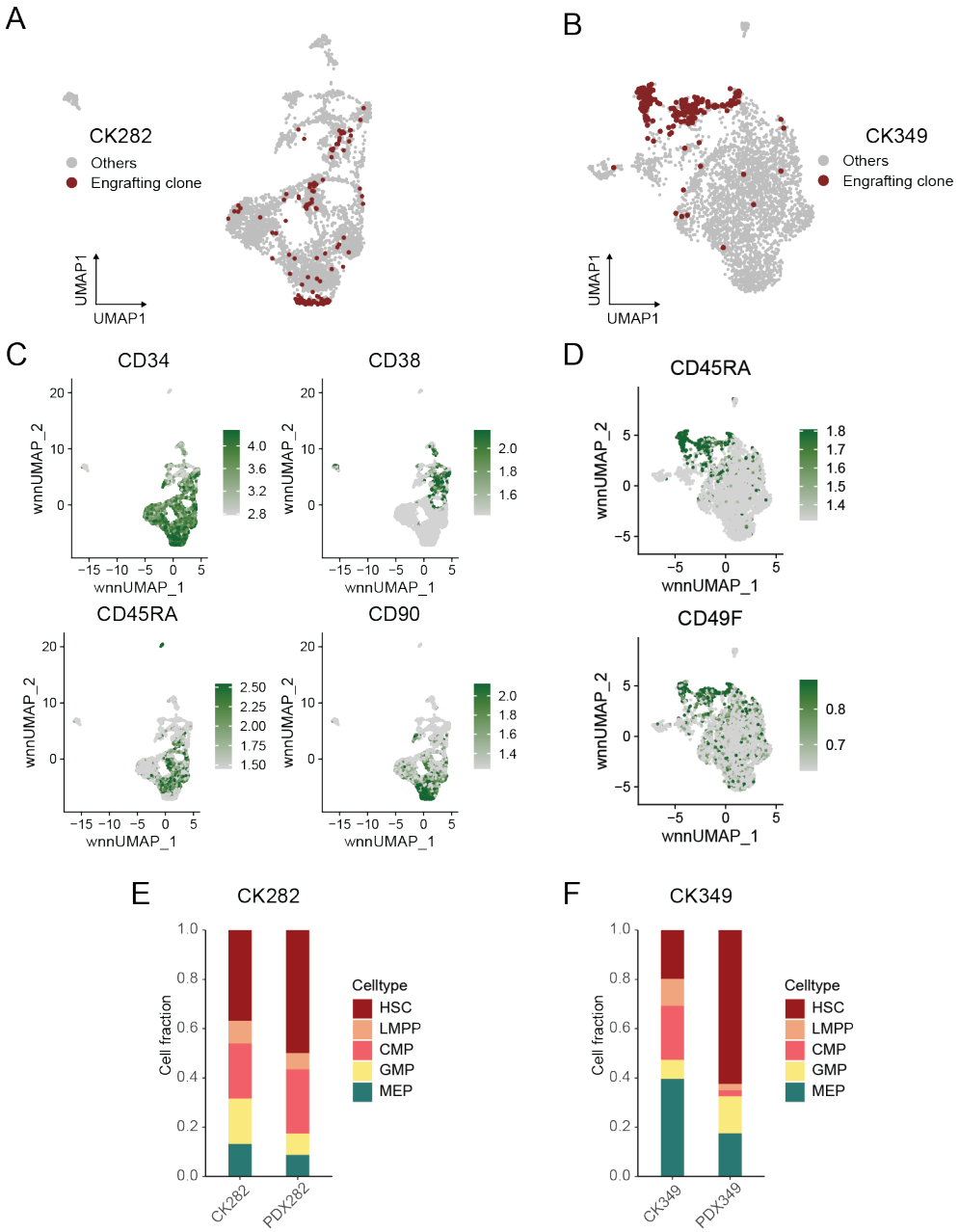


Figure 3.14 Cell surface characteristics and hematopoietic cell-like states of engraftment driving cells. A-B Weighted nearest neighbor (WNN)-based UMAP plot of CITE-seq data from leukemic cells at diagnosis from **A** CK282 and **B** CK349. Engraftment driving cells are colored in red based on SVs detected in engrafted cells. **C** UMAP plot showing cell surface protein expression of CD34, CD38, CD45RA and CD90 colored as centered log ratio-transformed counts in CK282. **D** UMAP plot showing cell surface protein expression of CD45RA and CD49F colored as centered log ratio-transformed counts in CK349. **E-F** Stacked bar plots showing the fraction of indicated hematopoietic cell-like states out of all cells between cells from diagnosis and PDX in **E** CK282 and **F** CK349. Cell types were annotated using a MNase-seq reference dataset from healthy sorted CD34+ bone marrow cells. Analysis of nucleosome occupancy data in E-F performed jointly by Dr. Hyobin Jeong and myself.

I next assessed the transcriptomic characteristics associated with leukemia initiation potential and inferred that 114 genes were significantly upregulated in the diagnosis subclone driving engraftment, including *DLK1*, *THY1* and *RAD21* (Figure 3.15A, Appendix Table 5). *RAD21*

resides in the genomic segments affected by the isochromosome formation at chromosome 8 and has been shown to be highly expressed in various types of undifferentiated cancer with poor prognosis[170,171]. It allows the repair of additional DNA damage, thus potentially increasing the fitness advantage of the cells overexpressing it. In CK349, I inferred 327 genes that were significantly upregulated in the diagnosis subclone driving engraftment, including *IGHM*, *ATP5MG*, *PRDX1* and *KMT2A* (Figure 3.15B, Appendix Table 5). As majority of these genes reside in the genomic segments affected by the somatic DNA rearrangements at chromosome 11, they were likely upregulated due to higher copy-number levels in the 'seismic amplification'. Thus, they may, to an extent, present passengers rather than solely disease drivers. Using bulk RNA-seq of the matched diagnosis-PDX samples, I verified that the upregulated genes in the engraftment driving subclones at diagnosis were also increased in the engrafted cells of CK282 and CK349 (Appendix Figure 10A-B).

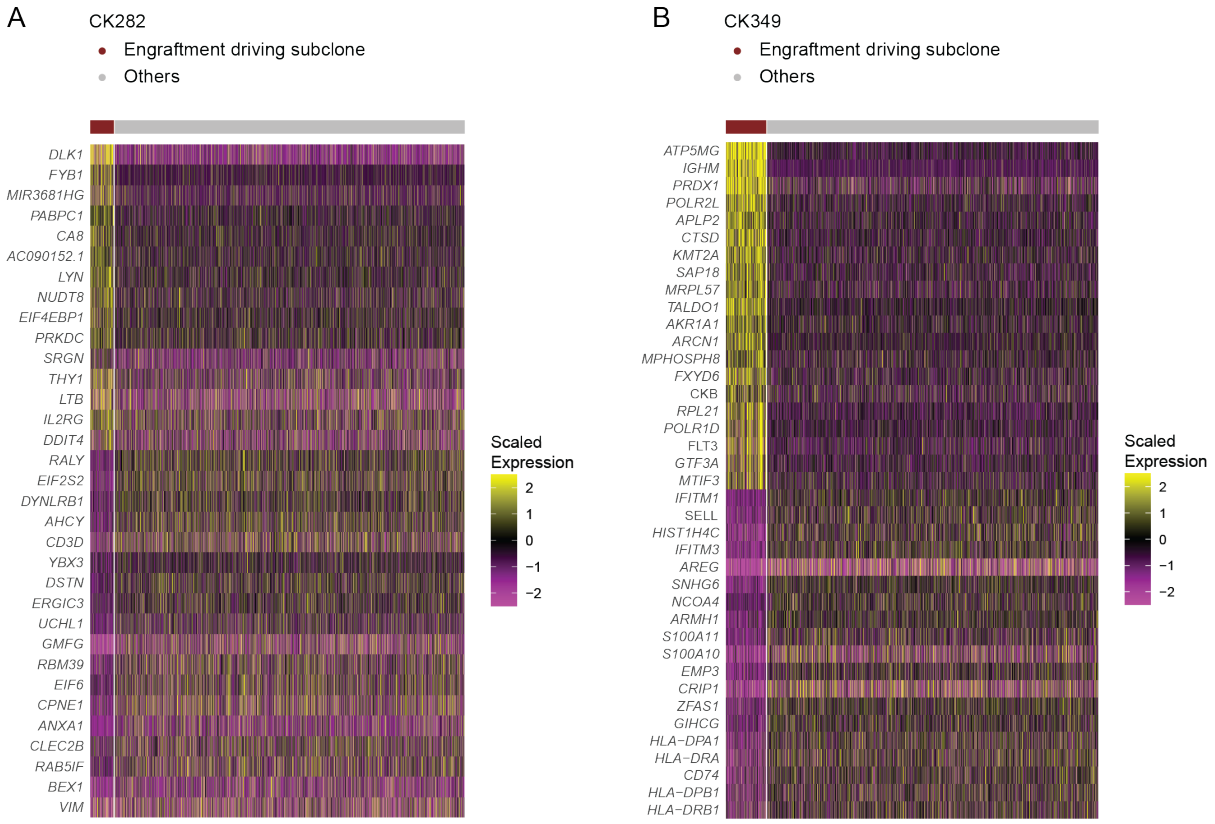


Figure 3.15 Transcriptomic characteristics of engraftment driving cells. A-B Heatmap showing top differentially expressed (DE) genes for engraftment driving subclone vs. others in **A** CK282 and **B** CK349. FDR < 0.05 and log-fold-change > 0.25 was considered significant. DE genes were calculated from normalized and variance stabilized counts using findMarkers function from scran with scaled expression values shown.

Last, I looked for commonalities between the engraftment driving subclones of CK282 and CK349. I identified eight genes that were significantly upregulated in both samples (FDR <

0.05 and logFC > 0.25), including the aldehyde dehydrogenase encoding gene *ALDH1A1* (Figure 3.16A). As high expression of *ALDH1A1* has been associated with cancer stemness and refractoriness to chemotherapy particularly in solid cancers as well as in a subset of AMLs[172-174], I sought to determine whether high *ALDH1A1* expression is also a more widespread phenomenon in CK-AML. I assessed the *ALDH1A1* gene expression levels in the BeatAML cohort[20], and observed a significant association with high *ALDH1A1* expression and reduced overall survival within CK-AML patient samples (Figure 3.16B), providing a potential vulnerability in a subset of CK-AMLs with poor outcome.

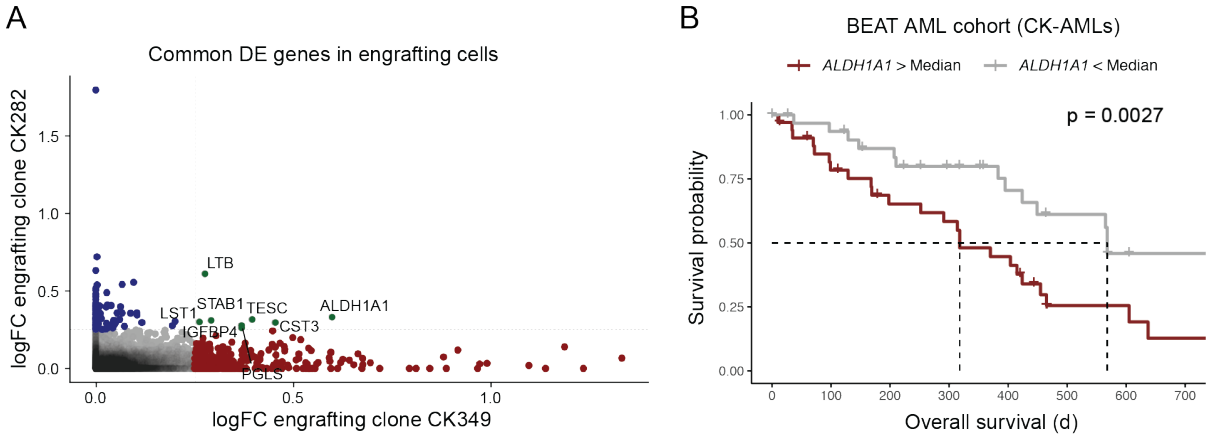


Figure 3.16 Common vulnerabilities in engraftment driving cells. A Scatter plot showing differentially expressed (DE) genes in engraftment driving subclones of CK282 and CK349. Shown are common DE genes (green and annotated) and DE genes specific for CK282 (blue) and CK349 (red). **B** Overall survival of CK-AML patients in the BeatAML cohort stratified by above and below median *ALDH1A1* expression. Log-rank test was used to compare the overall survival between the two groups. Abbreviations: logFC, log-fold-change; d, Day.

3.2.7 Targeting chemotherapy-resistant subclones that drive disease regeneration

To assess if I can target the leukemia regenerating cells, I performed single-cell drug sensitivity profiling at diagnosis using drug combinations consisting of standard chemotherapy regimens as well as BH3 mimetics shown to target cancer stemness[175-179] (Figure 3.17A, Methods Table 6.3). In CK282, I identified the engraftment driving cells to express high levels of CD90 and lack CD45RA (Figure 3.17B). Using these markers, I measured the *ex vivo* drug responses for each subpopulation in CK282, and observed a striking response to BCL-xL inhibitor A-1331852 (Figure 3.17C-D). All subpopulations showed a remarkable reduction in cell viability, with $CD90^{high}CD45RA^{-}$ cells having the strongest effect (Figure 3.17C-D). Intracellular staining of the BCL2 family members by flow cytometry showed the highest expression of BCL-xL in the $CD90^{high}CD45RA^{-}$ cells while the other anti-apoptotic proteins showed low

expression, supporting the *ex vivo* results (Figure 3.17E, Appendix Figure 11). Of note, the CD90 expressing cells were highly resistant to all other tested drugs, including standard chemotherapy regimens (Figure 3.17D, Appendix Figure 12A). I repeated the *ex vivo* drug screening in the PDX cells and observed the BCL-xL inhibitor to still be highly efficacious (Appendix Figure 12B). Collectively, I could effectively target the leukemia regenerating cells in CK282 *ex vivo* using a BCL-xL inhibitor, providing an alternative to standard chemotherapy.

To further exemplify the translational relevance of the analyses, I measured the *ex vivo* drug response of patient sample CK349 to cytarabine and daunorubicin – the same chemotherapy regimens the patient received as first-line treatment. Based on CITE-seq analysis, the engraftment driving subclone expressed high levels of CD49F and CD45RA (Figure 3.17F). I observed a distinct resistance to monotherapy as well as to combination therapy of cytarabine and daunorubicin exclusively in the subpopulation expressing CD49F and CD45RA ($CD45RA^{+}CD49F^{high}$) (Figure 3.17G-H). However, the hypomethylating agent azacitidine alone or together with the BCL-2 inhibitor venetoclax showed considerable response (Appendix Figure 12C). This suggests that the leukemia initiating cells in CK349 are highly resistant to standard chemotherapy *ex vivo* but may be targeted with azacytidine-based regimens.

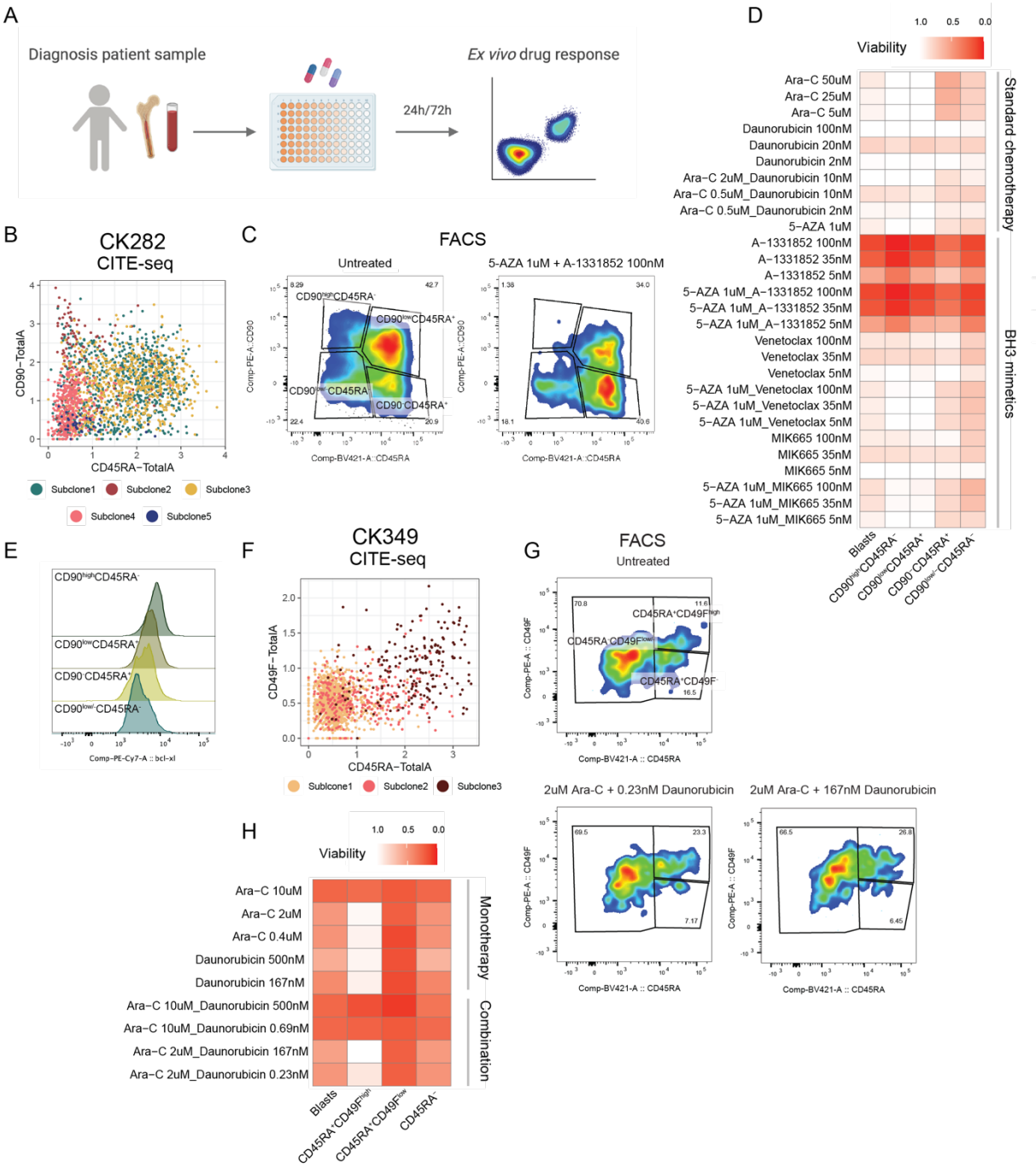


Figure 3.17 Targeting leukemia regenerating cells. A Schematic of the *ex vivo* single-cell drug sensitivity profiling assay. Mononuclear cells of primary patient and PDX samples were treated *ex vivo* for 24h and/or 72h with chemotherapy regimens. Following treatment, cells were stained for viability and a set of surface markers (Methods Table 6.4) and analyzed using flow cytometry-based drug profiling. **B** Scatter plot of CD45RA and CD90 cell surface expression as identified from CK282 CITE-seq data pre-gated to leukemic cells. Shown are centered log ratio-transformed counts highlighted according to subclones. **C** FACS plot displaying expression of CD45RA and CD90 in pre-gated leukemic cells. The gates highlight four populations with different CD45RA and CD90 expressions. Cells from untreated (left) and BCL-xL inhibitor-treated (A-1331852, 100nM) together with hypomethylating agent (5-AZA, 1uM, right) conditions are shown after 24h *ex vivo* exposure. **D** Viabilities (fraction of viable cells compared to untreated control) of different populations after 24h *ex vivo* exposure with indicated concentrations of standard chemotherapy regimens as well as BH3 mimetics. **E** Fluorescence intensity

of BCL-xL protein expression in four populations with different CD45RA and CD90 expressions. **F** Scatter plot of CD45RA and CD49F cell surface expression as identified from CK349 CITE-seq data pre-gated to leukemic cells. Shown are centered log ratio-transformed counts highlighted according to subclones. **G** FACS plot displaying expression of CD45RA and CD49F in pre-gated leukemic cells. The gates highlight three populations with different CD45RA and CD49F expressions. Cells from untreated (left) and cytarabine (2uM, middle and right) together with daunorubicin-treated (0.23nM, middle, and 167nM, right) conditions are shown after 72h *ex vivo* exposure. **H** Viabilities (fraction of viable cells compared to untreated control) of different populations after 72h *ex vivo* exposure with indicated concentrations of standard chemotherapy regimens.

3.2.8 Chemotherapy-resistant leukemic stem cells drive disease progression

To validate that the engraftment driving cells showing resistance *ex vivo* also mimic relapse, I analyzed cells from patient CK349 during relapse after treatment with cytarabine and daunorubicin followed by alloHSCT. Due to lack of viably stored material from relapse, I took advantage of the classical cytogenetics performed as part of routine diagnostics in the clinic (Figure 3.18A). I predicted that the $CD45RA^{+}CD49F^{high}$ clone harboring the 'seismic amplification' at chromosome 11 but lacking the chromosome 8 trisomy detected in the majority of cells, would drive disease progression, not the major clone. All SVs detected at diagnosis were present also at relapse, with the exception of chromosome 8 trisomy (Figure 3.18B). Based on the clinical karyograms, the fraction of chemotherapy-resistant cells, lacking the chromosome 8 trisomy, increased from 0% (0/25 cells) to 88% (22/25 cells) at relapse (Figure 3.18B). The remaining 12% had a normal female karyotype and originated from the alloHSCT donor. Instead, cells at relapse had acquired a large marker chromosome (Figure 3.18B). scNOVA analysis of the PDX revealed a complex 'seismic amplification' at chromosome 11 together with unaltered chromosome 8, as validated by M-FISH (Figure 3.18C-D). The fraction of cells with this karyotype increased from 5.5% (5/91 cells) at diagnosis to 97.5% (45/46 cells) in the PDX, as assessed by scNOVA (Figure 3.18C). M-FISH analysis of the PDX revealed these amplifications to result in a large ring chromosome, linearized marker chromosome or both, similar to the observations at relapse (Figure 3.18D, Figure 3.8B). These data demonstrate that the clone driving engraftment in the PDX can also drive relapse in the patient.

In summary, I dissected the intra-patient heterogeneity of CK-AML patient samples at genomic, epigenetic, transcriptional, and cell surface proteome levels and followed the disease progression longitudinally in PDX as well as in response to *ex vivo* drug treatment. These data highlight the clinical relevance of intra-patient heterogeneity in ongoing tumor evolution, and show how minor subclones with complex SVs can drive disease progression and relapse.

Moreover, I showed that the PDX system not only captures the cells with self-renewing capacity, as confirmed by the ability to regenerate AML in NSG mice, but also captures the chemotherapy-resistant cells, and may be used to model relapse.

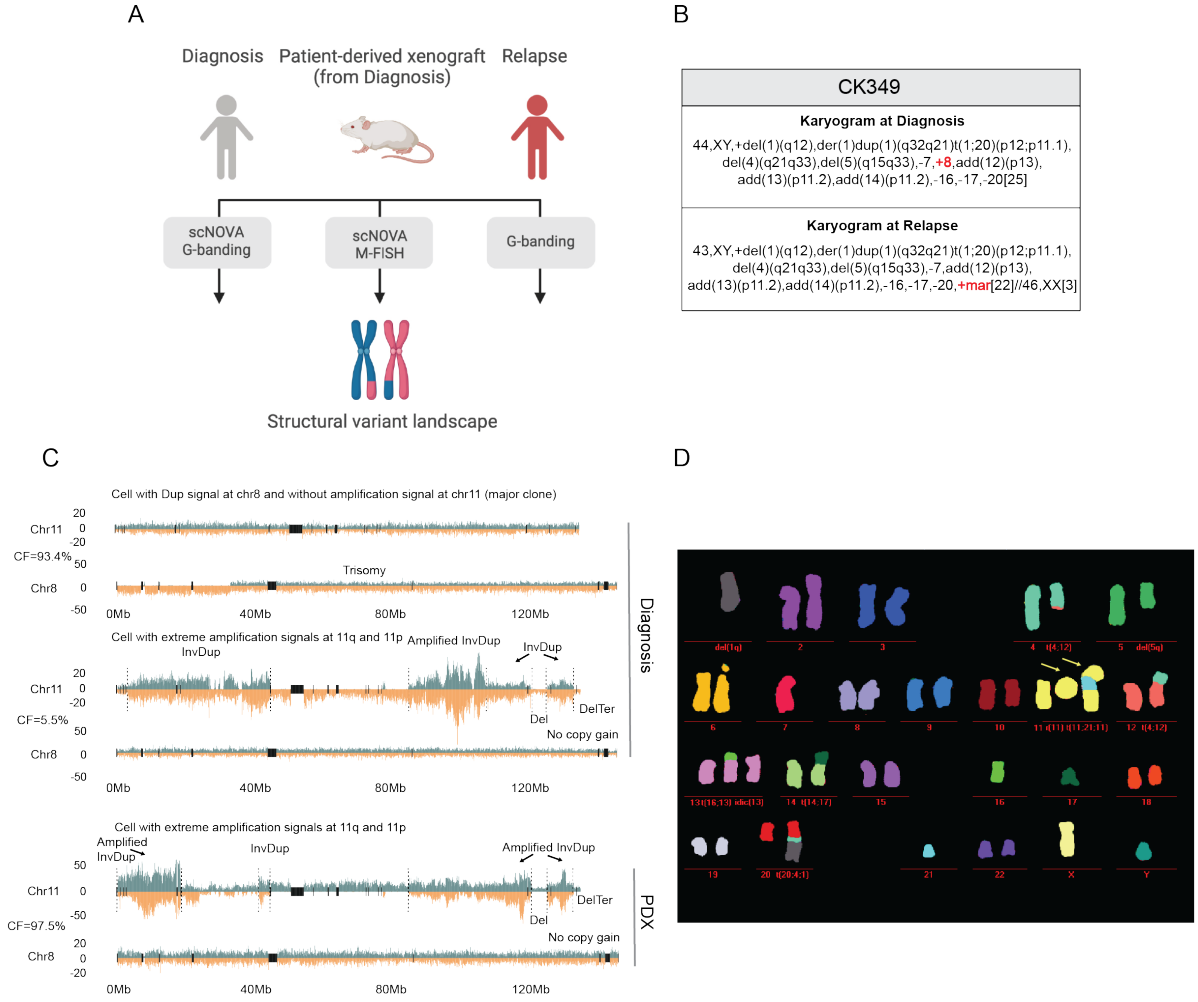


Figure 3.18 Engrafting cells resemble cells at relapse. A Schematic of the SV landscape comparison between diagnosis, PDX and relapse samples from CK349. **B** G-banding karyograms of CK349 at diagnosis and at relapse. SVs differing between the two time points are highlighted in red. **C** Depiction of two example CK349 cells at diagnosis and one in PDX with differing levels of amplification at chromosome 11, based on no amplification at diagnosis (upper, major clone), extreme amplification at diagnosis (middle, minor clone) and extreme amplification in PDX (lower, major clone). For each cell, the corresponding chromosome 8 trisomy status is shown beneath, which scNOVA inferred to be mutually exclusive with chromosome 11 amplification. **D** M-FISH of an engrafted cell from the PDX of CK349. Arrows (yellow) highlight relapse-mimicking SVs at chromosome 11, including ring chromosome 11 and linearized marker chromosome. Abbreviations: mar, Marker chromosome; CF, Cell fraction; r, Ring chromosome; t, Translocation.

3.3 Discussion I

In this chapter, I presented the efforts to dissect the intra-patient heterogeneity of CK-AML using two multi-modal single-cell technologies and to assess the functional impact of the detected heterogeneity. Advances in single-cell approaches allowed me to highlight the clinical relevance of assessing dynamic intra-patient heterogeneity and offer novel targeting options for patient specific disease-driving cells.

Genomic instability is a characteristic of most human cancers with chromosomal instability representing the major form of genomic instability[180]. The loss of p53 function in CK-AML as a result of mutations in *TP53* or deletion of the gene has been shown to result in simultaneous presence of multiple related subclones as well as complex derivative and marker chromosomes containing material from multiple chromosomes[28]. The work in this chapter extended on these findings and revealed high structural complexity and different degrees of chromosomal instability across single cells in CK-AML. As a plate-based method, Strand-seq enables whole genome sequencing in single cells but with a relatively low throughput[90]. I assessed SVs in hundreds of single cells and revealed complex clonal hierarchies represented by multiple subclones, while likely still underestimating the genetic intra-patient heterogeneity of CK-AML. Even though limited in throughput, the Strand-seq approach enabled the identification and detailed construction of complex genomic rearrangements in an unprecedented resolution, addressing several limitation of prior AML genomic studies[31,181].

While the scNOVA-CITE approach made it possible to study the functional consequences of these genomic rearrangements, others have also assessed the functional impact of clonal architecture in AML[22,182]. Xenotransplant assays have shown significant subclonal restriction in PDXs with one subclone often dominating the engraftment[182]. This is in line with the results described here. In contrast, a lasting clonal heterogeneity following transplantation has also been reported[22]. These contradicting results may represent heterogeneity in AML, but also suggest that the functional impact of mutations assessed by xenotransplantation studies is not completely clear. Importantly, others have focused on targeted sequencing approaches to assess clonal heterogeneity whereas I showed that assessing SVs in single cells is also important for a full picture of the genetic heterogeneity[22,182]. By assessing the consequences of genomic rearrangements, I not only demonstrated that specific subclones preferentially engrafted in mice but were also enriched for more stemness features, and showed resistance to standard

chemotherapy *ex vivo*. These findings support and add considerable detail to the genomic, phenotypic and functional landscape of CK-AML[28,31,32,34,35].

The importance of intra-patient heterogeneity during disease progression has been well described in solid cancers[183,184]. The findings of this chapter demonstrated this to be the case also in CK-AML. I observed substantial cellular complexity represented by a wide spectrum of SVs with complex genomic rearrangements present in all four CK-AML samples. Previous studies in AML have shown that complex genomic rearrangements, such as chromothripsis, are associated with adverse cytogenetics including alterations in *TP53* and complex karyotype[30,185]. Typically, chromothripsis is described as an early event in tumor evolution that posits rapid bursts of adaptive evolution in the form of hundreds of simultaneous genomic rearrangements[30,82]. I showed in single cells that punctuated evolution not only represents early events but also ongoing tumor evolution with complex genomic rearrangements present in minor subclones in addition to major clones. Such clonal diversification may give rise to subclones with increased fitness[31,186].

I provided evidence of the subclonal evolution of seismic amplifications resulting in selective advantage over other clones. The evolution of seismic amplifications has been proposed to follow a multi-step model that is initiated by chromothripsis followed by formation of circular DNA structure that undergoes several rounds of circular recombination via circular BFB cycles[85]. I showed the existence of intermediate cells with differences in copy number levels, supporting the concept of a step-wise acquisition of amplified regions via circular recombination[85,187]. However, the exact ordering of initial events remains unclear. This, to my knowledge, represents the first seismic amplification event reported and characterized in single cells with a longitudinal follow-up, adding considerable detail to the model of seismic amplifications[85,187].

Furthermore, I observed that acquisition of genomic rearrangements can generate subclones with converging characteristics leading to primitive, stem-like phenotypes. Functionally, these subclones showed characteristics associated with aggressive disease, namely resistance to standard chemotherapy and ability to regenerate leukemia *in vivo*[45]. As clonal and subclonal SVs tend to occur as ordered events, this may reflect the continuous optimization of the fitness landscape throughout tumor evolution[186]. Consistent with this concept, in chronic myeloid leukemia terminally differentiated cells transition into immature, primitive cells in the advanced

phase, resulting in an aggressive disease phenotype. This transition is associated, for example, with an increase in genomic instability[188], a phenomenon associated with poor outcome in different cancer entities[31,186]. Here, I also observed higher karyotypic complexity in subclones driving engraftment, which points towards a mechanistic link between stemness and karyotypic complexity also in CK-AML. The causal relationship between complex genomic rearrangements and stemness in CK-AML remains to be determined in future studies with more samples.

As an example of metabolic adaptability, I detected *ALDH1A1* as a shared upregulated gene in chemotherapy resistant subclones driving engraftment *in vivo*. ALDH plays a role in reactive oxygen species scavenging and reduces oxidative stress in stem cells[189]. This may give the cell the ability to cope with genomic instability. Clinically, ALDH1A1 expression contributes to drug resistance and has been associated with poor outcome especially in solid tumors[172]. Interestingly, in solid cancers All-Trans-Retinoic Acid, a vitamin A metabolite used to successfully treat acute promyelocytic leukemia, has been shown to re-sensitize cells to chemotherapeutics in tumors by reducing ALDH1A1 levels[190]. Therefore, targeting this reactive oxygen species scavenger might offer an intriguing alternative to overcome chemotherapy resistance also in AMLs with complex cytogenetics.

Last, I assessed if I could target the disease-driving leukemia re-initiating cells using known drug combinations consisting of BH3 mimetics shown to target cancer stemness[175-179]. Using a BCL-xL inhibitor, I was able to efficiently target the functional LSCs in a patient sample with secondary AML following primary myelofibrosis with highly adverse genetics. BCL-xL has been shown to be over-expressed in myeloproliferative neoplasms that can be targeted by BCL-xL-based inhibition[191,192]. These data together with the findings reported here provide exciting future research opportunities in a subset of AMLs largely lacking effective therapeutic options.

In summary, this is the first study assessing genetic and phenotypic heterogeneity as well as the resulting functional outcome in such detail in clinically relevant primary CK-AML patient samples. These findings provide important detail to the understanding of ongoing tumor evolution and show how minor subclones with complex SVs can generate LSCs that drive disease progression, mediate resistance and relapse. While this work offers novel targeting options for the disease driving LSCs in individual CK-AML patients, identifying clinically translatable biomarkers to predict broader response of disease-driving cells to targeted therapies

are further required. In **Chapter 4**, I will present a novel strategy to predict treatment response and facilitate clinical decision making by scoring the combinatorial expression of BCL-2 family members in disease-driving LSCs.

Chapter 4 – Flow Cytometry-based Combinatorial BCL-2 Family Expression in Acute Myeloid Leukemia Stem Cells Predicts Clinical Response to Venetoclax

4.1 Introduction and Motivation II

In recent years there has been talk of a golden era in cancer research as a high number of new anti-cancer drugs have entered the clinical setting[193]. Since 2015, the US Food and Drug Administration has approved more than 80 new anti-cancer drugs[193]. These include the BCL-2 inhibitor venetoclax in combination with azacytidine or decitabine for newly-diagnosed elderly or frail AML patients ineligible for intensive induction chemotherapy[59]. Venetoclax targets the anti-apoptotic protein BCL-2 that among other factors conveys survival dependency to AML cells[194-196].

As venetoclax-based therapies have shown good tolerability together with high efficacy, they are now under investigation as first-line treatment for adult AML patients eligible for intensive induction chemotherapy (NCT04801797, NCT05177731). Molecular pathology has become fundamental not only to inform on tumor diagnosis and prognosis but also to drive therapeutic decisions in daily practice[197]. Thus, biomarkers are needed to better identify patients benefitting from venetoclax-based therapies compared to standard chemotherapy and to assist with patient stratification[198].

Several determinants of venetoclax sensitivity have been reported, such as monocytic differentiation of leukemic blasts[198-200], apoptotic priming[201] and cell of origin[202]. Monocytic differentiation has been of particular interest in several studies that have defined differentiation based on the morphological FAB classification and/or expression of monocytic markers $CD11b^+$, $CD64^+$ or $CD68^+$ detected by flow cytometry[203]. However, studies assessing clinical AML cohorts treated with venetoclax in combination with azacytidine have not shown an association with monocytic differentiation and inferior outcome[59,204]. This suggests that responsiveness to venetoclax-based therapy in the clinical setting is more complex than suggested by pre-clinical studies alone.

In this chapter I explore the basis for the inconsistency between mechanistic and clinical findings in regards to monocytic differentiation and response to venetoclax and azacytidine

treatment. I describe that monocytic blasts express high levels of MCL-1 and are resistant to venetoclax and azacytidine treatment *ex vivo* as well as in patients, but lack significant disease-driving LSC-potential. In contrast, immature GPR56+ cells are enriched with functional LSCs, depend on BCL-2 and respond to venetoclax and azacytidine treatment *ex vivo* and in patients. Furthermore, I introduce a clinically translatable biomarker that can be used to predict response to venetoclax and azacytidine treatment by a flow cytometry-based assay. This biomarker integrates BCL-2, BCL-xL and MCL-1 protein expression levels in the disease-driving LSCs into a "Mediators of Apoptosis Combinatorial-Score" (MAC-Score) to predict individual patient response and duration on venetoclax and azacytidine therapy.

The results described in this section were part of a joint project together with Dr. Alexander Waclawiczek and Dr. Simon Renders. The generation and analysis of the data was mostly generated and analyzed together and is also stated as such in the text. The contents of this chapter are based on a manuscript that has been provisionally accepted in *Cancer Discovery*.

"Combinatorial BCL-2 family expression in Acute Myeloid Leukemia Stem Cells predicts clinical response to Azacitidine/Venetoclax"
Alexander Waclawiczek[*], Aino-Maija Leppä[*], Simon Renders[*], Karolin Stumpf[§], Cecilia Reyneri[§], Barbara Betz[§], Maike Janssen, Rabia Shahswar, Elisa Donato, Darja Karpova, Vera Thiel, Julia M. Unglaub, Susanna Grabowski, Stefanie Gryzik, Lisa Vierbaum, Richard F. Schlenk, Christoph Röllig, Michael Hundemer, Caroline Pabst, Michael Heuser, Simon Raffel, Carsten Müller-Tidow, Tim Sauer[#], and Andreas Trumpp[#]

[*,§,#]Contributed equally

4.2 Results II: BCL-2 family expression in LSCs predicts venetoclax response

4.2.1 Monocytic differentiation of AML is associated with *ex vivo* resistance to venetoclax and azacytidine

To identify predictive markers for the response of venetoclax and azacytidine treatment, I together with Dr. Alexander Waclawiczek selected 19 AML cell lines that differed in cell of origin, immunophenotype and cytogenetics (Methods Table 6.8). Based on the mean fluorescent intensity (MFI) of the monocytic cell surface marker CD64, we stratified the cell lines into monocyte-like AMLs (Mono-AMLs, n = 8) and primitive-like AMLs (Prim-AMLs, n = 11). We exposed the cell lines *in vitro* to venetoclax and azacytidine for 72h and evaluated the viability. Mono-AML cell lines were resistant to venetoclax and azacytidine whereas Prim-AML cell lines were largely sensitive to the treatment even at low concentrations of venetoclax (Figure 4.1A).

We next treated primary cells from 12 AML patients with increasing concentrations of venetoclax and azacytidine. Unsupervised clustering on the cell viability measured after 72h of drug exposure revealed two clusters. Samples with <20% of pre-treatment cells expressing CD64 and CD11b clustered together and showed sensitivity to venetoclax and azacytidine (Figure 4.1B). In contrast, samples with >40% of pre-treatment cells expressing CD64 and CD11b showed resistance to the treatment and formed the second cluster (Figure 1B). Thus, monocytic differentiation of bulk AML cells was associated with *ex vivo* resistance to venetoclax and azacytidine treatment, supporting the findings in cell lines. Henceforward, AML samples are termed either Prim-AML or Mono-AML based on the percentage of $CD64^+CD11b^+$ cells, with the label cutoffs obtained from the unsupervised clustering (<20% and >40%, respectively).

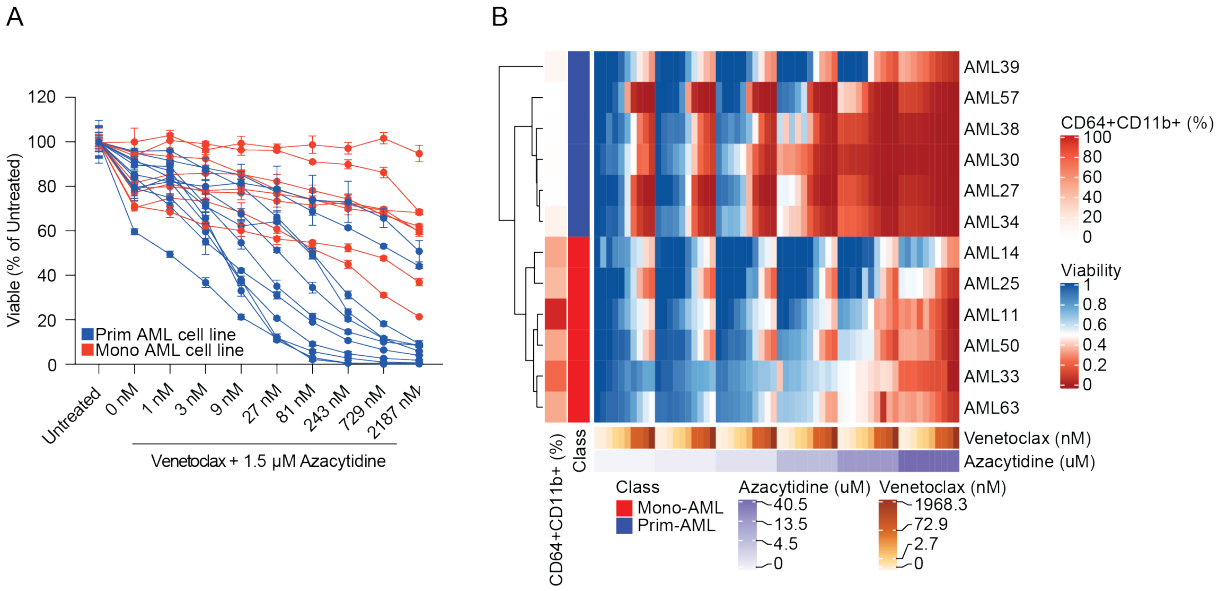


Figure 4.1 *In vitro* and *ex vivo* exposure of AML cells to venetoclax and azacytidine. A 19 AML cell lines classified as primitive (Prim-AML, n = 11) or monocytic (Mono-AML, n = 8) based on CD64 surface expression were treated *in vitro* with 1.5 µM of azacytidine and increasing concentrations of venetoclax for 72 hours. Representative data of two independent replicates. Mean ± SEM of technical replicates. **B** Mononuclear cells of AML patients (N = 12) were treated *ex vivo* for 72h on a drug matrix with increasing venetoclax and azacytidine concentrations. Unsupervised clustering was performed based on viability. Each quadrat represents one well with a specific venetoclax and azacytidine combination on the drug matrix. Experiments performed jointly by Dr. Alexander Waclawiczek and myself.

4.2.2 Clinical response of AML to venetoclax and azacytidine is independent of monocytic differentiation

To determine the clinical relevance of our findings, Dr. Simon Renders retrospectively analyzed a cohort of 54 newly-diagnosed AML patients who received venetoclax and azacytidine at Heidelberg University Hospital between 2019 and 2022 (Table 4.1). Out of the 54 patients, 14 (26%) did not achieve remission. He assessed parameters potentially associated with refractoriness to venetoclax and azacytidine treatment, including age, sex, previous MDS/MPN or previous HMA treatment, European LeukemiaNet risk group, genetics and percentage of $CD64^+$ cells. Using univariate logistic regression analysis, previous MDS/MPN, adverse risk based on European LeukemiaNet classification and complex karyotype were the only statistically significant factors predicting risk of refractory disease (Table 4.1). However, none of the assessed variables remained as independent predictors in the multivariate analysis. Importantly, percentage of $CD64^+$ cells did not predict therapy resistance, contradicting the *ex vivo* findings in bulk AML cells (Table 4.1). Taken together, these data show that monocytic differentiation based on CD64 surface staining was not a reliable predictor of clinical response to venetoclax and azacytidine in patients.

Table 4.1 Characteristics of patients treated first-line with venetoclax and azacytidine. Analysis of data performed by Dr. Simon Renders.

Patient characteristics	Total (n = 54)	Univariate ORR (CI 95%)	p-value	ORR (CI 95%)	p-values
Age (range)	70.2 (44-83)	1.05 (0.98 - 1.14)	0.18		
Sex (female)	17 (31.5%)	1.97 (0.51 - 9.8)	0.35		
Previous:					
- MDS/MPN	17 (31.5 %)	4.59 (1.3 - 17.6)	0.02	2.6 (0.54 - 12.8)	0.23
- HMA treatment	7 (13.0 %)	4.93 (0.95 - 28.7)	0.06	3.1 (0.43 - 23.4)	0.25
ELN good	12 (22.2 %)	0.20 (0.01 - 1.2)	0.14		
ELN intermediate	11 (20.4 %)	0.44 (0.06 - 2)	0.33		
ELN adverse	29 (53.7 %)	4.48 (1.2 - 22.1)	0.04	1.8 (0.28 - 12)	0.54
Genetics:					
Complex	16 (29.6 %)	5.33 (1.5 - 20)	0.01	3.8 (0.75 - 23)	0.12
NPM1	12 (22.2 %)	0.23 (0.01 - 1.4)	0.18		
IDH1/2	15 (27.8 %)	0.14 (0.008 - 0.83)	0.07	0.17 (0.008 - 1.2)	0.13
RUNX1	11 (20.4 %)	1.09 (0.21 - 4.57)	0.91		
N-/KRAS/PTPN1	9 (16.7 %)	0.79 (0.11 to 3.8)	0.78		
ASXL1	8 (14.8 %)	1.17 (0.15 - 6.2)	0.86		
TP53	8 (14.8 %)	3.6 (0.74 - 17.9)	0.11		
CD64+ blasts					
0-50%	43 (79.6 %)	0.53 (0.13 - 2.4)	0.38		
50-100%	11 (20.4 %)	1.89 (0.42 - 7.7)	0.38		

Abbreviations: HMA, Hypomethylating agent; ELN; European LeukemiaNet

4.2.3 LSCs are enriched in *Immature* GPR56$^+$ cells in Mono-AMLs and Prim-AMLs

Due to the discrepancy between clinical and pre-clinical data, I together with Dr. Alexander Waclawiczek and Dr. Simon Renders hypothesized that AML cells within each patient may differ in their importance for driving the disease. Thus, to identify the disease-driving LSCs we aimed to dissect 72 diagnostic AML patient samples into subpopulations using the previously introduced myeloid differentiation markers CD64 and CD11b. Using these markers, we identified two predominant cell populations: *Mature* cells and *Immature* cells. *Mature* cells had a CD64$^+$CD11b$^+$ cell surface expression and were the most prominent cells in Mono-AMLs, but were only present at a lower percentage in Prim-AMLs (40-97.6% and 0.1-20% of leukemic blasts, respectively) (Figure 4.2A). To further dissect the *Immature* population and enrich for disease-driving LSCs in CD34$^+$ and CD34$^-$ AMLs, we included the G-protein-coupled receptor GPR56 into our analysis (Figure 4.2A). High GPR56 expression has been associated with adverse outcome in AML and has been shown to enrich for functional LSCs[45,205,206]. Moreover, it is universally expressed in AMLs, irrespective of CD34 expression[205]. Using this combination of cell surface markers, we assessed the percentages of CD64$^+$CD11b$^+$ *Mature*, GPR56$^-$

Immature (*non-LSCs*) and GPR56^{+} *Immature* (*LSC-like*) cells in our cohort of 72 diagnostic AMLs (Figure 4.2B). All examined samples including Prim-AMLs and Mono-AMLs harbored a GPR56^{+} *Immature, LSC-like* population that ranged from 0.4% to 92.6% (Figure 4.2B).

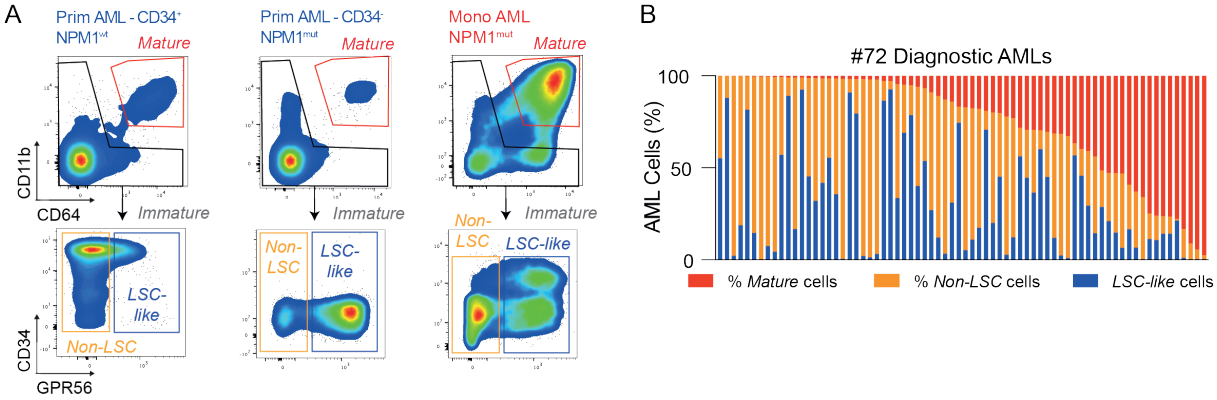


Figure 4.2 Phenotypic characterization of diagnostic AMLs. A FACS gating strategy for *Mature*, *non-LSC* and *LSC-like* subpopulations. Displayed are AML bulk cells from primitive CD34^{+} (*NPM1*-wildtype), primitive CD34^{-} (*NPM1*-mutated) and monocytic (*NPM1*-mutated) samples. **B** Percentages of *Mature*, *non-LSC* and *LSC-like* populations among bulk AML cells in 72 diagnostic AML samples sorted by the frequency of the *Mature* population. Experiments performed jointly by Dr. Alexander Waclawiczek and myself.

To confirm that the GPR56^{+} *Immature LSC-like* cells were predominantly enriched for functional LSCs compared to the other subpopulations, we FACS-sorted cells into *Mature*, *non-LSC* and *LSC-like* populations from 14 patient samples (Prim-AML: n = 7, Mono-AML: n = 7). Dr. Alexander Waclawiczek together with Dr. Simon Renders injected the different populations into NSG mice and assessed the leukemia re-initiating potential (Figure 4.3C). In 14/14 AML patient samples the *LSC-like* cells engrafted the mice with leukemic cells as evident by the predominant CD45^{+}CD33^{+} engraftment. In contrast, in only 2/14 of the patient samples the *Mature* and *non-LSC* cells gave rise to relevant percentages of leukemic engraftment (Figure 4.3D-E). Importantly, *LSC-like* cells between Prim-AMLs and Mono-AMLs showed no differences in their leukemia re-initiating potential as both AML classes showed preferable engraftment of the *LSC-like* cells.

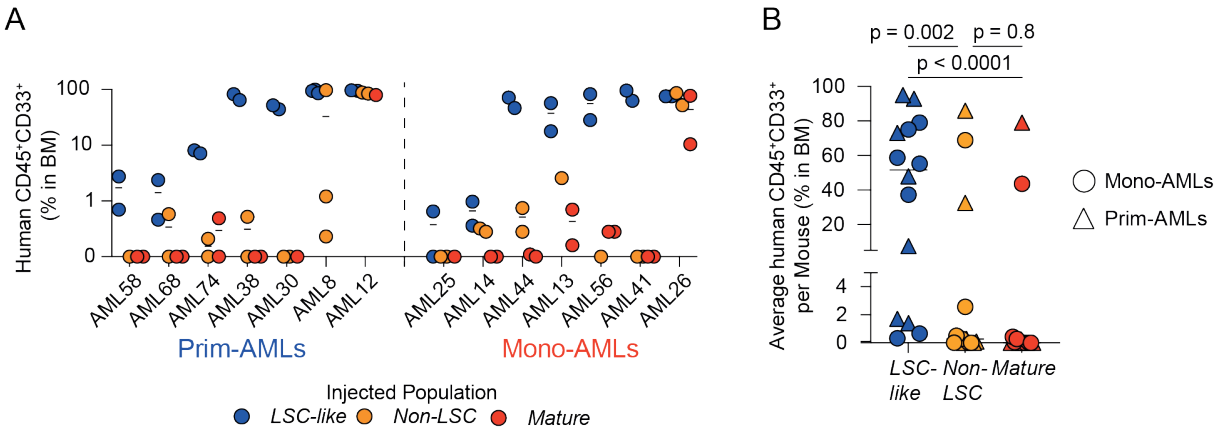


Figure 4.3 Leukemia re-initiation potential in phenotypically-defined AML subpopulations. A Percentage of human leukemic engraftment in bone marrow of NSG mice obtained from *Mature*, *non-LSC* and *LSC-like* subpopulations of 14 AML samples at endpoint. Each dot represents an individual mouse with the line marking mean engraftment levels. **B** Mean percentage of human engraftment in bone marrow per NSG mouse obtained from *Mature*, *non-LSC* and *LSC-like* subpopulations of 14 AML samples at endpoint. Each dot represents an individual AML patient with the line marking mean engraftment levels. Friedmann test was used to compare *LSC-like* with *non-LSC* and *Mature* subpopulations. Abbreviations: BM, bone marrow. Experiments performed by Dr. Alexander Waclawiczek and Simon Renders.

4.2.4 *LSC-like* cells are enriched for stemness-associated molecular programs

As Prim-AMLs and Mono-AMLs showed similar leukemia re-initiating potential in the different populations, we aimed to find other features in the *LSC-like* and *Mature* cells that would differ between the two AML classes. I performed RNA sequencing on 23 AML patients sorted for *LSC-like* and *Mature* cells and *non-LSC* cells when available. Based on dimensionality reduction using principal component analysis (PCA), the samples clustered according to population (*LSC-like* and *Mature*) but not according to AML class (Prim-AML and Mono-AML) (Figure 4.4A). For selected samples I also analyzed the *non-LSCs* that clustered either with the *LSC-like* or *Mature* cells. I performed differential gene expression analysis between these populations which revealed that in *LSC-like* cells several known cancer stem cell markers were upregulated, such as *ERG*, *PROM1* and *GPR56*[45,206-210] (Appendix Figure 13A-B). *Mature* cells showed upregulation of monocytic markers, including *S100A9*, *S100A8* and *CD14*[211-213] (Appendix Figure 13A-B). In line with these results, stemness-associated transcription was significantly higher in *LSC-like* cells compared to *Mature* cells, irrespective of the AML class. These findings support the preferential engraftment of the *LSC-like* cells over *Mature* cells also on transcriptomic level (Figure 4.4B). As expected, *LSC-like* cells between the two AML classes revealed only 59 differentially expressed genes, highlighting that *LSC-like* cells from Prim-AMLs and Mono-AMLs are transcriptionally very similar.

Even though I did not detect stark differences in the *LSC-like* cells between the two AML classes, different *LSC-like* samples showed subtle differences based on the PCA. I hypothesized that genetics would explain these differences, as seen in clustering of bulk AML samples based on mutations[214]. Indeed, *RUNX1*-mutated *LSC-like* cells and *NPM1*-mutated *LSC-like* cells clustered separately (Figure 4.4C), suggesting that transcriptomic clustering of LSCs is not determined by the differentiation state of the blast progeny but rather largely by the underlying mutational profile. A larger cohort of *LSC-like* samples is needed to further strengthen this observation.

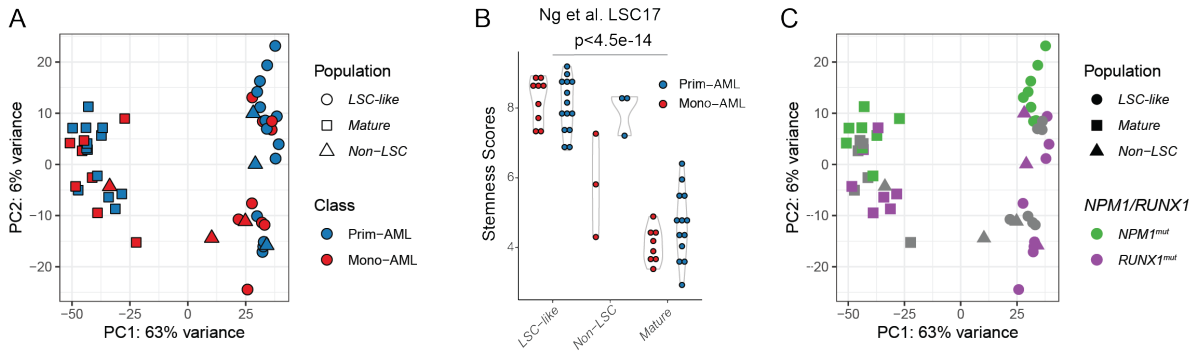


Figure 4.4 Transcriptomics of Prim-AML and Mono-AML. A PCA plot of bulk RNA-seq data from *LSC-like*, *non-LSC* and *Mature* subpopulations from Prim-AML (n = 14) and Mono-AML (n = 9). Each dot represents a subpopulation from one AML sample with samples annotated based on subpopulation and AML subclass. **B** LSC17 score in *LSC-like*, *non-LSC* and *Mature* subpopulations from Prim-AML (n = 14) or Mono-AML (n = 9) patient samples. LSC17 score was calculated for each AML sample as the mean expression of the 17 genes part of the LSC17 signature[45]. **C** PCA plot of bulk RNA-seq data from *LSC-like*, *non-LSC* and *Mature* subpopulations from Prim-AMLs (n = 14) and Mono-AMLs (n = 9) annotated based on mutation and subpopulation. Two-Way ANOVA with Type-III sums of squares was used to evaluate the effect of population and class on stemness score. Each dot represents an AML patient sample. Abbreviations: PC, Principal component.

4.2.5 *LSC-like* cells predominantly express *BCL2*/BCL-2 not *MCL1*/MCL-1

As venetoclax targets the apoptotic pathway by inhibiting BCL-2, I next focused on the gene expression of the BCL-2 family members to assess differences between *LSC-like* and *Mature* cells. Based on unsupervised clustering on the BCL-2 family member gene expression, the samples clustered again according to population (*LSC-like* and *Mature*) but not according to AML class. *Mature* cells showed higher expression of genes encoding for several pro-apoptotic proteins, including *BAK1, BAX, BMF* and *BBC3* (encoding PUMA), whereas *LSC-like* cells showed higher expression of genes encoding for anti-apoptotic proteins *BCL2* and *BCL2L2* (encoding BCL-w) (Figure 4.5A).

Comparison of the gene expression levels of the genes encoding for the three main anti-apoptotic proteins *BCL2*, *MCL1* and *BCL2L1* (encoding BCL-xL) showed largely population-specific expression. *BCL2* was 4.7-fold higher expressed in the *LSC-like* cells compared to *Mature* cells whereas *MCL1* was 2.3-fold higher expressed in the *Mature* cells (Figure 4.5B). *BCL2L1* showed no clear differences between the populations (Figure 4.5B). Of note, I did not detect differences in expression of the anti-apoptotic proteins between Prim-AMLs and Mono-AMLs, highlighting population-specific rather than AML class-specific expression patterns. Moreover, these data suggest that *LSC-like* cells express higher levels of the anti-apoptotic *BCL2* and lower levels of *MCL1*, hinting towards a higher dependency on BCL-2.

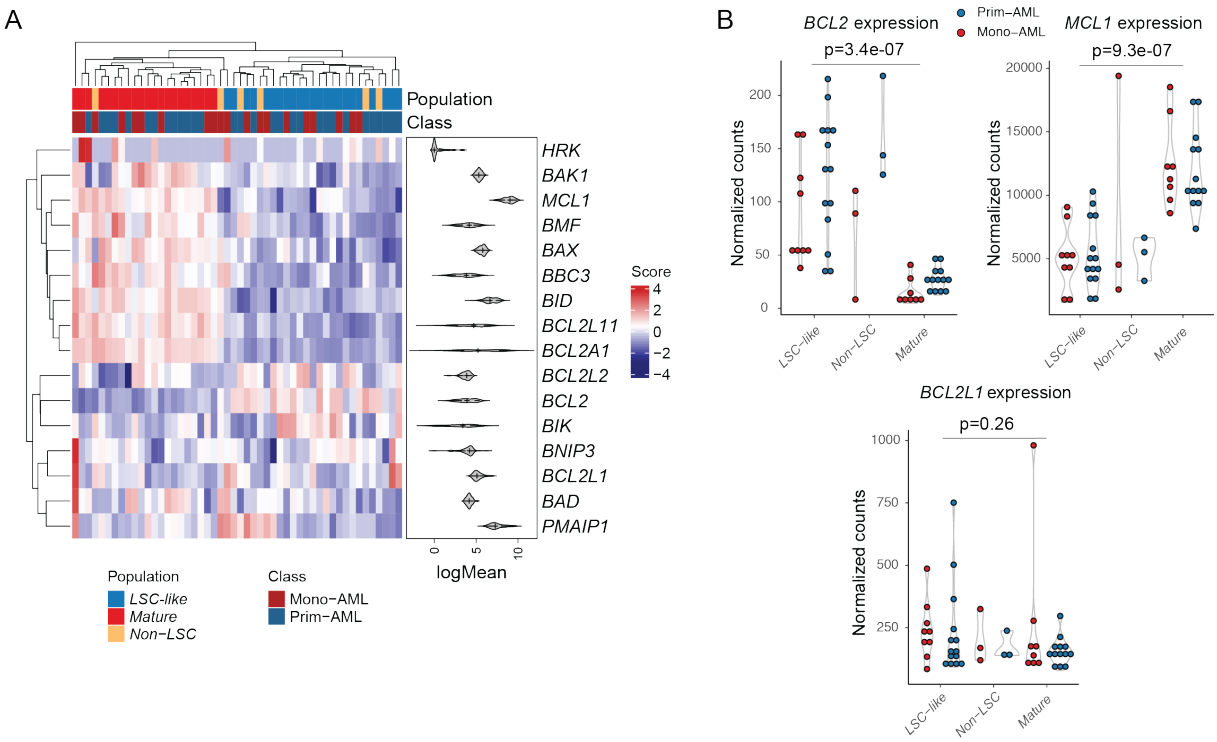


Figure 4.5 Subpopulation-specific BCL-2 family gene expression in AML. Heatmap of apoptosis regulators in *LSC-like, non-LSC* and *Mature* subpopulations from Prim- and Mono-AMLs with violin plots showing the distribution of the individual genes in log-scale. Expression levels of the individual genes were calculated from normalized counts. **B** Normalized counts of *BCL2*, *MCL1* and *BCL2L1* expression in *LSC-like*, *non-LSC* and *Mature* subpopulations from Prim-AML (n = 14) or Mono-AML (n = 9) patient samples. Two-Way ANOVA with Type-III sums of squares was used to evaluate the effect of population and class on gene expression. Each dot represents an AML patient sample.

To confirm the gene expression on the protein level, Dr. Alexander Waclawiczek and I established an intracellular staining protocol to measure intracellular protein expression of BCL-2, MCL-1 and BCL-xL together with cell surface protein expression, enabling the assessment of the anti-apoptotic protein levels in different populations. When we focused on

un-fractioned total leukemic cells, Prim-AMLs showed a 1.7-fold higher expression of BCL-2, whereas Mono-AMLs had a 2.1-fold higher expression of MCL-1 (Figure 4.6A-B). This suggests a lower dependency of BCL-2 in Mono-AMLs and a higher dependency in Prim-AMLs. However, assessment of the protein levels in pre-gated populations abrogated these differences and revealed a 2.7-fold higher expression of BCL-2 in *LSC-like* cells compared to *Mature* cells, irrespective of AML class (Figure 4.6A). Similarly, MCL-1 was 2.1-fold higher expressed in *Mature* cells compared to *LSC-like* cells, again irrespective of AML class (Figure 4.6B). BCL-xL showed no clear population-specific nor AML class-specific expression differences (Figure 4.6C). The discrepancy between total and population-level findings could be explained by the relative distributions of *LSC-like* and *Mature* cells in the two AML classes, as described in 4.2.3. Prim-AMLs contained a higher frequency of *LSC-like* cells and Mono-AMLs a higher frequency of *Mature* cells, blurring the population-specific expression when assessed in bulk. Collectively, assessment of the anti-apoptotic BCL-2, MCL-1 and BCL-xL protein levels extends the findings of the transcriptional data, and further highlights the similarities between *LSC-like* cells in Prim-AMLs and Mono-AMLs.

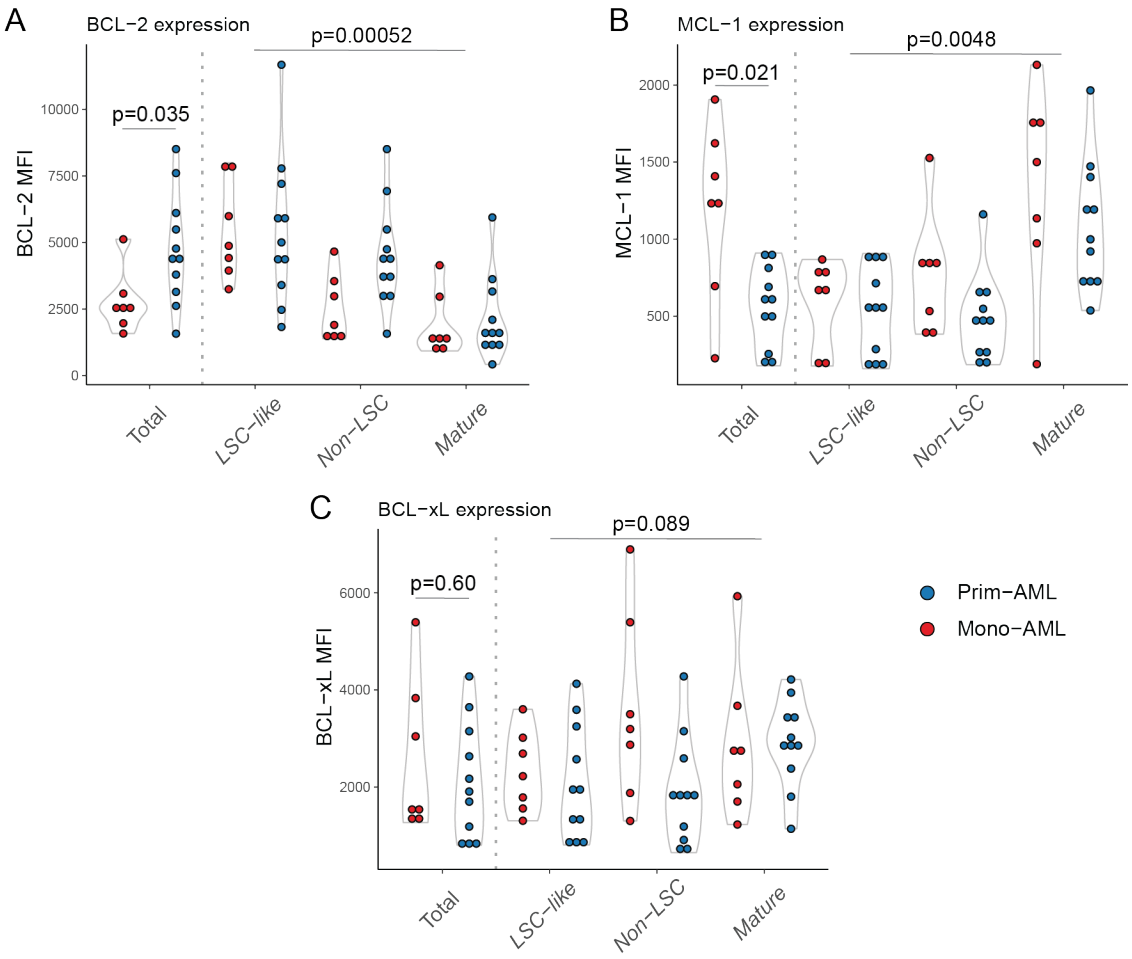


Figure 4.6 Subpopulation-specific BCL-2 family protein expression in AML. A-C MFI of **A** BCL-2, **B** MCL-1 and **C** BCL-xL in total AML cells and in *LSC-like, non-LSC* and *Mature* subpopulations from Prim-AML (n = 11) or Mono-AML (n = 7) patient samples. Each dot represents an AML patient sample. Wilcoxon test was used to compare the subclasses in total AML and Two-Way ANOVA with Type-III sums of squares was used to evaluate the effect of population and class on protein expression. Experiments performed jointly by Dr. Alexander Waclawiczek and myself.

4.2.6 Venetoclax and azacytidine exposure eradicates *LSC-like* but spares *Mature* cells in Prim-AML and Mono-AML

As the transcriptomic and protein-level findings revealed a higher expression of the anti-apoptotic *BCL2*/BCL-2 in *LSC-like* cells, Dr. Alexander Waclawiczek and I hypothesized that *LSC-like* cells would show vulnerability to BCL-2-specifc targeting with venetoclax. We exposed cells from 18 treatment-naïve diagnostic AML patients to the clinically-approved combination of venetoclax and azacytidine for 24h *ex vivo* and analyzed the cell viability using flow cytometry. Prim-AMLs showed significantly higher response to the treatment compared to Mono-AMLs that were largely resistant, when un-fractioned bulk AML cells were assessed (Appendix Figure 14). Similar to the anti-apoptotic protein analysis, assessment of cell viability in pre-gated populations abrogated these differences and revealed an efficient reduction of *LSC-like* cells in both AML classes (90 +/- 10.6% in Prim-AML and 79 +/- 19% Mono-AML) (Figure 4.7A). In contrast, *Mature* cells were highly resistant to the treatment with only marginal reduction in cell viability (40 +/- 32% in Prim-AML and 22.5 +/- 11.5% in Mono-AML) (Figure 4.7A). These data show that the differences in un-fractioned bulk AML sensitivity between Prim-AML and Mono-AML are largely driven by the initial difference in proportions of the treatment-resistant *Mature* population. Importantly, the *Mature* cells lacked consistent disease initiating potential, as described earlier (Figure 4.3). Moreover, the disease-driving *LSC-like* cells were efficiently eliminated by venetoclax and azacytidine treatment *ex vivo* in both AML classes.

To assess whether the *ex vivo* effect of venetoclax and azacytidine treatment could be recapitulated in patients, Dr. Simon Renders collected peripheral blood mononuclear cells (PBMCs) from three AML patients at the start (Day 0) and during venetoclax and azacytidine treatment (Days 1-6). At each collection day, I together with Dr. Alexander Waclawiczek assessed the absolute number and the proportion of different cell populations in the PBMCs. All three patients showed a reduction in the absolute number as well as the proportion of *LSC-like* cells within 24h from the start of the treatment with a continued decrease throughout the

length of the collection period (Figure 4.7B). In contrast, *Mature* cells and *non-LSC* cells persisted in all three patients during venetoclax and azacytidine treatment (Figure 4.7B).

Encouraged by these findings, we evaluated whether *ex vivo* treatment with venetoclax and azacytidine could be used to predict clinical response to the treatment. We exposed cells from 24 diagnostic patients treated first-line with venetoclax and azacytidine to the same compounds *ex vivo* for 24h, and assessed the cells by flow cytometry. As expected, *LSC-like* cells showed the strongest reduction in viability compared to *non-LSC* and *Mature* cells (Figure 4.7C). Importantly, *LSC-like* cells from patients refractory to the first-line treatment with venetoclax and azacytidine were significantly more resistant also to the *ex vivo* treatment (Figure 4.7C). Collectively, these findings show that the disease-driving *LSC-like* cells can be targeted *ex vivo* as well as in patients by BCL-2-specific inhibition with venetoclax in combination with azacytidine. These data further highlight the importance of assessing the disease-driving cells and raises the potential for identifying biomarkers specific for this groups of cells.

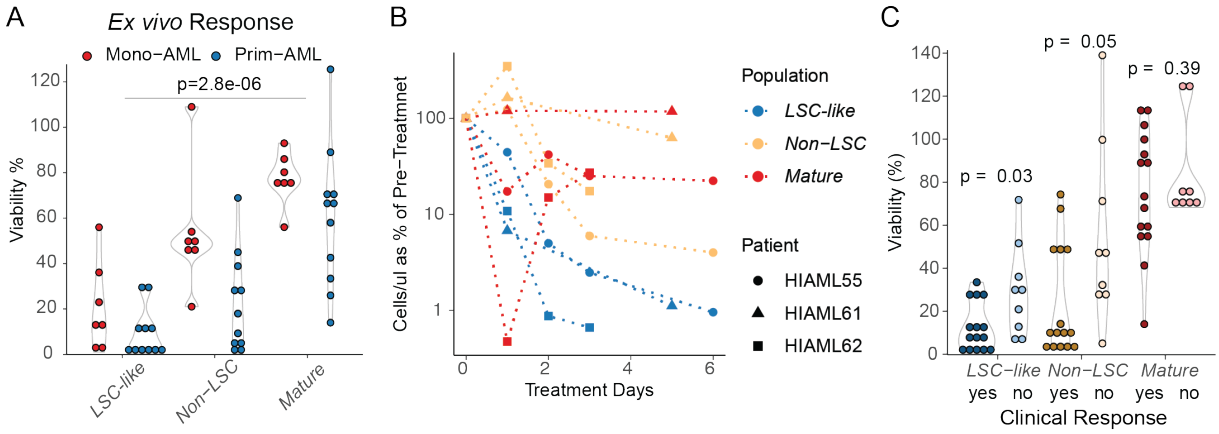


Figure 4.7 Targeting *LSC-like* cells *ex vivo* and in patients. A Relative viability of *LSC-like, non-LSC* and *Mature* subpopulations from Prim-AML (n = 11) or Mono-AML (n = 7) patient samples after 24h venetoclax and azacytidine treatment *ex vivo.* Viabilities were normalized to untreated controls. **B** Quantification of *LSC-like, non-LSC* and *Mature* cell counts from PBMCs relative to pre-therapy in the first week of venetoclax and azacytidine treatment in three patients undergoing therapy initiation. Each dotted line represents an individual patient with each shape on the line representing an individual timepoint of the patient. **C** Relative viability of *LSC-like, non-LSC* and *Mature* subpopulations after 24h venetoclax and azacytidine treatment *ex vivo* from patients who achieved ("yes") or did not achieve ("no") clinical response to first-line treatment with venetoclax and azacytidine. Each dot represents an AML patient sample. Two-Way ANOVA with Type-III sums of squares was used to evaluate the effect of population and class on viability and Wilcoxon test was used to compare the clinical response groups. Experiments performed jointly by Dr. Alexander Waclawiczek, Simon Renders and myself.

4.2.7 Rapid prediction of *ex vivo* response by MAC-Score in *LSC-like* cells

As venetoclax specifically targets the anti-apoptotic protein BCL-2, I together with Dr. Alexander Waclawiczek and Dr. Simon Renders hypothesized that BCL-2 family protein expression levels in the disease-driving *LSC-like* cells could be used to predict response to venetoclax-based therapies. We performed *ex vivo* venetoclax and azacytidine treatment with parallel intracellular staining of BCL-2, MCL-1 and BCL-xL in 54 diagnostic AML patient samples. AML samples with higher intracellular BCL-2 expression scores in *LSC-like* cells were more sensitive to *ex vivo* treatment while MCL-1 and BCL-xL scores showed no clear association with resistance or sensitivity (Figure 4.8A-C).

Similarly, others have also shown that BCL-2 conveys sensitivity to venetoclax, whereas MCL-1 and BCL-xL can promote survival independent of BCL-2[215]. Thus, to additionally consider the factors contributing resistance, we incorporated all three proteins into a singular response score, and termed it the "Mediators of Apoptosis Combinatorial Score" (MAC-Score). MAC-Score calculates the ratio between the normalized MFI of the drug target (BCL-2) and the normalized MFI of the resistance factors (sum of MCL-1 and BCL-xL) as follows: $BCL\text{-}2^{Norm.\ MFI}/(MCL\text{-}1^{Norm.\ MFI} + BCL\text{-}xL^{Norm.\ MFI})$ (for details see Methods 6.2.5). Furthermore, the MAC-Score can be calculated for different populations defined by flow cytometry, enabling the prediction of response specifically in disease-driving cells.

The MAC-Score was able to improve the separation between AML samples stratified based on the viability of *LSC-like* cells following venetoclax and azacytidine treatment *ex vivo*, outperforming the assessment of individual protein levels in the same population (Figure 4.8A-D). These data highlight the benefit of accounting for resistance factors of venetoclax in addition to its target protein levels when predicting *ex vivo* response.

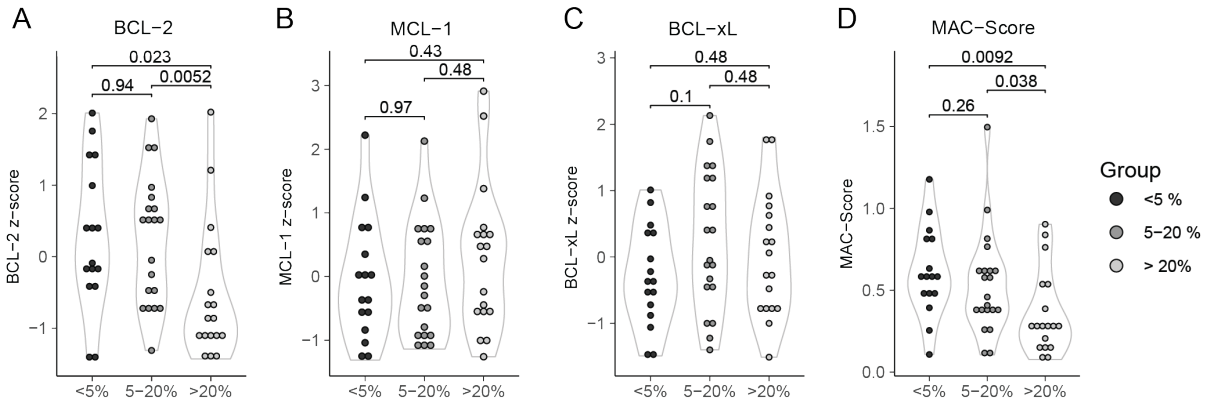


Figure 4.8 Prediction of *ex vivo* response to venetoclax and azacytidine based on BCL-2 family protein levels and MAC-Score in *LSC-like* cells. A-D Treatment-naïve AML patient samples were stratified based on *ex vivo* cell viability *in LSC-like* cells after 24h treatment with venetoclax and azacytidine and plotted for pre-culture **A** BCL-2, **B** MCL-1 or **C** BCL-xL MFI-based z-scores or **D** MAC-Score calculated based on normalized BCL-2 family protein expression levels in *LSC-like* cells. Each dot represents an AML patient sample. Wilcoxon test was used to compare the groups. Experiments were performed jointly by Dr. Alexander Waclawiczek and myself.

4.2.8 Rapid and robust prediction of clinical response and remission duration by MAC-Score in *LSC-like* cells

Encouraged by the prediction results of the *ex vivo* response, I together with Dr. Alexander Waclawiczek and Dr. Simon Renders further assessed the use of BCL-2 family protein levels in predicting clinical response to venetoclax and azacytidine. We measured the expression of the anti-apoptotic proteins BCL-2, MCL-1 and BCL-xL together with a set of cell surface markers in 35 old or frail diagnostic AML patients treated first-line with venetoclax and azacytidine. The samples originated from two independent multicenter cohorts processed separately (Cohort 1: n = 17, Cohort 2: n = 18). Supporting the *ex vivo* results, expression of BCL-2 in *LSC-like* cells was significantly higher in patients who achieved complete remission, complete remission with incomplete count recovery or morphologic leukemia-free state (responder), compared to patients who had a stable disease, partial response or progressive disease (non-responder) (Figure 4.9A). In contrast, MCL-1 and BCL-xL showed higher expression in non-responders (Figure 4.9A). While the individual proteins showed promise in predicting response, all three proteins showed high variation in their z-scores and did not clearly separate the responders from non-responders.

We next calculated the MAC-Score in the *LSC-like* cells of the two cohorts and detected an improved separation between responders and non-responders (Figure 4.9B). Patients who responded to venetoclax and azacytidine treatment had a higher MAC-Score compared to non-responders (Figure 4.9B). To verify that MAC-Score has the highest performance in *LSC-like* cells, we compared MAC-Scores calculated in *LSC-like*, *non-LSC*, *Mature* and un-fractioned bulk cells. As expected, MAC-Score had the best separation of responders and non-responders in *LSC-like* cells, again confirming the importance of assessing the disease-driving cells (Figure 4.9B).

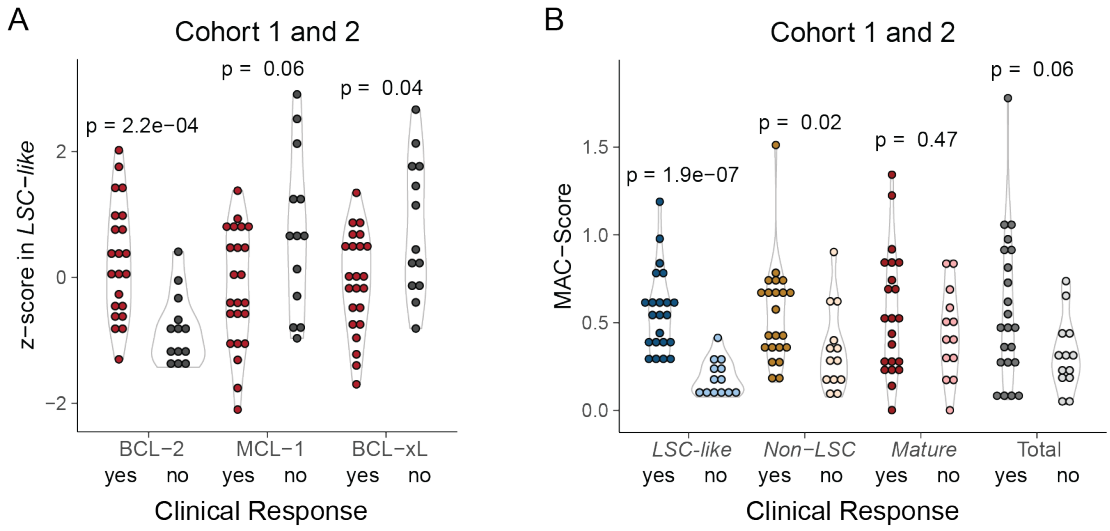


Figure 4.9 Prediction of clinical response to venetoclax and azacytidine based on BCL-2 family protein levels and MAC-Score. A Expression of BCL-2, MCL-1 and BCL-xL in *LSC-like* cells of AML patients from cohorts 1 and 2 combined who achieved ("yes") or did not achieve ("no") clinical response to first-line treatment with venetoclax and azacytidine. Protein expression shown as MFI-based z-scores. **B** MAC-Score in *LSC-like* cells of AML patients from cohorts 1 and 2 combined who achieved ("yes") or did not achieve ("no") clinical response to first-line treatment with venetoclax and azacytidine. Each dot represents an AML patient sample. Wilcoxon test was used to compare the groups. Experiments performed jointly by Dr. Alexander Waclawiczek, Simon Renders and myself.

Even though MAC-Score in *LSC-like* cells performed well as a binary predictor of clinical response to venetoclax and azacytidine, we further evaluated if it could also be used to assess the duration of response. We assessed the event-free survival (EFS) of the same 35 diagnostic AML samples and stratified them into two groups based on the median MAC-Score of the two cohorts (>0.4 and <0.4). Patients who discontinued treatment due to other reasons than disease progression were censored. We observed a significantly longer EFS in patients with an over-the-median MAC-Score in the *LSC-like* cells compared to patients with a below-the-median MAC-Score (Figure 4.10A). Assessment of EFS based on singular BCL-2, MCL-1 or BCL-xL expression in *LSC-like* cells did not reach the same predictive potential as MAC-Score, even though high BCL-2 expression and low MCL-1 expression showed longer EFS (Figure 4.10B-D). To validate these findings, we analyzed a third independently processed cohort (Cohort 3: n = 24) where again MAC-Score performed well as a binary predictor of venetoclax and azacytidine response and showed longer EFS in patients with a MAC-Score > 0.4, the median threshold defined based on the first two cohorts (Figure 4.10E-F).

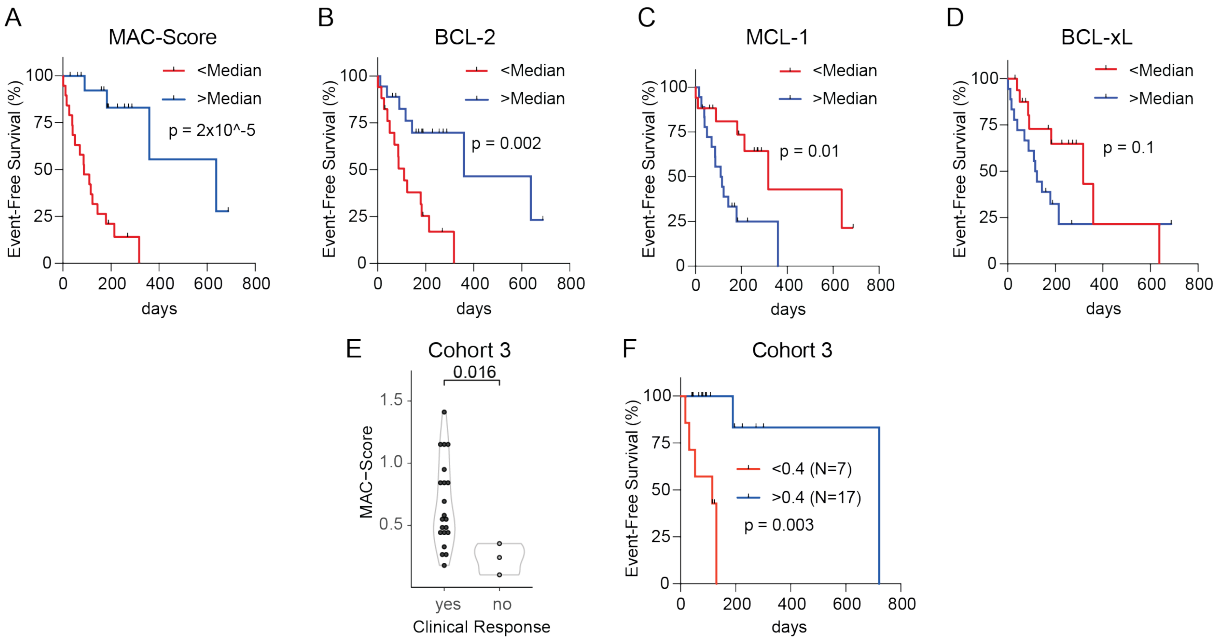


Figure 4.10 Evaluation of response duration using MAC-Score. A-D Event-free survival of AML patients treated first-line with venetoclax and azacytidine from cohorts 1 and 2 combined with above and below median **A** MAC-Score, **B** BCL-2 expression, **C** MCL-1 expression or **D** BCL-xL expression in *LSC-like* cells. **E** MAC-Score in *LSC-like* cells of AML patients from cohort 3 who achieved ("yes") or did not achieve ("no") clinical response to first-line treatment with venetoclax and azacytidine. Each dot represents an AML patient sample. **F** Event-free survival of AML patients treated first-line with venetoclax and azacytidine from cohort 3 with above (>0.4) and below (<0.4) median MAC-Score based on cohorts 1 and 2. Wilcoxon test was used to compare the groups and log-rank test to compare therapy durations of AML patients. Experiments performed jointly by Dr. Alexander Waclawiczek, Simon Renders and myself.

To further evaluate the prognostic accuracy of MAC-Score, we combined the three cohorts of diagnostic AML patients treated first-line with venetoclax and azacytidine, increasing the cohort size to 59 patients. As expected, MAC-Score performed well as a binary predictor of clinical response with responders showing significantly higher scores than non-responders (Figure 4.11A). Receiver operating characteristic curve analysis revealed a C statistic of 0.95, indicating a high accuracy to predict venetoclax and azacytidine response using MAC-Score (Figure 4.11B). The combined EFS analysis of all 59 first-line treated patients, showed a 4-fold prolongation of EFS in patients with > 0.4 MAC-Score, extending the median EFS from 3 months to 12 months (Figure 4.11C). Moreover, EFS analysis within patients who initially responded to venetoclax and azacytidine treatment also revealed differences in response duration. Initial responders with <0.4 MAC-Score had a significantly shorter EFS compared to initial responders with >0.4 MAC-Score (Figure 4.11D).

Finally, as venetoclax together with azacytidine is more often given as salvage therapy rather than first-line therapy, we assessed whether MAC-Score in *LSC-like* cells could also be used to predict response in young, fit patients refractory to initial induction therapy. We evaluated the MAC-Score in this fourth cohort consisting of 23 refractory AML patients with material available from the time of diagnosis or before the start of venetoclax and azacytidine treatment. Similar to findings in the first-line treated patients, MAC-Score was again able to separate responders from non-responders, with responders showing higher MAC-Scores (Figure 4.11E). Moreover, EFS was also significantly longer in patients with a MAC- Score > 0.4, the median threshold defined based on the first two first-line-treated cohorts (Figure 4.11F).

Taken together, MAC-Score in *LSC-like* cells can be used as a robust biomarker for binary response to venetoclax and azacytidine treatment in patients receiving first-line as well as salvage therapy. Moreover, it can be used to identify patients with long-lasting response, making MAC-Score a promising biomarker towards selecting the best therapy for each patient on an individual basis.

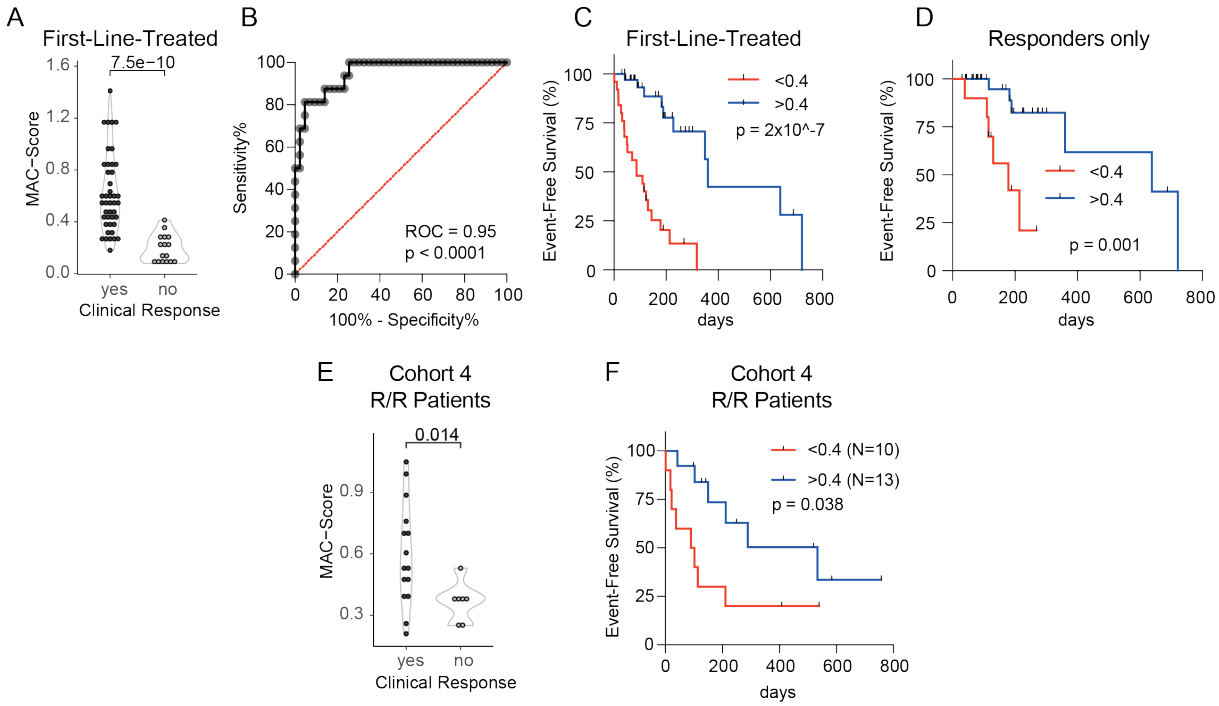


Figure 4.11 MAC-Score predicts response to venetoclax and azacytidine in patients receiving first-line and salvage therapy. A MAC-Score in *LSC-like* cells of AML patients from all first-line-treated patients combined (cohorts 1, 2 and 3) who achieved ("yes") or did not achieve ("no") clinical response to first-line treatment with venetoclax and azacytidine. **B** Receiver operating characteristic (ROC) curve of MAC-Score and therapy outcomes of all first-line-treated venetoclax and azacytidine AML patients combined (cohorts 1, 2 and 3). **C** Event-free survival of AML patients treated first-line with venetoclax and azacytidine from all first-line-treated patients

combined (cohorts 1, 2 and 3) with above (>0.4) and below (<0.4) median MAC-Score determined based on cohorts 1 and 2. **D** Event-free survival of all AML patients from combined cohorts (cohorts 1, 2 and 3) who achieved complete remission to first-line venetoclax and azacytidine treatment with above (>0.4) and below (<0.4) median MAC-Score determined based on cohorts 1 and 2. **E** MAC-Score in *LSC-like* cells of AML patients from refractory/relapsed patients who achieved ("yes") or did not achieve ("no") clinical response to salvage treatment with venetoclax and azacytidine. **F** Event-free survival of refractory/relapsed AML patients treated with venetoclax and azacytidine as salvage therapy with above (>0.4) and below (<0.4) median MAC-Score determined based on cohorts 1 and 2. Wilcoxon test was used to compare the groups and log-rank test to compare therapy durations of AML patients. Abbreviations: ROC, Receiver operating characteristic; R/R, Refractory/Relapse. Experiments performed jointly by Dr. Alexander Waclawiczek, Simon Renders and myself.

4.3 Discussion II

In this chapter, I presented the collaborative efforts to shed light to the inconsistency between mechanistic and clinical findings in regards to monocytic differentiation and response to venetoclax and azacytidine treatment in AML. I revealed this inconsistency to be largely based on the different proportions of AML subpopulations and highlighted the importance to identify and assess the disease-driving LSCs. I also introduced a BCL-2 family protein-based biomarker score termed MAC-Score that can be used in the disease-driving LSCs to predict binary clinical response and response duration to venetoclax and azacytidine treatment in AML patients.

As already shown and discussed in **Chapter 3**, assessing LSCs is crucial for the elimination of the disease-driving cells[50]. LSCs have been considered the source of therapy resistance and relapse in part due to their ability to adapt to the changing environment following chemotherapy using different strategies. These include, for example, genomic instability as shown in **Chapter 3**, the entrance into a senescent or dormant-like cell state or metabolic re-wiring[48,50,51]. The results of this chapter highlight that the understanding of LSC features present already at diagnosis are also important for prediction of initial response to treatments such as venetoclax and azacytidine. Others have shown that LSCs from *de novo* AML patients rely on amino acid metabolism for increased oxidative phosphorylation and cell survival[48,216]. By comparison, the results of this chapter show that LSCs express high levels of BCL-2, which is further connected to BCL-2 dependency and sensitivity to venetoclax and azacytidine treatment[179].

Xenotransplantation assays are considered the golden standard to identify and assess functional LSCs. In this chapter I showed that cell surface expression of GPR56 in *Immature* cells could be used to enrich for engraftment-driving cells, which is in line with findings from others[45,205,206]. While this strategy resulted in the most consistent engraftment across all AML samples and served as a robust strategy, GPR56$^-$ *Immature* and *Mature* cells also showed engraftment in a few samples. While AML, similar to healthy hematopoiesis, is considered to be organized in a hierarchy with LSCs enriched at the apex[38,217], there has been a debate whether such a cancer stem cell model truly is that rigid and unidirectional[218]. Others have demonstrated plasticity in differentiated cells that reinvigorate their leukemia initiation potential by de-differentiating[218]. In contrast, treatment with venetoclax and azacytidine has also revealed strong selection of cells with more differentiated cell surface phenotypes with a parallel acquisition of stem-like transcriptional profiles upon relapse[200]. These data suggest that some

functional LSCs may also reside in the cells lacking GPR56 expression, and poses an interesting basis for future studies.

The high clinical relevance of the results described in this chapter culminated in the introduction of the BCL-2 family protein-based MAC-Score. While the correlation between high BCL-2 levels and response to BCL-2 targeting by venetoclax and azacytidine treatment seems clear, combinatorial reliance on other BCL-2 family members plays a central role for the control of apoptosis, as discussed in **Section 1.5**. Indeed, based on gene expression analyses, studies have suggested that the gene expression ratio of the BCL-2 family members rather than *BCL2* expression alone predicts efficacy to venetoclax-based treatment in lymphoid malignancies[219-221]. Moreover, increased mitochondrial priming in AML blasts mediated by the combination of different BCL-2 family members have been shown to predict clinical response to venetoclax-monotherapy and combination therapy with azacytidine[201,215]. These findings together with the work described in this chapter support and strengthen the rational of combining BCL-2, MCL-1 and BCL-xL levels for accurate prediction of venetoclax and azacytidine response in AML patients.

Predictive biomarkers that rely on genomic sequencing have enabled most of the recent advances in terms of biomarker-guided therapeutic decision[55]. In contrast, flow cytometry is a routine diagnostic tool for diagnosis and monitoring of AML and can support clinical decision-making within several hours after sampling. Here, MAC-Score allowed the assessment of BCL-2 family members in the disease-driving LSCs that were present in varying proportions in different AMLs. It could further be used to predict treatment response with high accuracy and could easily be implemented into routine diagnostics. By comparison, the turnaround time for alternative strategies to predict clinical response, such as RNA analysis, BH3-profiling or *ex vivo* drug screening[201,222], is more time consuming as well as resource and labor-intensive. These attributes make the implementation of these assays into routine diagnostics difficult. In contrast, MAC-Score provides a rapid and robust strategy to support clinical decision-making within several hours after sampling. Most importantly, MAC-Score has superior predictive accuracy compared to alternative methods[201,222]. Prospective studies to further validate the MAC-Score will provide more insights into its applicability to guide therapy in front-line and relapsed/refractory disease.

In summary, the work described in this chapter, builds on the results from **Chapter 3** by further highlighting the importance to identify and assess the disease-driving LSCs. Although only focusing on venetoclax and azacytidine treatment, this chapter describes the crucial role of LSCs and their features in initial therapy response. However, many AML patients eventually relapse whether treated with venetoclax and azacytidine or with conventional chemotherapy. To extend biomarker-guided personalized medicine beyond diagnosis, the identification and assessment of LSCs following treatment would be important. Future studies to locate and characterize disease-driving LSCs in serially collected AML samples from diagnosis, remission and relapse will be of interest for extended assessment of alternative treatment strategies.

Chapter 5 – Conclusions and Future Outlook

The work in this thesis outlined the efforts to identify, characterize and target the disease-driving LSCs in AML. Here, I presented two studies with differing approaches to study these cells. In **Chapter 3,** I highlighted the power and potential of using single-cell approaches to dissect the heterogeneity in CK-AML with the aim to identify and target the disease-driving LSCs. In **Chapter 4**, I showed that also more affordable and accessible techniques such as flow cytometry are of high value for predicting clinical treatment response and may be applied into routine diagnostics.

One of the key outcomes of the thesis was the assessment of longitudinal subclonal dynamics in CK-AML using an integrated single-cell multi-omics framework (**Chapter 3**). This revealed that genetically heterogenous CK-AML samples at the time of diagnosis show clonal outgrowth of minor subclones with complex genetic rearrangements when transplanted into immunocompromised mice. Importantly, the cell surface phenotype of the engraftment-driving cells at diagnosis could be used to identify and target these cells *ex vivo*. These findings are of high clinical relevance as they demonstrate the role of minor subclones in disease progression; a feature that cannot be assessed solely using bulk sequencing approaches. Furthermore, the ability to identify the disease-driving LSCs not only *in vivo* but also *ex vivo* opens up avenues for personalized medicine approaches in genetically heterogenous AMLs lacking effective therapeutic options.

This work paved way for using a novel single-cell multi-omics framework together with functional assays to characterize different levels of intra-patient heterogeneity in CK-AML and to further target the disease-driving LSCs. Indeed, the findings in this small cohort of CK-AML patient samples encourage a broader scale investigation into the genetic and non-genetic landscapes of CK-AML. Due to the high inter-patient heterogeneity between CK-AMLs, the study of commonalities between CK-AML patients was not feasible in this work. A larger cohort would make it possible to look for common genetic and molecular profiles between CK-AMLs, and to further stratify them into functional groups and identify common biological mechanisms that mediate therapy resistance

While the use of PDXs was a highly informative approach to model clonal dynamics longitudinally (**Sections 3.2.5**), paired patient samples from diagnosis and relapse would allow a more accurate study of longitudinal clonal dynamics in patients during disease progression.

Using paired patient samples, it would be of high interest to create evolutionary pathways for each patient and use these to outline how the original disease evolved in response to therapy stress. This would enable a detailed assessment of several interesting evolutionary aspects that shape cancer growth and development, including the degree of chromosomal instability, epigenetic and transcriptomic changes and shifts in cell differentiation states during disease progression. These data would not only be important for a better molecular understanding of CK-AML but would also be of clinical relevance for identification of novel targetable vulnerabilities in the disease-driving LSCs.

Another key outcome of the thesis was the development of the MAC-Score for prediction of clinical response to the BCL-2 inhibitor venetoclax in combination with azacytidine (**Chapter 4**). MAC-Score not only enabled a highly accurate prediction of binary clinical response but also response duration to the newly-approved therapy. Similar to the results in Chapter 3, the role of the disease-driving LSCs was the key in achieving high prediction accuracy. Prediction assessment in un-fractioned leukemic cells or non-disease driving cells showed worse or even lacked predictive value entirely. The findings in this study can have a direct translational relevance and help guide therapy decision between the highly toxic standard induction therapy and a more tolerable therapy with venetoclax and azacytidine. Future prospective studies will shed further light into the predictive potential of the MAC-Score.

While the assessment of initial response and response duration to venetoclax and azacytidine is important, patients who originally respond to BCL-2 inhibition-based treatment eventually relapse. Thus, an improved understanding of the mechanisms driving the disease progression is important. The findings of this work encourage a broader scale investigation of resistance factors of AML to venetoclax and azacytidine treatment. I together with Dr. Alexander Waclawiczek and Dr. Simon Renders have already started to work on this by transcriptionally characterizing a larger cohort of diagnosis AML samples and assessing their MAC-Scores in parallel. We hope that this will shed further light to what drives relapse to venetoclax and azacytidine treatment and whether we can identify ways to re-sensitize the resistant cells or find alternative treatment strategies. We have also begun analyzing paired samples from diagnosis and relapse AML samples from patients treated with venetoclax and azacytidine to better assess the clonal evolution during disease progression upon therapy stress.

Together, the two studies described in this thesis represent important advances in the field of AML to better identify, characterize and target the disease-driving LSCs; with a potential to be applied to other hematological malignancies and even solid cancers. The results described here will function as an outline to decipher how genetic heterogeneity evolves in highly rearranged genomes, what the role of ongoing karyotypic heterogeneity is, how it affects the function and phenotype of a cancer cell and why this leads to an aggressive treatment-resistant malignancy. Furthermore, the results pave way for flow cytometry-based prediction scores for therapies targeting BCL-2 family members to enter routine diagnostics.

Chapter 6 – Materials and Methods

6.1 Experimental Practices Related to Chapter 3

The text of the following chapter has been taken from Leppä et al. (manuscript in preparation) and was originally written by myself:

6.1.1 Primary AML patient samples

AML samples were collected from diagnostic bone marrow and/or peripheral blood aspirations at the University hospital in Heidelberg in accordance with the Declaration of Helsinki after obtaining written consent from each patient. The project was approved by the Ethics Committee of the Medical Faculty of Heidelberg (NCT-MASTER Platform S-206/2011). Bone marrow and peripheral blood mononuclear cells were isolated by density gradient centrifugation and stored in liquid nitrogen until further use. Patient characteristics are listed in Appendix Table 1.

6.1.2 Processing of primary AML cells for single-cell sequencing

Viably cryopreserved AML bone marrow and/or peripheral blood samples were thawed at 37 °C in Iscove's modified Dulbecco's medium (IMDM) containing 10% FBS, and treated with DNase I for 15min (100 μg/ml).

6.1.2.1 Strand-seq in primary leukemia cells

For Strand-seq analysis, recovered cells were cultured using previously established protocols[206,223] using IMDM, 15% BIT (bovine serum albumin, insulin, transferrin; Stem Cell Technologies, cat # 09500), 100 ng/ml SCF (PeproTech, cat # 300-07), 50 ng/ml FLT3-L (PeproTech, cat # 300-19), 20 ng/ml IL-3 (PeproTech, cat # 200-03), 20 ng/ml G-CSF (PeproTech, cat # 300-23), 100 μM β-mercaptoethanol (ThermoFisher, cat # 31350010), 500 nM SR1 (StemRegenin 1, STEMCELL Technologies, cat # 72342), 500 nM UM729 (STEMCELL Technologies, cat # 72332), and 1% penicillin-streptomycin (Sigma, cat # P4458-100ML). BrdU (40 μM; Sigma) was incorporated for the duration one cell division (52 h to 62 h) to perform non-template strand labeling. Single nuclei from the appropriate timepoint were sorted into 96-well plates using a BD FACSMelody cell sorter, followed by Strand-seq library preparation (described below).

6.1.2.2 CITE-seq in primary leukemia cells

For combined single-cell RNA and antibody derived tag sequencing (CITE-seq) analysis, recovered cells were stained with a total of 38 antibody-derived tags (ADT) (Table 6.1), and sorted for live CD45+ cells using a BD FACSAria III cell sorter. CITE-seq library preparation was performed as previously reported[105] using the Chromium Single Cell 3' Library and Gel Bead Kit (10X Genomics, cat # 1000128). Five to ten thousand cells were targeted for each sample, and processed according to the manufacturer's instructions up until the cDNA amplification step (10X Genomics). 0.2μM of ADT additive oligonucleotides were spiked into the cDNA amplification PCR (13 cycles). Following PCR, the large cDNA fraction derived from cellular mRNAs (retained on beads) was separated from the ADT (in supernatant) using 0.6X SPRI. The cDNA fraction was processed according to the 10X Genomics Single Cell 3′ v3.1 protocol to generate the transcriptome libraries. To generate the ADT libraries, ADTs were indexed with Truseq Small RNA RPIx primers by PCR for 10 cycles, followed by library purification and reamplification for five additional cycles with P5/P7 generic primers. ADT and scRNA-seq libraries were pooled in a ratio of 25% ADT and 75% RNA and sequenced together on an Illumina NovaSeq 6000 S1 (300pM with 1% PhiX loading concentration, 28+94 bp read configuration).

Table 6.1 Antibody-derived tag and fluorophore panel for CITE-seq in Chapter 3.

Antibody	Clone	Catalogue No.	Barcode/Fluorophore	Company
CD26	BA5b	302720	GGTGGCTAGATAATG	BioLegend
CD45	HI30	304064	TGCAATTACCCGGAT	BioLegend
TIM3	F38-2E2	345047	TGTCCTACCCAACTT	BioLegend
CD99	3B2/TA8	371317	ACCCGTCCCTAAGAA	BioLegend
CD33	P67.6	366629	TAACTCAGGGCCTAT	BioLegend
CD38	HIT2	303541	TGTACCCGCTTGTGA	BioLegend
CD44	IM7	103045	TGGCTTCAGGTCCTA	BioLegend
CD117	104D2	313241	AGACTAATAGCTGAC	BioLegend
CD34	581	343537	GCAGAAATCTCCCTT	BioLegend
CD90	5E10	328135	GCATTGTACGATTCA	BioLegend
CD49F	GoH3	313633	TTCCGAGGATGATCT	BioLegend
CD10	HI10a	312231	CAGCCATTCATTAGG	BioLegend
CD135	BV10AH2	313317	CAGTAGATGGAGCAT	BioLegend
CD123	6H6	306037	CTTCACTCTGTCAGG	BioLegend
CD371 (CLEC12A)	50C1	353613	CATTAGAGTCTGCCA	BioLegend
CD7	CD7-6B7	343123	TGGATTCCCGGACTT	BioLegend
HLA-DR	L243	307659	AATAGCGAGCAAGTA	BioLegend

GPR56	CG4	358207	GCCTAGTTTCCGTTT	BioLegend
CD45RA	HI100	304157	TCAATCCTTCCGCTT	BioLegend
CD64	10.1	305037	AAGTATGCCCTACGA	BioLegend
CD11b	ICRF44	301353	GACAAGTGATCTGCA	BioLegend
CD3	UCHT1	300475	CTCATTGTAACTCCT	BioLegend
CD4	SK3	344649	GAGGTTAGTGATGGA	BioLegend
CD8	SK1	344751	GCGCAACTTGATGAT	BioLegend
CD25	BC96	302643	TTTGTCCTGTACGCC	BioLegend
CD19	HIB19	302259	CTGGGCAATTACTCG	BioLegend
CD56	5.1H11	362557	TCCTTTCCTGATAGG	BioLegend
CD16	3G8	302061	AAGTTCACTCTTTGC	BioLegend
CD274 (PD-L1)	29E.2A3	329743	GTTGTCCGACAATAC	BioLegend
CD223 (LAG-3)	11C3C65	369333	CATTTGTCTGCCGGT	BioLegend
CD152 (CTLA-4)	BNI3	369619	ATGGTTCACGTAATC	BioLegend
CD279 (PD-1)	EH12.2H7	329955	ACAGCGCCGTATTTA	BioLegend
CD86	IT2.2	305443	GTCTTTGTCAGTGCA	BioLegend
CD226 (DNAM-1)	11A8	338335	TCTCAGTGTTTGTGG	BioLegend
CD314 (NKG2D)	1D11	320835	CGTGTTTGTTCCTCA	BioLegend
CD119 (IFNGR1)	GIR-208	308607	TGTGTATTCCCTTGT	BioLegend
CD155 (PVR)	SKII.4	337623	ATCACATCGTTGCCA	BioLegend
Streptavidin	-	405251	AACCTTTGCCACTGC/PE	BioLegend
pan-NK2GDL	-	1299-NK-050	Biotin	R&D Systems
CD45	HI30	560566	AF700	BD Biosciences
CD3	OKT3	317332	BV510	BioLegend
7-AAD	-	559925	-	BD Biosciences

6.1.3 Strand-seq data preprocessing

Paired-end sequencing reads were aligned to the human reference genome (GRCh38) using BWA[224] and duplicated reads were marked using biobambam[225] as described before for the Strand-seq data analysis[93]. Good quality (MAPQ>=10) and non-duplicated reads were used in the downstream analysis. Reads aligned to Watson strand and Crick strand were counted separately in the 100kb genomic bins. Based on the read depth, strand orientation, and haplotype information, SV calling was performed using the scTRIP method[93].

6.1.4 Cell type classification of CK-AML cells using scNOVA

Using single-cell Strand-seq libraries of CK-AML, scNOVA[98] analysis was performed to obtain nucleosome occupancy at gene bodies for each single-cell. As genetic copy-number alteration can confound the nucleosome occupancy measurement at gene bodies, copy-number normalization of nucleosome occupancy based on the ploidy status inferred by PloidyassignR

using 1Mb bins and 500kb sliding window (Christiansen et al. in preparation) was performed. The copy-number normalized nucleosome occupancy matrix was used as input for the nucleosome occupancy-based cell-type classifier of HSPCs (Grimes et al. in preparation) to predict the most likely cell type for each single-cell Strand-seq library.

6.1.5 Identification of clone-specific differentially occupied genes and their over-represented pathways using scNOVA

Using the copy-number normalized nucleosome occupancy measurement at gene bodies, as described above, differential gene activity analysis of scNOVA[98] was performed for CK282 and CK349 subclones. To infer differentially active genes for each subclone, the single cells in certain subclones with all other single cells in the same tumor were compared using an alternative mode of scNOVA based on PLS-DA. The inferred cell-type was considered as a confounding factor in the differential analysis. After obtaining clone-specific differentially occupied genes, functional enrichment analysis of the identified gene lists was performed using the gene-set over-representation analysis provided by consensusPathDB[226].

6.1.6 Haplotype-specific nucleosome occupancy analysis of CK397 to explore local effect of balanced SVs

First, the chromosome-wide haplotype of nucleosome occupancy at gene bodies was resolved. The nucleosome occupancy of two haplotypes for each gene were compared using Wilcoxon ranksum test followed by Benjamini-Hochberg multiple correction. Using 10% FDR cutoff, genes showing haplotype-specific nucleosome occupancy were identified.

6.1.7 CITE-seq data pre-processing and integration

CITE-seq data pre-processing was done using Cell Ranger (v.6.0.0) (10X Genomics), including aligning the sequencing reads to the GRCh38 human reference genome build, distinguishing cells from the background, and generating unified feature-barcode matrices. For each cell barcode, the generated feature-barcode matrices contained gene expression counts alongside cell surface protein feature counts.

6.1.7.1 Quality control of CITE-seq data

The R package Seurat (v.4.0.4) was used to calculate the quality control metrics[227]. Cells were removed from the analysis if fewer than 200 or more than 8,000 distinct genes, fewer than 1,000 counts or more than 15% of reads mapping to mitochondrial genes were detected.

6.1.7.2 Pre-processing and dimensional reduction of CITE-seq data

Pre-processing and dimensional reduction of CITE-seq data was performed independently on both RNA and ADT assays. Gene counts were normalized by applying regularized negative binomial regression using the Seurat sctransform function[228], followed by principal component analysis with highly variable genes as input. Cell surface protein counts were center log ratio-transformed across cells using the Seurat NormalizeData function with 'CLR' method, followed by scaling and principal component analysis.

6.1.7.3 Weighted Nearest Neighbor analysis of CITE-seq data

For each cell, its closest neighbors in the dataset were calculated based on a weighted combination of RNA and protein similarities, using the Seurat FindMultiModalNeighbors function[107]. For the RNA modality 30 dimensions were used and for the protein modality 18 dimensions. Downstream analysis including UMAP visualization and tSNE visualization of the data as well as clustering was performed based on a weighted combination of RNA and protein data. Clustering of the cells was done using the FindClusters function.

6.1.7.4 Integration of CITE-seq datasets from different patients

For joint analysis of CITE-seq data from all patients, all cells were projected into a shared embedding using Harmony R package[136]. The pre-processed Seurat object with computed PCA was fed into RunHarmony function with patient ID set as covariate for correction. Downstream analyses, including UMAP visualization of the data as well as clustering was performed based on Harmony embeddings.

6.1.8 scNOVA-CITE workflow: coupling SVs with gene and cell surface protein expression

6.1.8.1 Copy Number Variation analysis from CITE-seq gene expression data

Copy number variation (CNV) calling from the gene expression counts from CITE-seq data was done using the CONICSmat R package. Briefly, to infer copy number status of each cell, CONICSmat fits a two-component Gaussian Mixture Model for each provided chromosomal region. The mixture model is fit to the average gene expression of genes within a region, and cells with a deletion in that region show on average lower gene expression compared to cells without the deletion. The copy number status for each cell can be deciphered from the posterior probabilities for each cell belonging to one of the components[148].

Here, the SV discovery from scNOVA was used to construe a list of chromosomal regions containing SVs. These together with chromosomal coordinates of all chromosome arms on autosomes were used to infer the copy number status of each cell for each chromosomal region using the log2(CPM/10+1) normalized gene counts from CITE-seq. To be able to detect CNVs affecting smaller regions, posterior probabilities were computed for regions with more than 10 expressed genes (modified VisualizePosterior.R script; line 107 if(length(chr_genes)>10)). After obtaining the mixture model results, the results were restricted to regions of known copy number events from the scNOVA SV calling. Posterior probability cutoff of 0.8 was used for a confident CNV assignment.

6.1.8.2 Assignment of single cells from CITE-seq data to genetic subclones

CNVs from CONICSmat reaching confidence cutoff were used to identify 'marker SVs' matching subclone-specific SVs identified using scNOVA. These 'marker SVs' were used to assign each cell to its corresponding genetic subclone.

6.1.8.3 "Reference-based" mapping of leukemic cells onto a multimodal bone marrow atlas

Single leukemic cells were assigned to their corresponding healthy counterparts by determining similarity to the reference bone marrow cell types. The previously published CITE-seq dataset, which consists of 30,672 scRNA-seq profiles measured alongside a panel of 25 antibodies from bone marrow, was used as the reference bone marrow atlas[229]. "Supervised principal component analysis" (sPCA)[230] was applied to the transcriptome measurements in the reference dataset using the RunSPCA function. After computing the transformation, it was projected onto the CK-AML CITE-seq dataset. Anchors between each CK-AML dataset and the multimodal bone marrow reference dataset were computed using the FindTransferAnchors function, followed by individual mapping of each of the datasets using the MapQuery function[229]. The resulting supervised annotations were leveraged to test for differences in cell type abundance across different genetic subclones using Fisher's exact test.

6.1.8.4 Finding differentially expressed features between subclones

Marker genes that defined each genetic subclone by differential expression were identified using the scran findMarkers function. To account for the biases driven by different cell types in the subclones, cell type variable together with the subclone variable were used as predictors in the linear model via the design argument of findMarkers. Only upregulated marker genes

were considered. Genes with an FDR-corrected p-value ≤ 0.05 and at least a 0.25-fold change in expression (log2FC ≥ 0.25) were considered as differentially expressed.

6.1.8.5 Molecular phenotype analysis in gene-sets

Gene-set over-representation analysis using enricher function from clusterProfiler was performed to model gene expression changes across Hallmark and KEGG modules from MSigDB[146,147]. For each gene set, the significance of overlap between the target gene set and genes exhibiting differential gene expression between subclones was computed using hypergeometric tests, followed by controlling the FDR at 10%. AUCell[163] was used for signature score calculations between subclones with default parameters, using Hallmark modules from MSigDB[146,147]. Stemness scores were calculated for each cell as the mean expression of the normalized gene counts of the signature genes from Ng, Mitchell [45]

6.1.9 Determination of *in vivo* leukemia-initiating potential

NSG mice were bred and housed under specific pathogen-free conditions at the central animal facility of the German Cancer Research Center (DKFZ). Animal experiments were conducted in compliance with all relevant ethical regulations. All experiments were approved and performed in accordance with all regulatory guidelines of the official committee at the Regierungspräsidium Karlsruhe (G42/18).

6.1.9.1 Xenotransplantations and analysis of leukemic engraftment

Female mice 8-12 weeks of age were sublethally irradiated (175 cGy) 24 h before xenotransplantation assays. AML samples were stained with human CD3 MicroBeads (Miltenyi Biotec, cat # 130-050-101) for depletion of CD3+ T cells. Magnetic-activated cell sorting (MACS) was performed according to manufacturer's instructions, and unlabeled cells run through the MACS column were collected. 1 x 10^6 – 2 x 10^6 bulk CD3-depleted AML cells were injected into the femoral bone marrow cavity of sublethally irradiated mice. Human leukemic engraftment in mouse bone marrow was evaluated by flow cytometry at 10 weeks, at 16 weeks and at end point (maximum 30 weeks unless end point criteria were reached earlier) using anti-human-CD45-FITC (clone HI30), anti-human-CD34-BUV395 (clone 581), anti-human-CD38-BUV496 (clone HIT2), anti-human-GPR56-PE (clone CG4), anti-human-CD19-APC (clone HIB19), anti-human-CD33-APC (clone WM53), and anti-mouse-CD45-FITC (clone 30-F11). Mice were considered 'engrafted' if human cells represented >1% of the bone marrow cell population and 'leukemic/myeloid' if the human cells showed >80% CD33

positivity. At end point, bone marrow cells were harvested from tibiae, femurs, iliac crests, and spine by bone crushing. Spleen cells were harvested by mincing the spleen with a plunger. Following red blood cell lysis, cells were resuspended in Cryostore (Sigma, cat # C2874-100) and stored in liquid nitrogen until further use.

6.1.10 Fluorescent activated cell sorting (FACS) of 'mini-bulk'

Primary AML cells as well as AML cells from xenotransplantations showing myeloid engraftment were stained with a total of 14 and 12 fluorescent antibodies, respectively (Table 6.2). Cells were sorted into five populations according to CD34, GPR56 and pan-NKG2DL expression within the lineage-negative live cell gate. Cells from each population and from lineage-negative bulk were sorted directly into RNA extraction buffer (Thermo Fisher, cat # KIT0214), snap-frozen and stored at -80 °C until RNA extraction. Cells from each population and from lineage-negative bulk were sorted into PBS/BSA, spun-down, and the pellets stored at -80 °C until DNA extraction.

Table 6.2 Antibody panels for FACS in Chapter 3.

Primary AMLs				
Marker	**Clone**	**Manufacturer**	**Fluorophore**	**Catalogue No.**
CD45	HI30	BioLegend	FITC	304006
CD34	581	BD Biosciences	BUV395	563778
CD38	HIT2	BD Biosciences	BUV496	564657
CD123	7G3	BD Biosciences	PECy7	560826
CD117	104D2	BioLegend	BV785	313238
GPR56	CG4	BioLegend	PE	358204
pan-NKG2DL	-	R&D Systems	Biotin	1299-NK-050
Streptavidin	-	BD Biosciences	BUV805	564923
Cell Rox	-	ThermoFisher	Deep Red	C10422
CD3	UCHT1	BD Biosciences	BV480	566105
CD4	SK3	BD Biosciences	BV480	566104
CD8	RPA-T8	BD Biosciences	BV480	566121
CD20	2H7	BD Biosciences	BV480	566132
CD235a	GAR2	BD Biosciences	BV480	746358
Live/Dead	-	BD Biosciences	7AAD	559925
Xenotransplantations				
Marker	**Clone**	**Manufacturer**	**Fluorophore**	**Catalogue No.**
CD45	HI30	BD Biosciences	AF700	560566
CD34	581	BD Biosciences	BUV395	563778
CD38	HIT2	BD Biosciences	BUV496	564657
CD123	7G3	BD Biosciences	PE-CF594	562391

CD117	104D2	BioLegend	BV785	313238
GPR56	CG4	BioLegend	PE	358204
pan-NKG2DL	-	R&D Systems	Biotin	1299-NK-050
Streptavidin	-	BD Biosciences	BV421	563259
CD19	HIB19	ebioscience	APC	17-0199-42
CD33	WM53	LIFE Technologies	PE-Cy7	25-0338-42
mCD45	30-F11	eBioscience	FITC	11-0451-82
CD3	UCHT1	BD Biosciences	BV510	317332
Live/Dead	-	BD Biosciences	7AAD	559925

6.1.11 RNA sequencing of 'mini-bulk' AML populations

RNA extraction and purification of FACS-sorted cells was done using PicoPure RNA Isolation Kit according to manufacturer's instructions (Thermo Fisher, cat # KIT0214). Purified RNA was eluted in nuclease free water. RNA quality assessment and quantification were performed with Bioanalyzer using Agilent RNA 6000 Pico Kit (Agilent, cat # 5067-1513). Full-length cDNA synthesis was done using SMART-Seq v4 Ultra Low Seq kit (Takara Bio Clontech, cat # 634888) following manufacturer's instructions with 0.5ng to 1ng of total RNA as input and 11 cycles used for cDNA amplification PCR. Tagmentation of cDNA was done using Nextera XT DNA Library Preparation Kit (Illumina, cat # FC-121-1030). All RNA libraries were pooled and sequenced together on an Illumina NextSeq 550 high output sequencer (1.4 pM with 1% PhiX loading concentration, paired-end 75bp read configuration).

6.1.11.1 Fusion transcript detection

STAR-aligner-based Arriba fusion detection tool[231] was used to detect fusion transcripts from 'mini-bulk' AML populations. First, STAR aligner 2.5.3a was used to demultiplex the reads and to align the FASTQ files containing reads for individual samples by two-pass alignment[232]. Reads were aligned to a STAR index generated using the GRCh38 genome build. Detection of chimeric reads was enabled. Next, the Arriba fusion detection tool was used to extract the Chimeric.out.sam and Aligned.out.bam files and to create a list of fusion predictions passing Arriba's filters.

6.1.12 Infinium MethylationEPIC DNA methylation arrays of 'mini-bulk' AML populations

Genomic DNA was extracted from FACS-sorted human CD45-positive AML cells from xenotransplantation samples using the QIAamp DNA Micro Kit (Qiagen, cat # 56304) and QIAmp DNA Mini Kit (Qiagen, cat # 51304) according to manufacturer's instructions. DNA

methylation data were generated using Infinium MethylationEPIC BeadChip arrays at the Genomics and Proteomics Core Facility of the DKFZ.

6.1.12.1 Infinium MethylationEPIC array processing

To calculate methylation ratios and the log2 methylation ratios, raw intensity files were obtained using the minfi package[233]. The data were normalized using Illumina preprocessing method implemented in minfi. Several quality control measures were applied to check for arrays with low quality. Median methylated and unmethylated signals were calculated for each array; arrays with detection P value greater than 0.05 were excluded. Probes that failed in one or more arrays as well as probes containing an annotated single-nucleotide polymorphism (SNP) at the single-base extension or CpG sites were removed (116,734 probes removed). Minfi version 1.30.0 was used. CNV calling was performed using conumee (v.1.30.0), a tool for somatic CNV calling from DNA methylation arrays.

6.1.13 Optical Genome Mapping

Optical genome mapping was performed on xenotransplantation samples from CK282 (three mice) and CK397 (one mouse). Ultra-high molecular weight DNA was extracted from AML cells recovered from bone marrow or spleen following manufacturer's protocols (Bionano Genomics, USA). Briefly, the cells were digested followed by DNA precipitation and binding with nanobind magnetic disk. Labeling of the ultra-high molecular weight DNA was performed following manufacturer's instructions (Bionano Genomics, USA), with 750 ng of DNA labeled using the Standard Direct Labeling Enzyme 1 (DLE-1). The fluorescently labeled DNA molecules were imaged sequentially across nanochannels on a Saphyr instrument. For all tested samples a coverage of approximately 300X was achieved.

Analysis of the somatic SVs was done using the Rare Variant Analyses software provided by Bionano Genomics. First, molecules were aligned against the GRCh38 human reference genome build, with no assumption about ploidy. Next, the clustered sets of molecules that identified the same variant were used to assemble consensus genome maps (*.cmaps). Finally, the cmaps were realigned to GRCh38a and the consensus forming final SV calls was used to confirm the SV data. Fractional copy number analysis was performed from the alignment of molecules and labels against GRCh38 (alignmolvrefsv). A sample's raw label coverage was normalized against relative coverage from normal human controls, segmented, and baseline copy number state estimated from calculating the mode of coverage of all labels. Last,

significant increase/decrease from the baseline was used to assess the copy number states of segmented genomic intervals. SV and copy number calls occurring in GRC38 regions found to be high variance were masked.

6.1.14 QIAseq Targeted DNA Panel sequencing

Targeted DNA panel libraries were prepared from primary bulk CK-AML samples and engrafted human CD45-sorted xenotransplantation samples. DNA libraries were built using the Human Myeloid Neoplasms Targeted DNA Panel kit (Qiagen, cat # DHS-003Z-12) following manufacturer's protocol. Briefly, genomic DNA samples were first fragmented, end-repaired, and A-tailed within a single, controlled multi-enzyme reaction. The prepared DNA fragments were then ligated at their 5' ends with a sequencing platform-specific adapter containing unique molecular identifiers (UMIs) and sample index, followed by target enrichment using one region-specific primer and one universal primer complementary to the adapter. All targeted DNA panel libraries were pooled and sequenced together on an Illumina NextSeq 550 high-output sequencer (1.4 pM with 1% PhiX loading concentration, paired-end 150bp read configuration).

6.1.14.1 Data analysis of Targeted DNA Panel sequencing

Post-sequencing analysis of the FASTQ files was performed using the Genomics Workbench (v.12.0.3) (Qiagen) with the Biomedical Genomics Analysis Plugin (v.1.2.1), corresponding target and hotspot BED-files for the hg19 reference genome. Briefly, UMIs were trimmed, and reads were annotated with the corresponding UMI. UMI annotated reads were mapped to hg19, and a single consensus read, a UMI read, was created from aligned reads that had the same UMI. Subsequently, ligation artifacts were removed.

The following criteria were subsequently applied to account for sequencing artifacts: (1) The pipeline-specific quality parameter "Filter" was set to "PASS" and variants in tandem repeats were rejected. (2) A minimum of 80 UMI reads per locus and a minimum of 5 UMI reads carrying the variant were used as a threshold. (3) Synonymous variants and common SNPs with a minor allele frequency (MAF) of >=0.01 based on the dbSNP common_all_20160601 database were excluded. (4) Given the size of the cohort, variants (same gene and position) found to occur more than 4 times were considered to be panel artifacts and were excluded from the final analysis. The same applied to genes mutated more than 2 times in the same sample. Filtering was run using R (v.4.0.0).

6.1.15 *Ex vivo* drug screening

Ex vivo drug screening was performed on thawed cells from diagnosis samples of CK282 and CK349, and human CD45+ cells from the PDX of CK282. Cells were cultured using previously established protocol[206,223] using IMDM, 15% BIT (bovine serum albumin, insulin, transferrin; Stem Cell Technologies, cat # 09500), 100 ng/ml SCF (PeproTech, cat # 300-07), 50 ng/ml FLT3-L (PeproTech, cat # 300-19), 20 ng/ml IL-3 (PeproTech, cat # 200-03), 20 ng/ml G-CSF (PeproTech, cat # 300-23), 100 µM β-mercaptoethanol (ThermoFisher, cat # 31350010), 500 nM SR1 (StemRegenin 1, STEMCELL Technologies, cat # 72342), 500 nM UM729 (STEMCELL Technologies, cat # 72332), and 1% penicillin-streptomycin (Sigma, cat # P4458-100ML). $0.5x10^5$ AML cells/well were seeded in flat-bottom 96-well plates, and cells were treated with different concentrations of BH3 mimetics alone or in combination with 1 µM of azacytidine for 24h. Cells were also treated with different concentrations of cytarabine and daunorubicin alone or in combination for 24h and 72h (Table 6.3). After 24h, the cells were stained with cell surface antibodies (Table 6.4). Same amount of CountBright Absolute Counting Beads (Thermo Fisher Scientific, cat # C36950) together with 7-AAD (BD Biosciences, cat # 559925) were added to each sample prior to analysis with BD LSRFortessa Cell Analyzer. After 72h, viability was assessed using CellTiter-Glo (Promega, cat # G7571) Luminescent Cell Viability Assay or by flow cytometry for a set of concentrations (Table 6.3).

Table 6.3 Drug concentrations of *ex vivo* drug screening in Chapter 3.

24h				
Drug1	**Concentration1**	**Drug2**	**Concentration2**	**Assay**
Untreated	0uM			FC
Azacitidine	1uM			FC
Azacitidine	1uM	A-1331852	100nM	FC
Azacitidine	1uM	A-1331852	5nM	FC
Azacitidine	1uM	A-1331852	35nM	FC
A-1331852	100nM			FC
A-1331852	5nM			FC
A-1331852	35nM			FC
Cytarabine	50uM			FC
Cytarabine	5uM			FC
Cytarabine	25uM			FC
Cytarabine	2uM	Daunorubicin	10nM	FC
Cytarabine	0,5uM	Daunorubicin	2nM	FC
Cytarabine	0,5uM	Daunorubicin	10nM	FC

Daunorubicin	100nM			FC
Daunorubicin	2nM			FC
Daunorubicin	20nM			FC
Azacitidine	1uM	MIK665	100nM	FC
Azacitidine	1uM	MIK665	5nM	FC
Azacitidine	1uM	MIK665	35nM	FC
MIK665	100nM			FC
MIK665	5nM			FC
MIK665	35nM			FC
Azacitidine	1uM	Venetoclax	100nM	FC
Azacitidine	1uM	Venetoclax	5nM	FC
Azacitidine	1uM	Venetoclax	35nM	FC
Venetoclax	100nM			FC
Venetoclax	5nM			FC
Venetoclax	35nM			FC
72h				
Drug1	**Concentration1**	**Drug2**	**Concentration2**	**Assay**
Untreated	0uM			FC and CTG
Cytarabine	10uM	Daunorubicin	500nM	FC
Daunorubicin	500nM			FC
Cytarabine	2uM	Daunorubicin	167nM	FC
Daunorubicin	167nM			FC
Cytarabine	10uM	Daunorubicin	0,69nM	FC
Cytarabine	2uM	Daunorubicin	0,23nM	FC
Cytarabine	10uM			FC
Cytarabine	2uM			FC
Cytarabine	0,4uM			FC
Cytarabine	250uM	Daunorubicin	500nM	CTG
Cytarabine	50uM	Daunorubicin	500nM	CTG
Cytarabine	2uM	Daunorubicin	500nM	CTG
Cytarabine	0,4uM	Daunorubicin	500nM	CTG
Cytarabine	250uM	Daunorubicin	166,7nM	CTG
Cytarabine	50uM	Daunorubicin	166,7nM	CTG
Cytarabine	10uM	Daunorubicin	166,7nM	CTG
Cytarabine	0,4uM	Daunorubicin	166,7nM	CTG
Cytarabine	250uM	Daunorubicin	55,6nM	CTG
Cytarabine	50uM	Daunorubicin	55,6nM	CTG
Cytarabine	10uM	Daunorubicin	55,6nM	CTG
Cytarabine	2uM	Daunorubicin	55,6nM	CTG
Cytarabine	0,4uM	Daunorubicin	55,6nM	CTG
Daunorubicin	55,6nM			CTG
Cytarabine	250uM	Daunorubicin	18,5nM	CTG
Cytarabine	50uM	Daunorubicin	18,5nM	CTG

Cytarabine	10uM	Daunorubicin	18,5nM	CTG
Cytarabine	2uM	Daunorubicin	18,5nM	CTG
Cytarabine	0,4uM	Daunorubicin	18,5nM	CTG
Daunorubicin	18,5nM			CTG
Cytarabine	250uM	Daunorubicin	6,2nM	CTG
Cytarabine	50uM	Daunorubicin	6,2nM	CTG
Cytarabine	10uM	Daunorubicin	6,2nM	CTG
Cytarabine	2uM	Daunorubicin	6,2nM	CTG
Cytarabine	0,4uM	Daunorubicin	6,2nM	CTG
Daunorubicin	6,2nM			CTG
Cytarabine	250uM	Daunorubicin	2nM	CTG
Cytarabine	50uM	Daunorubicin	2nM	CTG
Cytarabine	10uM	Daunorubicin	2nM	CTG
Cytarabine	2uM	Daunorubicin	2nM	CTG
Cytarabine	0,4uM	Daunorubicin	2nM	CTG
Daunorubicin	2nM			CTG
Cytarabine	250uM	Daunorubicin	0,7nM	CTG
Cytarabine	50uM	Daunorubicin	0,7nM	CTG
Cytarabine	2uM	Daunorubicin	0,7nM	CTG
Cytarabine	0,4uM	Daunorubicin	0,7nM	CTG
Daunorubicin	0,7nM			CTG
Cytarabine	250uM	Daunorubicin	0,2nM	CTG
Cytarabine	50uM	Daunorubicin	0,2nM	CTG
Cytarabine	10uM	Daunorubicin	0,2nM	CTG
Cytarabine	0,4uM	Daunorubicin	0,2nM	CTG
Daunorubicin	0,2nM			CTG
Cytarabine	250uM			CTG
Cytarabine	50uM			CTG

Abbreviations: FC, Flow cytometry; CTG, Cell-Titer-Glo

Table 6.4 Antibody panel for *ex vivo* drug screening in Chapter 3.

Marker	Clone	Manufacturer	Fluorophore	Catalogue No.	Sample stained
mCD45	30-F11	eBioscience	FITC	11-0451-82	PDX282
CD45RA	HI100	BD Biosciences	BV421	562885	CK349, CK282, PDX282
CD3	UCHT1	BD Biosciences	BV510	317332	CK349, CK282, PDX282
CD20	2H7	BioLegend	BV510	302340	CK349, CK282, PDX282
CD235a	GAR2	BD Biosciences	BV510	740174	CK349, CK282, PDX282
CD45	HI30	BD Biosciences	AF700	560566	CK349, CK282, PDX282
CD49F	GoH3	BioLegend	PE	313611	CK349
GPR56	CG4	BioLegend	PE-Cy7	358206	CK349, CK282, PDX282
CD34	581	BD Biosciences	BUV395	563778	CK349, CK282, PDX282
CD38	HIT2	BD Biosciences	BUV496	564657	CK349, CK282, PDX282

CD99	HCD99	BioLegend	FITC	318006	CK282
CD90	5E10	BioLegend	PE	328109	CK282

Abbreviations: m, Mouse; PDX, Patient-derived xenograft

6.1.16 Intracellular staining for BCL-2 family members

Intracellular staining was performed on thawed cells from diagnosis samples of CK282 and CK349, and human CD45+ cells from the PDX of CK282. Thawed cells were stained with Zombie NIR Fixable Viability stain in PBS (BioLegend, cat # 423105), followed by staining with cell surface antibodies (Table 6.5). Stained cells were fixed and permeabilized using the Fixation/Permeabilization Solution Kit (BD Biosciences, cat # 554714) according to manufacturer's instructions, followed by a secondary permeabilization step for enhanced intracellular staining using Permeabilization Buffer Plus (BD Biosciences, cat # 561651). Fixed and permeabilized cells were stained for anti-human-BCL-2-AF647 (clone 124, Cell Signaling, cat # 82655), anti-human-MCL-1-AF488 (clone D2W9E, Cell Signaling, cat # 58326) and anti-human-BCL-xL-PE-Cy7 (clone 54H6, Cell Signaling, cat # 81965) (Table 6.5). Samples were analyzed with BD LSRFortessa Cell Analyzer.

Table 6.5 Antibody panel for intracellular staining of BCL-2 family members in Chapter 3.

Marker	Clone	Manufacturer	Fluorophore	Catalogue No.	Sample stained
CD45RA	HI100	BD Biosciences	BV421	562885	CK349, CK282, PDX282
CD3	UCHT1	BD Biosciences	BV510	317332	CK349, CK282, PDX282
CD20	2H7	BioLegend	BV510	302340	CK349, CK282, PDX282
CD235a	GAR2	BD Biosciences	BV510	740174	CK349, CK282, PDX282
CD45	HI30	BD Biosciences	AF700	560566	CK349, CK282, PDX282
CD49F	GoH3	BioLegend	PE	313611	CK349
CD90	5E10	BioLegend	PE	328109	CK282
CD34	581	BD Biosciences	BUV395	563778	CK349, CK282, PDX282
CD38	HIT2	BD Biosciences	BUV496	564657	CK349, CK282, PDX282
Zombie NIR	-	BioLegend	-	423105	CK349, CK282, PDX282
MCL-1	D2W9E	Cell Signaling	AF488	58326	CK349, CK282, PDX282
BCL-2	124	Cell Signaling	AF647	82655	CK349, CK282, PDX282
BCL-xL	54H6	Cell Signaling	PE-Cy7	81965	CK349, CK282, PDX282

Abbreviations: PDX, Patient-derived xenograft

6.1.17 Quantification and statistical analysis

Methods used for statistical analyses were detailed in figure legends. In all figures, NS indicates a not significant P value of >0.05; *, **, ***, and **** indicate $P < 0.05$, $P < 0.01$, $P < 0.001$, and $P < 0.0001$, respectively. All statistical analyses were done using R 4.0.0. Flow cytometry data analysis was done using FlowJo (v.10.5.3).

6.1.18 Code availability

The computational code of the analytical framework for scNOVA is available open source at https://github.com/jeongdo801/scNOVA.

Other software used in the study:

- Mosaicatcher (https://github.com/friendsofstrandseq/mosaicatcher-pipeline),
- Strand-PhaseR (https://github.com/daewoooo/StrandPhaseR),
- CONICSmat (https://github.com/diazlab/CONICS),
- Delly2 (https://github.com/dellytools/delly),

BWA (v.0.7.15), STAR (v.2.7.9a and v.2.5.3a), SAMtools (v.1.3.1), biobambam2 (v.2.0.76), Sambamba (v.0.6.5), R (v.4.0.0), DESeq2, FlowJo (v.10.5.3) and BD FACSDiva.

6.2 Experimental Practices Related to Chapter 4

The text of the following chapter has been taken from Waclawiczek, Leppä [179] and was originally written by myself:

6.2.1 Primary AML patient samples and clinical data

AML samples were collected from diagnostic peripheral blood aspirations at the University hospital in Heidelberg or Hannover in accordance with the Declaration of Helsinki and based on institutional approvals after obtaining written informed consent from each patient. The project was approved by the Ethics Committee of the Medical Faculty of Heidelberg (S-169/2017, S-648/2021) and the local Ethics Review Committee of Hannover Medical School (ethical vote No.7972_BO_K_2018). Peripheral blood mononuclear cells were isolated by density gradient centrifugation using Ficoll Paque Plus (GE Healthcare, cat # GE17-1440-03), and stored in liquid nitrogen until further use. Detailed patient and specimen characteristics are listed in Waclawiczek, Leppä [179] and are based on the clinical diagnostic reports.

6.2.2 Processing of primary AML cells

Viably cryopreserved AML peripheral blood samples were thawed at 37 °C in IMDM containing 10% FBS, and treated with DNase I for 15min (100 µg/ml).

6.2.3 *Ex vivo* drug screening in primary leukemia cells

Recovered cells were cultured using previously established protocols[206,223] using IMDM, 15% BIT (bovine serum albumin, insulin, transferrin; Stem Cell Technologies, cat # 09500), 100 ng/ml SCF (PeproTech, cat # 300-07), 50 ng/ml FLT3-L (PeproTech, cat # 300-19), 20 ng/ml IL-3 (PeproTech, cat # 200-03), 20 ng/ml G-CSF (PeproTech, cat # 300-23), 100 µM β-mercaptoethanol (ThermoFisher, cat # 31350010), 500 nM SR1 (StemRegenin 1, STEMCELL Technologies, cat # 72342), 1 µM UM729 (STEMCELL Technologies, cat # 72332), and 1% penicillin-streptomycin (Sigma, cat # P4458-100ML). For drug assay in Figure 4.1, $0.5x10^5$ AML cells/well were seeded in flat-bottom 96-well plates, and cells were treated with increasing concentration of azacytidine (0.5 µM, 1.5 µM, 4.5 µM, 13.5 µM, 40.5 µM) and venetoclax (0.3 nM, 0.9 nM, 2.7 nM, 8.1 nM, 24.3 nM, 72.9 nM, 218.7 nM, 656.1 nM, 1968.3 nM) alone and in combination for 72h. After 72h, viability was assessed using CellTiter-Glo (Promega, cat # G7571) Luminescent Cell Viability Assay. For drug assay in Figure 4.7 and Figure 4.8, $0.5x10^5$ AML cells/well were seeded in flat-bottom 96-well plates, and cells were

treated with 1.5 µM of azacytidine and 100 nM of venetoclax for 24h. After 24h, the cells were stained with fluorescent cell surface antibodies (Table 6.6). Same amount of CountBright Absolute Counting Beads (Thermo Fisher Scientific, cat # C36950) together with 7-AAD (BD Biosciences, cat # 559925) were added to each sample prior to analysis with BD LSRFortessa Cell Analyzer.

Table 6.6 Antibody panel for *ex vivo* drug screening and patient sample phenotyping in Chapter 4.

Marker	Clone	Manufacturer	Fluorophore	Catalogue No.
CD64	10.1	BioLegend	APC	305014
CD34	581	BD Biosciences	BUV395	563778
CD33	WH53	BioLegend	BV421	303416
CD11b	ICRF44	BioLegend	BV711	301344
GPR56	CG4	BioLegend	PE	358204
CD56	5.1H11	BioLegend	BV785	362550
CD45	HI30	BioLegend	FITC	304006
CD3	OKT3	BioLegend	BV510	317332
CD20	2H7	BioLegend	BV510	302340
CD235a	GA-R2	BD Biosciences	BV510	740174

6.2.4 Intracellular staining for BCL-2 family members

Thawed cells were stained with Zombie NIR Fixable Viability stain in PBS (BioLegend, cat # 423105), followed by staining with fluorescent cell surface antibodies (Table 6.7). Stained cells were fixed and permeabilized using the Fixation/Permeabilization Solution Kit (BD Biosciences, cat # 554714) according to manufacturer's instructions, followed by a secondary permeabilization step for enhanced intracellular staining using Permeabilization Buffer Plus (BD Biosciences, cat # 561651). Fixed and permeabilized cells were stained separately for anti-human-BCL-2-PE (clone 124, Cell Signaling, cat # 26295S), anti-human-MCL-1-PE (clone D2W9E, Cell Signaling, cat # 65617S) and anti-human-BCL-xL-PE (clone 54H6, Cell Signaling, cat # 13835S), or together for anti-human-BCL-2-AF647 (clone 124, Cell Signaling, cat # 82655), anti-human-MCL-1-AF488 (clone D2W9E, Cell Signaling, cat # 58326) and anti-human-BCL-xL-PE (clone 54H6, Cell Signaling, cat # 13835S) (Table 6.7). Samples were analyzed with BD LSRFortessa Cell Analyzer.

Table 6.7 Antibody panel for intracellular staining of BCL-2 family members in Chapter 4.

Marker	Clone	Manufacturer	Fluorophore	Catalogue No.

CD64	10.1	BioLegend	APC	305014
CD34	581	BD Biosciences	BUV395	563778
CD11b	ICRF44	BioLegend	BV711	301344
GPR56	CG4	BioLegend	PE-Cy7	358206
CD45	HI30	BioLegend	BV421	304032
CD3	OKT3	BioLegend	BV510	317332
CD20	2H7	BioLegend	BV510	302340
CD235a	GA-R2	BD Biosciences	BV510	740174
Zombie NIR	-	BioLegend	-	423105
MCL-1	D2W9E	Cell Signaling	PE / AF488	65617S / 58326
BCL-2	124	Cell Signaling	PE / AF647	26295S / 82655
BCL-xL	54H6	Cell Signaling	PE	13835S

6.2.5 MAC-Score Calculation

To ensure consistent and comparable MFI measurements of samples processed and analyzed on separate days, a reference AML sample was processed and run along with each cohort. Detector voltages were adjusted to keep MFI for each BCL-2, MCL-1 and BCL-xL of the *LSC-like* population in the reference sample constant. Small fluctuations of reference sample MFI were adjusted by normalizing the measurement day's reference sample to match previous reference sample measurements. For each sample, normalized MFI for each BCL-2, MCL-1 and BCL-xL of the *LSC-like* population were divided by the respective median MFI of AML patients classified as responders within the cohort to obtain relative MFI-values (rel.MFI). Afterwards MAC-Score was calculated using the following formula:

$$[\mathrm{MAC-Score} = \frac{\mathrm{rel.MFI\ (BCL2)}}{\mathrm{rel.MFI\ (MCL1)+rel.MFI\ (BCL-xL)}}].$$

6.2.6 Longitudinally collected primary AML samples

From three newly diagnosed AML patients (AML55, AML61 and AML62), peripheral blood was drawn prior to treatment with venetoclax and azacytidine, followed by blood draws every to every other day till up to day six on venetoclax and azacytidine treatment. peripheral blood mononuclear cells were isolated as explained above. Prior to freezing, $0.2x10^6$ cells were stained with fluorescent cell surface antibodies (Table 6.6) and analyzed with BD LSRFortessa Cell Analyzer.

6.2.7 Processing of AML cell lines

Twenty-four authenticated and mycoplasma-screened AML cell lines were received from DSMZ and from Cell Services at Francis Crick Institute (courtesy of Dr. Dominque Bonnet). AML cell lines were cultivated at 37 °C in a humidified incubator with 5% CO_2 following the German Collection of Microorganisms and Cell Cultures (DSMZ) culture recommendations. Cell lines were stratified into monocyte-like or primitive-like based on the MFI of the monocytic cell surface marker CD64 (Table 6.8).

Table 6.8 Characteristics of AML cell lines in Chapter 4.

Cell Line	FAB of original patient	Class	(log CD64 MFI)
KG1a	M1	Prim	2.00
ME-1	M4	Prim	2.02
GDM1	M4	Prim	2.43
MOLM13	M5a	Prim	2.54
SKNO1	M2	Prim	2.64
Kasumi6	M2	Prim	2.79
Kasumi1	M2	Prim	2.83
GF-D8	M1	Prim	2.93
HL60	M2	Prim	2.94
ML1	M4	Prim	2.96
ML2	M4	Prim	2.98
NOMO1	M5	Mono	3.58
OCI-AML3	M4	Mono	3.63
PL21	M3	Mono	3.69
YNH-1	M1	Mono	3.72
U937	M5	Mono	3.72
MONOMAC6	M5	Mono	3.77
THP1	M5	Mono	4.02
SKM-1	M5	Mono	4.10

Abbreviations: FAB, French-American-British Classification; Prim, primitive-like; Mono, monocyte-like; log MFI, Logarithm of Mean Fluorescence Intensity

6.2.7.1 *In vitro* drug screening in leukemia cell lines

$0.1x10^5$ cells/well from each cell line were seeded in flat-bottom 96-well plates, and cells were treated with increasing concentration of venetoclax (1 nM, 3 nM, 9 nM, 27 nM, 81 nM, 243 nM, 729 nM, 2187 nM) in combination with a single dose of azacytidine (1.5 μM) for 72h. After 72h, viability was assessed using CellTiter-Glo (Promega, cat # G7571) Luminescent Cell Viability Assay.

6.2.8 Fluorescent activated cell sorting

Primary AML cells were stained with fluorescent cell surface antibodies (Table 6.9). Cells were sorted into four populations according to CD11b, CD64 and GPR56 expression within the lineage-negative gate. Cells from each population and from lineage-negative bulk were sorted directly into RNA extraction buffer (Thermo Fisher, cat # KIT0214), snap-frozen and stored at -80 °C until RNA extraction. For xenotransplantations, cells from each population were sorted into PBS.

Table 6.9 Antibody panel for FACS in Chapter 4.

Marker	Clone	Manufacturer	Fluorophore	Catalogue No.
CD64	10.1	BioLegend	APC	305014
CD33	WH53	LIFE Technologies	PE-Cy7	25-0338-42
CD11b	ICRF44	BioLegend	BV711	301344
GPR56	CG4	BioLegend	PE	358204
CD56	5.1H11	BioLegend	BV785	362550
CD45	HI30	BioLegend	BV421	304006
CD3	OKT3	BioLegend	BV510	317332
CD20	2H7	BioLegend	BV510	302340
CD235a	GA-R2	BD Biosciences	BV510	740174

6.2.9 Determination of *in vivo* leukemia-initiating potential

NSG mice were bred and housed under specific pathogen-free conditions at the central animal facility of the German Cancer Research Center (DKFZ). Animal experiments were conducted in compliance with all relevant ethical regulations. All experiments were approved and performed in accordance with all regulatory guidelines of the official committee at the Regierungspräsidium Karlsruhe (G-140-21 and G42/18.).

6.2.9.1 Xenotransplantations and analysis of leukemic engraftment

Female mice 8-12 weeks of age were sublethally irradiated (175 cGy) 24 h before xenotransplantation assays. Up to $0.1x10^5$ cells from FACS sorted primary AML samples were injected into the femoral bone marrow cavity of sublethally irradiated mice. Mice were monitored daily and at the end point, when mice reached defined termination criteria, bone marrow cells were harvested from tibiae, femurs, and iliac crests by bone crushing. Spleen cells were harvested by mincing the spleen with a plunger. Following red blood cell lysis, cells were resuspended in FBS + 10% DMSO (Sigma, cat # D2650-100) and stored in liquid nitrogen until further use. Human leukemic engraftment in mouse bone marrow was evaluated by flow

cytometry (maximum 45 weeks unless end point criteria were reached earlier) using anti-human-CD45-FITC (clone HI30), anti-human-CD34-BUV395 (clone 581), anti-human-GPR56-PE (clone CG4), anti-human-CD19-FITC (clone HIB19), anti-human-CD33-PE-Cy7 (clone WM53), CD64-APC, CD11b-BV711and anti-mouse-CD45-Alexa700 (clone 30-F11).

6.2.10 RNA sequencing AML populations

RNA extraction and purification of FACS-sorted cells was done using PicoPure RNA Isolation Kit according to manufacturer's instructions (Thermo Fisher, cat # KIT0214). RNA quality assessment and quantification were performed with Bioanalyzer using Agilent RNA 6000 Pico Kit (Agilent, cat # 5067-1513). Whole transcriptome amplification was performed using a modified smart-seq2 protocol[234], with 5µl of a modified RT buffer containing 1× SMART First Strand Buffer (Takara Bio Clontech, cat # 639538), 1 mM dithiothreitol (Takara Bio Clontech), 1 µM template switching oligo (IDT), 10 U μl^{-1} SMARTScribe (Takara Bio Clontech, cat # 639538) and 1 U µl−1 RNasin Plus RNase Inhibitor (Promega, cat # N2615). Tagmentation of cDNA was done using Nextera XT DNA Library Preparation Kit (Illumina, cat # FC-121-1030). All RNA libraries were pooled and sequenced together on an Illumina NextSeq 550 high output sequencer (1.4 pM with 1% PhiX loading concentration, single-end 75bp read configuration).

6.2.11 Raw data processing and quality control of RNA sequencing data

Reads were demultiplexed, and FASTQ files containing reads for individual samples were aligned by two-pass alignment using STAR aligner v. 2.5.3a[235]. Reads were aligned to a STAR index generated from the 1000 Genomes Project human genomes assembly (hs37d5), using GENCODE v.19 gene models. Default alignment call parameters were used with the following modifications: --outSAMtype BAM Unsorted SortedByCoordinate --limitBAMsortRAM 100000000000 --outBAMsortingThreadN=1 --outSAMstrandField intronMotif --outSAMunmapped Within KeepPairs --outFilterMultimapNmax 1 --outFilterMismatchNmax 5 --outFilterMismatchNoverLmax 0.3 --chimSegmentMin 15 --chimScoreMin 1 --chimScoreJunctionNonGTAG 0 --chimJunctionOverhangMin 15 --chimSegmentReadGapMax 3 --alignSJstitchMismatchNmax 5 -1 5 5 --alignIntronMax 1100000 --alignMatesGapMax 1100000 --alignSJDBoverhangMin 3 --alignIntronMin 20.

Alignment file sorting, duplicate marking and BAM index generation was done using Sambamba v.0.6.5 using eight threads[236]. Quality control analysis was performed using the

sambamba flagstat command and rnaseqc v.1.1.8 with the hs37d5 assembly and GENCODE v.19 gene models. Depth of Coverage analysis for rnaseqc was turned off. FeatureCounts v.1.5.1 was used for gene-specific gene counting over exon features based on GENCODE v.19 gene models[237]. Quality threshold was set to 255, which indicates that STAR found a unique alignment. Strand-unspecific counting was used.

6.2.12 Analysis of RNA sequencing data

DESeq2[238] was used for statistical analysis of the read counts to identify differentially-expressed genes between the *LSC-like* and *Mature* populations in Prim-AMLs and Mono-AMLs. Genes with an FDR-corrected p-value ≤ 0.05 and at least a 2-fold change in expression ($|\log2FC| \geq 1$) were considered differentially expressed. Gene expression estimates for PCA visualization were adjusted by variance stabilization. Gene set enrichment analysis for Hallmark gene sets between *LSC-like* and *Mature* cells was performed based on log fold change order-ranked gene list. LSC17 scores were calculated for each AML sample as the mean expression of the normalized log2-transformed gene counts of the 17 LSC signature genes from Ng, Mitchell [45].

6.2.13 Quantification and statistical analysis

Methods used for statistical analyses were detailed in figure legends. Flow cytometry data analysis was done using FlowJo v.10.5.3. Statistical analyses were done using R.4.0.0 and Prism v.9.

Contributions

1.1 Dissection of intra-patient heterogeneity in CK-AML patients with ongoing karyotypic instability identifies targetable relapse-driving leukemic stem cell clones

Co-authors:
Aino-Maija Leppä[*], Karen Grimes[*], Hyobin Jeong[*], Tobias Boch, Alexander Waclawiczek, Darja Karpova, Florian Grünschläger, Anna Jauch, Anna Dolnik, Vera Thiel, Bernardo Rodriguez-Martin, Lars Bullinger, Alwin Krämer, Ashley D. Sanders[#], Jan O. Korbel[#], Andreas Trumpp[#]

[*,#]Contributed equally

I together with Dr. Karen Grimes, Dr. Hyobin Jeong, Dr. Ashley D. Sanders, Dr. Jan O. Korbel and Prof. Andreas Trumpp conceptualized this project. I together with Dr. Karen Grimes performed the Strand-seq experiments. I together with Dr. Karen Grimes and Dr. Hyobin Jeong carried out the analysis of Strand-seq data. Alignment of the CITE-seq data was done with the help of Florian Grünschläger and I performed the downstream analysis of the data. I carried out the xenotransplantation experiments. I performed the *ex vivo* drug screenings with the help of Dr. Alexander Waclawiczek. I carried out the M-FISH experiments together with Prof. Anna Jauch. I performed all other sample processing and data analysis. Dr. Tobias Boch, Dr. Darja Karpova and Vera Thiel provided technical support for the experiments. Prof. Alwin Krämer provided samples and clinical information.

I would like to thank all technicians and in particular Markus Sohn of the Trumpp laboratory for technical assistance; Steffen Schmitt, Marcus Eich, Klaus Hexel, Tobias Rubner, and Florian Blum from the DKFZ Flow Cytometry Core Facility for their assistance; and K. Reifenberg, P. Prückl, M. Durst, A. Rathgeb, and animal caretaker of the DKFZ Central Animal Laboratory for excellent animal welfare and husbandry. I also thank the Single Cell Open Lab for their help as well as DKFZ Genomics and Proteomics Core Facility and DKFZ ODCF System Administration. Graphical illustrations were created with BioRender.com.

This first project of the thesis described in **Chapter 3** is currently in preparation for a manuscript, with the aim to submit by March 2023 to *Nature.*

1.2 Combinatorial BCL-2 family expression in Acute Myeloid Leukemia Stem Cells predicts clinical response to Azacytidine/Venetoclax

Co-authors:
Alexander Waclawiczek[*], Aino-Maija Leppä[*], Simon Renders[*], Karolin Stumpf[§], Cecilia Reyneri[§], Barbara Betz[§], Maike Janssen, Rabia Shahswar, Elisa Donato, Darja Karpova, Vera Thiel, Julia M. Unglaub, Susanna Grabowski, Stefanie Gryzik, Lisa Vierbaum, Richard F. Schlenk, Christoph Röllig, Michael Hundemer, Caroline Pabst, Michael Heuser, Simon Raffel, Carsten Müller-Tidow, Tim Sauer[#], and Andreas Trumpp[#]

[*,§,#]Contributed equally

I together with Dr. Alexander Waclawiczek, Dr. Simon Renders, Dr. Tim Sauer, Prof. Carsten Müller-Tidow and Prof. Andreas Trumpp conceived the project. I together with Dr. Alexander Waclawiczek and Dr. Simon Renders carried out the experiments, including patient sample processing, *in vitro* drug screening and intracellular staining experiments. Karolin Stumpf, Cecilia Reyneri and Barbara Betz helped with these experiments. The *in vivo* transplantation experiments were performed by Dr. Alexander Waclawiczek and Dr. Simon Renders. The RNA-seq libraries were generated with the help of Dr. Simon Renders and Cecilia Reyneri and I analyzed the data. I together with Dr. Alexander Waclawiczek and Dr. Simon Renders performed all other data analysis. Dr. Elisa Donato, Dr. Darja Karpova and Vera Thiel contributed to patient sample processing. Dr. Richard F. Schlenk contributed to statistical analyses. Dr. Simon Renders collected patient data and Dr. Simon Renders, Dr. Julia M. Unglaub, Dr. Susanna Grabowski, Dr. Stefanie Gryzik, Dr. Richard F. Schlenk, Prof. Christoph Röllig, Prof. Michael Hundemer, Dr. Caroline Pabst, Prof. Michael Heuser, Dr. Simon Raffel, Prof. Carsten Müller-Tidow, and Dr. Tim Sauer provided samples and clinical information.

I would like to thank all technicians and in particular Markus Sohn of the Trumpp laboratory for technical assistance; Steffen Schmitt, Marcus Eich, Klaus Hexel, Tobias Rubner, and Florian Blum from the DKFZ Flow Cytometry Core Facility for their assistance; and K Reifenberg, P. Prückl, M. Durst, A. Rathgeb, and animal caretaker of the DKFZ Central Animal Laboratory for excellent animal welfare and husbandry. I also thank the DKFZ Genomics and Proteomics Core Facility for their assistance, as well as the DKFZ ODCF System Administration.

This second project of the thesis described in **Chapter 4** has been provisionally accepted to *Cancer Discovery*.

Acknowledgements

The last four years and nine months I have called Heidelberg, DKFZ and HI-STEM my home. During this time, I have learned more than I could ever have imagined, experienced the highs of new exciting results and the lows of failed experiments, worried that the PhD will never end, distressed that eventually it does all come to an end, and gained friendships that I hope to last the rest of my life. I believe that most people who have undertaken the task of doing a PhD, will agree with the rollercoaster of feelings I am trying to describe here. While the results of this thesis cover the main scientific findings of my PhD, they fail to cover the spectrum of emotions that have gone into the last years as a PhD student. Importantly, none of the work described here nor the great time as a PhD student could have been possible without a team of people supporting me.

First of all, I would like to thank my PhD supervisor Prof. Andreas Trumpp for the opportunity to do my PhD in his group. Andreas' enthusiasm on science, excitement about new findings, and table tennis skills are a true inspiration for me. He gave me the freedom to pursue several interesting projects and take them to directions I found interesting. He also trusted me with significant tasks such as helping with the ERC grant, all of which has helped me grow as a scientist; for this I am very thankful. I am also forever grateful for the atmosphere Andreas has created at HI-STEM, making the workplace a fun and supportive environment filled with great science and amazing people. I would also like to thank my thesis advisory committee members, Dr. Michael Milsom, Dr. Jan Korbel and Dr. Ashely Sanders, for their feedback and support at the yearly meetings but also for all the guidance outside the mandatory meetings and advice especially on the CK-AML project. For this, I am very appreciative.

I would also like to show my continuous appreciation and admiration to several former and current members of the Trumpp group. First, Dr. Darja Karpova who from day one not only helped and supported me with all science-related but was also there for me as a friend. I admire Darja's positive energy and kindness, willingness to take the time to listen, her vast knowledge on anything biology and science-related and excitement to test out some of my dance-moves. Thank you for being the biotin to my streptavidin. Second, I would like to thank Dr. Alexander Waclawiczek and Dr. Simon Renders as well as the whole Venetoclax-team. It has been highly inspiring to be part of this team with Alex and Simon. They show endless excitement towards science, often resulting in lengthy discussion in WhatsApp pinballing ideas and newest results.

This has not only contributed to some of the exciting science described in this thesis but also to fun times and adventures outside the lab. Many thanks also to all other current and past members of the Trumpp group who I have had the opportunity to work with.

Outside the lab, I am very grateful for the DKFZ core facilities, in particular the FACS core facility and the scOpenLab. Thank you for all the help and support with the important experiments and for teaching me the secrets of flow cytometry and single-cell libraries. I am grateful for the helpful administrative and financial support by Dagmar Wolf and Erika Krückel. Many thanks also to the whole of HI-STEM; I have enjoyed working with all of you. I would also like to thank my close collaborators Dr. Karen Grimes and Dr. Hyobin Jeong with whom I haves shared the excitement as well as frustration towards the complexity of CK-AML. They have contributed not only enthusiasm but also great science to the CK-AML project described in this thesis.

Importantly, my PhD life would not have been the same without my fellow PhD students, including but not limited to the members of the Milk Club and the Lunch Crew: Vera Thiel, Andreas Narr, Jonas Schwickert, Sarah-Jane Neuberth, Carolin Andresen, Manuel Mastel, Tasneem Cheytan, Tim Vorberg, Paul Schwerd-Kleine and Shubhankar Sood. I would like to thank all of them for the fun times in as well as outside the lab. A special glowing thank you to Vera, Andi and Jonas who made my time in Heidelberg such a great experience filled with so many good memories. Vera, for the jogging-vents, hikes to the monastery and discussions about anything and everything; Andi, for the tennis, biking and all the times spent on your balcony; Jonas, for the bakery club, fun facts about Finnish people and times on the roof top; to all three for the shared laughs, the (occasional) hugs, the struggles, frequent parties and bigger adventures around Heidelberg and abroad. I could not have managed without you, and I will forever cherish our time together at HI-STEM.

Finally, I would like to thank my family for all the support not only during my PhD but also throughout my life. They have inspired me, encouraged me to chase my dreams and believed in me to succeed; for this I am truly grateful. Last, thank you to Constantin for being my calming rock, everyday support and making my life brighter.

Appendix

Supplementary Items Related to Chapter 3

Appendix Table 1 Patient characteristics in Chapter 3.

Patient ID	CK282	CK295	CK397	CK349
Gender	Female	Female	Female	Male
Diagnosis	sAML	AML	AML	AML
Previous malignancy	PMF	-	-	-
Age at diagnosis	77	29	79	59
Treatment	not treated	7+3 | alloHSCT	Azacytidine	7+3 | Decitabine | alloHSCT
Initial response	-	CR	refractory	PR
Karyogram	46,XX,del(5)(q14-21q35)[4]/41~50,X,-X[16],-5[17],del(5)(q14~21q35)[6],add(6)(p12)[17],-7[3],-8[4],-9[19],add(11)(p15)[20],-12[20],add(12)(q24.3)[16],-13[3],-16[21],-17[20],-20[3],add(20)(q13.3)[10],del(20)(q11.2)[2],-21[20],+2~8mar[22][cp23]	45,XX,add(8)(p11.2),-11,-14,-16,+21,+mar[6]/46,sl,+add(8)(p11.2)[18]/46,XX[1]	42,X,-X,?3q,del(5)(q31),-7,-17,-18,der(21)t(7;21)(q11.2;q11.2)[27]	44,XY,+del(1)(q12),der(1)dup(1)(q32q21)t(1;20)(p12;p11.1),del(4)(q21q33),del(5)(q15q33),-7,+8,add(12)(p13),add(13)(p11.2),add(14)(p11.2),-16,-17,-20[25]
M-FISH	der(6)t(6;17), der(11)t(6;11), der(12)t(X;12), der(20)t(20;20), der(5)t(5;21;12;17;21), der(21)t(9;21)			
TP53	p.R196P (VAF=0.05)	wt	frameshift	p.Y220C
NPM1	wt	wt	wt	wt
CEBPA	wt	wt	wt	wt

IDH1/2	wt	wt	wt	wt
FLT3-ITD	neg	neg	NA	neg
NRAS	wt	p.G12D	wt	wt
CALR	frameshift	wt	wt	wt

Abbreviations: sAML, Secondary AML; PMF, Primary myelofibrosis; wt, Wild-type; neg, negative; CR, Complete response; PR, Partial response; alloHSCT, allogeneic hematopoietic stem cell transplantation; VAF, Variant allele frequency

Appendix Table 2 Co-segregation segments for CK397 using pq option of TranslocatoR.

Segment A	Segment B	Correlation	p-value	x	n	pBH
chrX.p.H2	chr11.p.H1	0.93	5.49E+07	28	29	2.16E+09
chr21.p.H2	chr7.p.H1	0.86	2.91E+08	27	29	9.82E+09
chr21.q.H2	chr7.p.H1	0.79	2.19E+08	26	29	0.00072
chr11.q.H1	chrX.p.H2	0.78	5.00E+09	26	29	0.0016
chr8.q.H1	chr2.p.H1	-0.73	0.00012	4	29	0.0036

Abbreviations: x, Number of matching states; n, Number of cells; pBH, FDR-adjusted p-value

Appendix Table 3 Top 20 differentially upregualted genes in each subclone in CK282.

p-value	FDR	logFC	Gene	Subclone
3,8E-132	7,4E-128	1,04	*LST1*	Subclone1
1,5E-81	2,2E-78	0,89	*CD3D*	Subclone1
1,9E-108	7,5E-105	1,07	*IL2RG*	Subclone1
9,3E-116	7,3E-112	0,87	*CD164*	Subclone1
1,7E-39	4,5E-37	0,92	*CSTB*	Subclone1
4,0E-35	9,2E-33	1,28	*ETS2*	Subclone1
5,7E-30	9,6E-28	0,82	*SEM1*	Subclone1
1,4E-94	4,0E-91	0,85	*TESC*	Subclone1
3,0E-27	4,2E-25	1,27	*MARCKSL1*	Subclone1
1,4E-25	1,7E-23	0,96	*SNHG15*	Subclone1
3,2E-78	3,8E-75	0,91	*S100A11*	Subclone1
2,4E-77	2,7E-74	0,86	*GSN*	Subclone1
1,0E-22	1,0E-20	0,95	*CD44*	Subclone1
1,8E-21	1,6E-19	1,03	*GNAI1*	Subclone1
5,5E-21	4,9E-19	0,82	*STMP1*	Subclone1
6,3E-21	5,6E-19	0,84	*CDK6*	Subclone1
5,7E-18	4,0E-16	0,92	*CYTL1*	Subclone1
9,6E-41	2,7E-38	0,86	*LTB*	Subclone1
9,4E-12	3,5E-10	0,90	*ZFP36L2*	Subclone1
5,9E-11	2,0E-09	0,83	*ID2*	Subclone1

7,5E-73	1,4E-68	1,95	*DLK1*	Subclone2
1,6E-36	1,3E-33	0,99	*CSTB*	Subclone2
9,1E-56	2,9E-52	1,02	*MIR3681HG*	Subclone2
8,0E-31	4,0E-28	0,93	*SEM1*	Subclone2
4,6E-54	1,3E-50	1,02	*LST1*	Subclone2
5,8E-24	1,7E-21	0,83	*MTPN*	Subclone2
3,3E-46	4,5E-43	0,93	*TESC*	Subclone2
5,9E-45	7,5E-42	1,07	*IL2RG*	Subclone2
2,6E-19	4,8E-17	1,04	*ETS2*	Subclone2
1,4E-17	2,1E-15	1,13	*MARCKSL1*	Subclone2
8,3E-32	5,2E-29	1,23	*LTB*	Subclone2
9,2E-32	5,5E-29	0,90	*THY1*	Subclone2
9,7E-32	5,7E-29	0,83	*PDLIM1*	Subclone2
6,0E-17	8,4E-15	0,86	*SNHG15*	Subclone2
1,6E-31	8,8E-29	0,84	*CST3*	Subclone2
3,9E-16	5,0E-14	1,03	*TFPI*	Subclone2
1,1E-14	1,1E-12	0,94	*GNAI1*	Subclone2
1,2E-14	1,2E-12	0,88	*HOPX*	Subclone2
1,2E-14	1,2E-12	0,92	*CYTL1*	Subclone2
6,1E-11	3,8E-09	0,93	*ID2*	Subclone2
3,4E-199	3,3E-195	1,24	*HM13*	Subclone3
2,9E-216	5,6E-212	1,27	*CPNE1*	Subclone3
8,4E-139	1,8E-135	0,93	*COMMD7*	Subclone3
1,7E-88	2,0E-85	1,18	*RALY*	Subclone3
6,5E-199	4,2E-195	1,09	*RAB5IF*	Subclone3
3,8E-187	1,8E-183	1,00	*EIF6*	Subclone3
1,4E-183	5,5E-180	1,03	*ROMO1*	Subclone3
5,0E-97	8,1E-94	0,91	*AHCY*	Subclone3
4,6E-175	1,5E-171	0,92	*RBM39*	Subclone3
7,5E-41	2,0E-38	0,94	*CSTB*	Subclone3
1,1E-93	1,7E-90	0,99	*PDLIM1*	Subclone3
8,4E-163	2,3E-159	0,92	*SCAND1*	Subclone3
9,0E-153	2,2E-149	0,94	*PHF20*	Subclone3
6,3E-87	6,7E-84	0,98	*GSN*	Subclone3
1,6E-82	1,5E-79	1,00	*S100A11*	Subclone3
2,9E-31	4,4E-29	1,14	*CD44*	Subclone3
1,3E-30	1,8E-28	1,20	*ETS2*	Subclone3
6,5E-27	8,2E-25	0,99	*SNHG15*	Subclone3
4,3E-20	3,7E-18	1,00	*GNAI1*	Subclone3
2,2E-15	1,2E-13	0,94	*MARCKSL1*	Subclone3
0,0E+00	0,0E+00	0,86	*RPS6*	Subclone4
1,7E-258	1,1E-254	1,42	*MAN1A1*	Subclone4
1,5E-40	4,8E-38	1,43	*ETS2*	Subclone4
3,0E-30	5,1E-28	0,86	*PPDPF*	Subclone4

1,7E-156	6,5E-153	1,47	*MDK*	Subclone4
1,5E-134	4,1E-131	1,39	*GMFG*	Subclone4
7,9E-106	1,2E-102	0,95	*CEBPB*	Subclone4
2,6E-110	4,2E-107	0,97	*SELL*	Subclone4
2,9E-78	2,4E-75	0,95	*BEX1*	Subclone4
8,5E-84	9,1E-81	0,81	*SMARCA2*	Subclone4
1,5E-87	2,1E-84	0,81	*ZFAS1*	Subclone4
9,6E-23	1,1E-20	0,89	*CD99*	Subclone4
6,3E-21	6,3E-19	0,89	*SNHG15*	Subclone4
1,8E-18	1,6E-16	1,07	*MARCKSL1*	Subclone4
6,9E-62	4,4E-59	0,83	*MLLT3*	Subclone4
1,4E-17	1,1E-15	0,96	*GNAI1*	Subclone4
2,0E-16	1,4E-14	1,07	*ID2*	Subclone4
1,0E-14	6,5E-13	0,85	*CYTL1*	Subclone4
7,0E-49	2,9E-46	0,90	*SOX4*	Subclone4
3,0E-43	1,0E-40	1,15	*AREG*	Subclone4
3,1E-246	3,0E-242	1,82	*AVP*	Subclone5
4,4E-111	4,7E-108	1,33	*BEX2*	Subclone5
2,0E-89	1,3E-86	1,40	*XIST*	Subclone5
4,6E-60	1,7E-57	1,17	*LGALS3BP*	Subclone5
7,4E-60	2,6E-57	1,85	*HLA-DPA1*	Subclone5
1,6E-51	4,1E-49	2,30	*ANXA1*	Subclone5
1,9E-51	4,8E-49	1,60	*HLA-DPB1*	Subclone5
4,7E-49	1,1E-46	1,59	*HLA-B*	Subclone5
1,9E-53	5,3E-51	1,44	*ALDH1A1*	Subclone5
2,8E-53	7,6E-51	1,12	*IGFBP2*	Subclone5
1,7E-35	2,6E-33	1,58	*NUCB2*	Subclone5
1,6E-38	2,8E-36	1,76	*GMFG*	Subclone5
2,8E-33	4,0E-31	1,38	*ATF7IP2*	Subclone5
1,7E-28	2,1E-26	1,18	*CORO1A*	Subclone5
2,7E-28	3,2E-26	1,33	*HLA-DRB1*	Subclone5
3,4E-27	3,8E-25	1,31	*HLA-DRA*	Subclone5
4,0E-26	4,3E-24	1,31	*SRGN*	Subclone5
1,9E-23	1,8E-21	1,27	*NPW*	Subclone5
8,4E-21	7,2E-19	1,12	*HCST*	Subclone5
1,3E-15	7,9E-14	1,12	*PRSS57*	Subclone5

Abbreviations: FDR, False Discovery Rate.

Appendix Table 4 Top 20 differentially upregualted genes in each subclone in CK349.

p-value	FDR	logFC	Gene	Subclone
1,43E-53	2,66E-49	0,32	*RPL30*	Subclone1
2,75E-50	2,55E-46	0,35	*RPL7*	Subclone1

2,55E-44	1,58E-40	0,33	*DPPA4*	Subclone1
3,99E-43	1,85E-39	0,56	*PABPC1*	Subclone1
1,65E-38	6,14E-35	0,37	*RPS20*	Subclone1
1,68E-29	4,46E-26	0,39	*SNHG6*	Subclone1
6,64E-26	1,37E-22	0,46	*LY6E*	Subclone1
1,05E-23	1,78E-20	0,38	*EIF3E*	Subclone1
5,88E-19	8,41E-16	0,45	*TCEA1*	Subclone1
1,13E-17	1,50E-14	0,38	*IER2*	Subclone1
3,26E-17	4,04E-14	0,38	*YWHAZ*	Subclone1
5,91E-17	6,87E-14	0,31	*KLF10*	Subclone1
3,36E-16	3,47E-13	0,35	*ANGPT1*	Subclone1
1,52E-14	1,35E-11	0,69	*FOS*	Subclone1
9,25E-14	6,88E-11	0,31	*SARAF*	Subclone1
2,63E-13	1,81E-10	0,43	*S100A10*	Subclone1
5,60E-13	3,59E-10	0,31	*JUND*	Subclone1
2,11E-11	1,12E-08	0,44	*HSPA1A*	Subclone1
6,29E-11	2,79E-08	0,45	*CRIP1*	Subclone1
1,22E-10	4,72E-08	0,35	*HSPA1B*	Subclone1
9,92E-17	1,85E-12	0,39	*COMMD6*	Subclone2
2,80E-09	2,61E-05	0,30	*SPINT2*	Subclone2
5,78E-09	3,58E-05	0,08	*RPS15*	Subclone2
9,90E-09	4,61E-05	0,25	*PBXIP1*	Subclone2
5,23E-08	1,95E-04	0,20	*SELL*	Subclone2
6,69E-08	2,08E-04	0,07	*RPL28*	Subclone2
1,07E-07	2,83E-04	0,10	*PTMA*	Subclone2
3,80E-07	7,79E-04	0,25	*CTSW*	Subclone2
4,19E-07	7,79E-04	0,32	*CLEC11A*	Subclone2
1,18E-06	1,98E-03	0,14	*NOP53*	Subclone2
1,32E-06	1,98E-03	0,09	*ATP5MC2*	Subclone2
1,39E-06	1,98E-03	0,11	*RPS28*	Subclone2
1,59E-06	2,12E-03	0,07	*RPL18*	Subclone2
3,94E-06	4,58E-03	0,07	*RPL38*	Subclone2
5,74E-06	6,26E-03	0,22	*IGLL1*	Subclone2
6,06E-06	6,26E-03	0,07	*RPS26*	Subclone2
1,19E-05	1,17E-02	0,17	*SPINK2*	Subclone2
1,38E-05	1,29E-02	0,17	*LEPROT*	Subclone2
1,96E-05	1,74E-02	0,16	*GAPDH*	Subclone2
4,56E-05	3,86E-02	0,12	*LINC01268*	Subclone2
1,06E-55	1,96E-51	1,01	*ATP5MG*	Subclone3
2,11E-51	1,31E-47	0,86	*APLP2*	Subclone3
9,85E-46	4,58E-42	0,97	*KMT2A*	Subclone3
5,83E-44	2,17E-40	1,03	*IGHM*	Subclone3
2,33E-39	7,23E-36	1,03	*PRDX1*	Subclone3
4,37E-38	1,16E-34	0,79	*POLR2L*	Subclone3

1,30E-37	3,02E-34	0,82	*CTSD*	Subclone3
2,25E-30	4,64E-27	0,58	*ZBTB44*	Subclone3
2,04E-29	3,45E-26	0,57	*TALDO1*	Subclone3
2,01E-28	3,11E-25	0,58	*TSPAN4*	Subclone3
2,94E-28	4,21E-25	0,63	*ARCN1*	Subclone3
1,28E-27	1,70E-24	0,68	*CKB*	Subclone3
1,96E-26	2,28E-23	0,79	*FXYD6*	Subclone3
3,11E-25	3,05E-22	0,92	*RPL21*	Subclone3
1,23E-24	1,15E-21	0,58	*CTHRC1*	Subclone3
8,66E-20	5,55E-17	0,98	*POLR1D*	Subclone3
4,80E-18	2,55E-15	0,58	*ALDH1A1*	Subclone3
4,47E-16	1,73E-13	0,92	*FLT3*	Subclone3
6,04E-16	2,29E-13	0,70	*MTIF3*	Subclone3
7,92E-15	2,20E-12	0,81	*GTF3A*	Subclone3

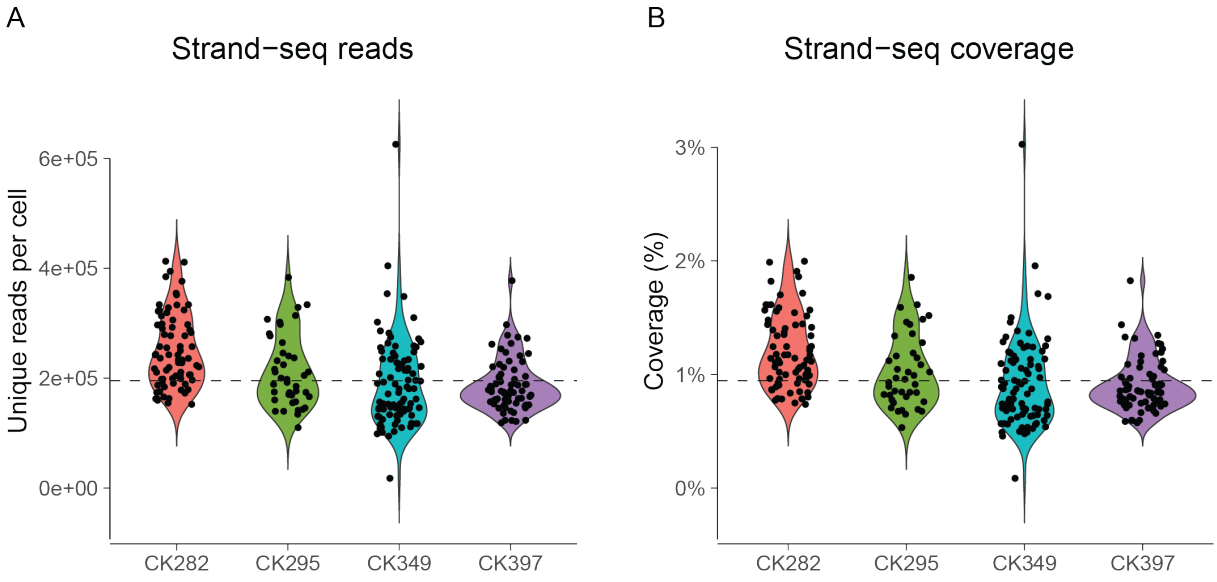
Abbreviations: FDR, False Discovery Rate.

Appendix Table 5 Top 60 differentially upregulated genes in the engraftment driving cells in CK282 and CK349.

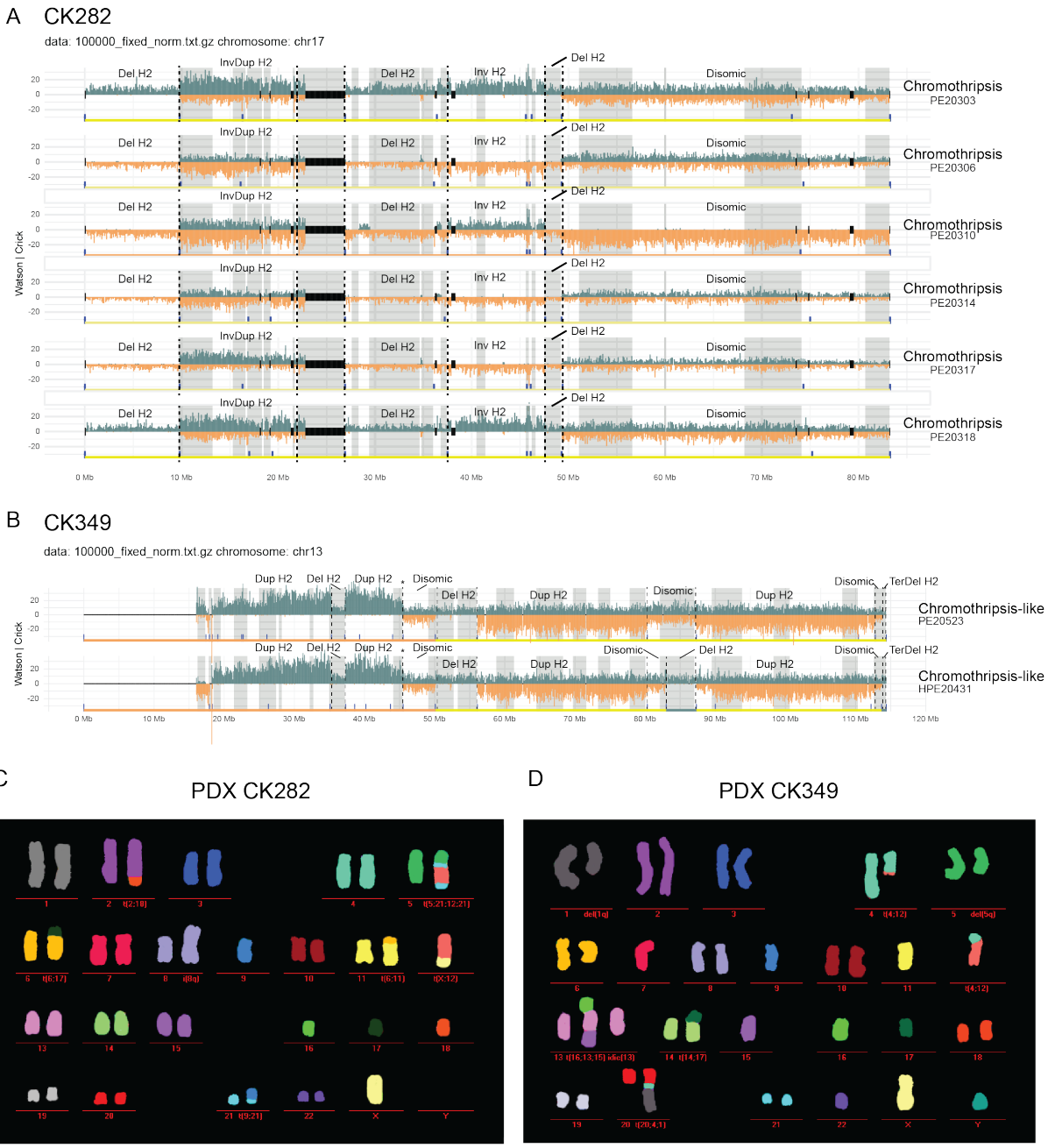
CK282				CK349			
p-value	**FDR**	**logFC**	**Gene**	**p-value**	**FDR**	**logFC**	**Gene**
2,2E-72	4,2E-68	0,40	*RPL8*	6,2E-71	1,1E-66	1,20	*ATP5MG*
3,5E-71	3,3E-67	1,80	*DLK1*	8,8E-66	6,8E-62	1,30	*IGHM*
2,8E-67	1,8E-63	0,31	*CHRDL1*	1,1E-65	6,8E-62	0,63	*RPLP2*
8,3E-60	4,0E-56	0,63	*FYB1*	1,8E-57	8,5E-54	1,20	*PRDX1*
6,3E-52	2,0E-48	0,35	*RPL30*	1,8E-55	6,6E-52	0,88	*POLR2L*
1,0E-51	2,8E-48	0,40	*RPL7*	1,2E-54	3,6E-51	0,89	*APLP2*
9,8E-45	2,3E-41	0,48	*UQCRB*	2,7E-52	7,1E-49	0,97	*CTSD*
4,5E-41	8,7E-38	0,28	*FCGR2B*	5,5E-51	1,3E-47	0,97	*KMT2A*
7,2E-40	1,3E-36	0,30	*LY86*	1,9E-47	3,9E-44	0,81	*SAP18*
7,4E-38	1,1E-34	0,27	*JAML*	4,5E-45	8,4E-42	0,85	*MRPL57*
7,8E-38	1,1E-34	0,48	*SNHG6*	1,6E-44	2,7E-41	0,67	*TALDO1*
2,2E-36	3,0E-33	0,72	*MIR3681HG*	5,0E-36	7,7E-33	0,51	*RPL27A*
9,0E-35	1,1E-31	0,42	*PABPC1*	6,5E-35	9,4E-32	0,68	*AKR1A1*
2,4E-34	2,7E-31	0,28	*IGFBP4*	4,2E-34	5,6E-31	0,64	*TEX30*
4,0E-30	3,8E-27	0,32	*RPS20*	8,0E-34	9,9E-31	0,66	*ARCN1*
5,4E-30	5,0E-27	0,41	*COX6C*	1,2E-33	1,3E-30	0,68	*MPHOSPH8*
6,2E-30	5,4E-27	0,39	*DMKN*	3,5E-32	3,8E-29	0,65	*CD164*
6,8E-27	5,7E-24	0,52	*CA8*	3,9E-32	3,8E-29	0,56	*MYCBP2*
9,9E-27	8,0E-24	0,33	*CYP7B1*	3,9E-32	3,8E-29	0,60	*SRPRA*
4,3E-25	3,3E-22	0,54	*AC090152.1*	4,5E-32	4,2E-29	0,59	*ZBTB44*
3,6E-24	2,6E-21	0,54	*LYN*	5,9E-32	5,2E-29	0,55	*ARGLU1*
1,0E-23	7,3E-21	0,41	*TCEA1*	1,1E-30	9,0E-28	0,79	*FXYD6*

3,9E-23	2,7E-20	0,56	*NUDT8*	2,6E-30	2,1E-27	0,59	*EI24*
2,6E-22	1,7E-19	0,29	*EEF1D*	6,8E-30	5,3E-27	0,59	*TSPAN4*
7,8E-22	4,9E-19	0,47	*VPS28*	3,1E-29	2,3E-26	0,69	*CKB*
1,3E-21	7,8E-19	0,40	*NDUFB9*	3,4E-29	2,4E-26	0,99	*RPL21*
1,7E-20	9,9E-18	0,26	*TSPAN7*	1,3E-28	9,1E-26	0,49	*PBX3*
1,6E-19	8,8E-17	0,33	*MTDH*	6,3E-28	4,2E-25	0,54	*DCUN1D5*
3,3E-19	1,6E-16	0,42	*RAB2A*	7,7E-28	4,9E-25	0,59	*ZMYM2*
3,6E-18	1,7E-15	0,41	*LSM1*	2,5E-27	1,5E-24	0,47	*ATP5F1A*
4,2E-18	1,9E-15	0,32	*EIF3H*	3,5E-27	2,1E-24	0,62	*CTHRC1*
5,3E-18	2,4E-15	0,36	*PUF60*	5,5E-27	3,2E-24	0,45	*NFRKB*
5,6E-18	2,5E-15	0,29	*TNNI2*	8,7E-27	4,9E-24	0,54	*UBE4A*
7,8E-18	3,3E-15	0,38	*ELOC*	6,1E-26	3,3E-23	0,50	*NFIA*
9,7E-18	4,1E-15	0,41	*EIF4EBP1*	2,1E-25	1,1E-22	0,55	*CTSC*
2,4E-17	9,3E-15	0,40	*TRAM1*	4,1E-25	2,1E-22	0,53	*CAVIN1*
4,5E-17	1,7E-14	0,40	*LYPLA1*	1,1E-24	5,6E-22	0,48	*CDKN2C*
5,5E-17	2,1E-14	0,39	*COPS5*	1,7E-24	8,0E-22	1,10	*POLR1D*
1,2E-16	4,2E-14	0,25	*XIRP2*	2,8E-24	1,3E-21	0,38	*SERBP1*
1,3E-16	4,7E-14	0,40	*ZHX1*	8,4E-24	3,8E-21	1,10	*FLT3*
1,5E-16	5,1E-14	0,37	*RBIS*	1,0E-23	4,4E-21	0,47	*LAMP1*
1,5E-16	5,1E-14	0,39	*UBE2V2*	1,2E-23	5,0E-21	0,54	*FCGRT*
1,7E-16	5,7E-14	0,37	*FAM49B*	2,3E-23	9,9E-21	0,47	*NAP1L4*
5,4E-16	1,8E-13	0,41	*PRKDC*	5,7E-23	2,4E-20	0,61	*PAN3*
2,1E-15	6,5E-13	0,38	*COL6A2*	1,1E-22	4,5E-20	0,44	*NDUFS8*
2,5E-15	7,7E-13	0,35	*ZNF706*	2,5E-22	9,9E-20	0,51	*ST3GAL4*
3,1E-15	9,1E-13	0,40	*SQLE*	7,8E-22	3,0E-19	0,51	*PDLIM1*
1,1E-14	3,1E-12	0,39	*LPAR6*	9,4E-22	3,6E-19	0,47	*PSPC1*
1,2E-14	3,3E-12	0,36	*GRINA*	1,3E-21	4,9E-19	0,60	*ALDH1A1*
1,6E-14	4,5E-12	0,36	*POLR2K*	2,4E-21	8,8E-19	0,55	*PAN3-AS1*
1,7E-14	4,6E-12	0,35	*CYC1*	4,7E-21	1,7E-18	0,52	*TAF10*
3,7E-14	9,9E-12	0,32	*AGO2*	6,7E-21	2,4E-18	0,34	*SELENOH*
4,9E-14	1,3E-11	0,51	*SRGN*	6,8E-21	2,4E-18	0,40	*CACNB4*
6,9E-14	1,8E-11	0,39	*RAD21*	9,8E-21	3,3E-18	0,37	*PPP1R14A*
8,8E-14	2,2E-11	0,49	*THY1*	1,3E-20	4,2E-18	0,54	*SMC4*
9,2E-14	2,3E-11	0,36	*ARHGAP4*	2,2E-20	7,1E-18	0,35	*FOXRED1*
1,0E-13	2,5E-11	0,35	*FNTA*	5,2E-20	1,7E-17	0,49	*REXO2*
3,0E-13	7,3E-11	0,33	*MAF1*	7,7E-20	2,4E-17	0,33	*TOLLIP*
5,6E-13	1,3E-10	0,30	*IFI30*	8,2E-20	2,5E-17	0,35	*POMC*
1,1E-12	2,4E-10	0,32	*EPSTI1*	9,8E-20	3,0E-17	0,51	*EIF3F*

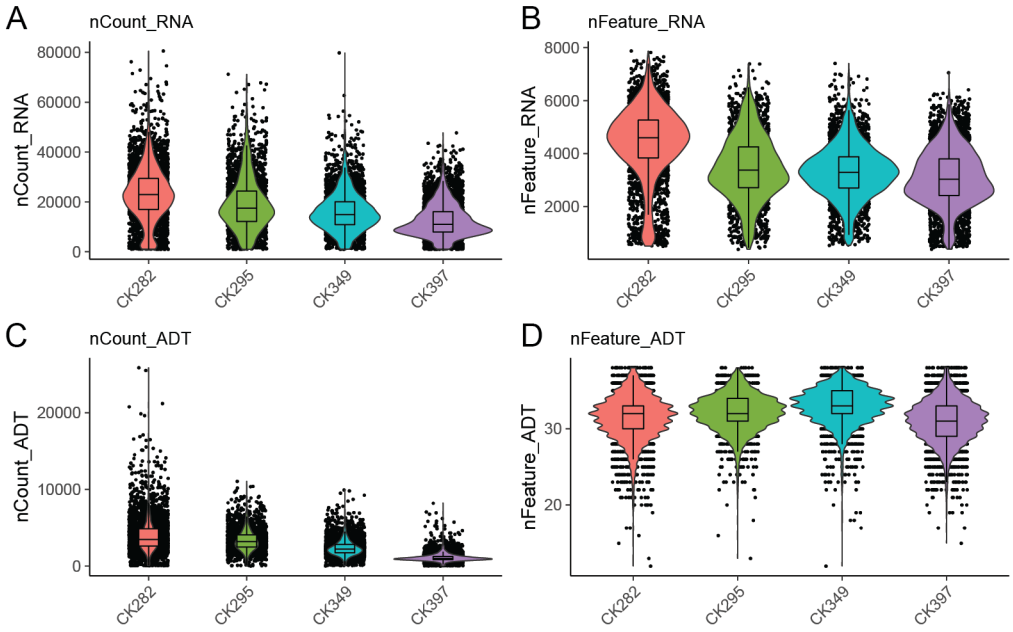
Abbreviations: FDR, False Discovery Rate.



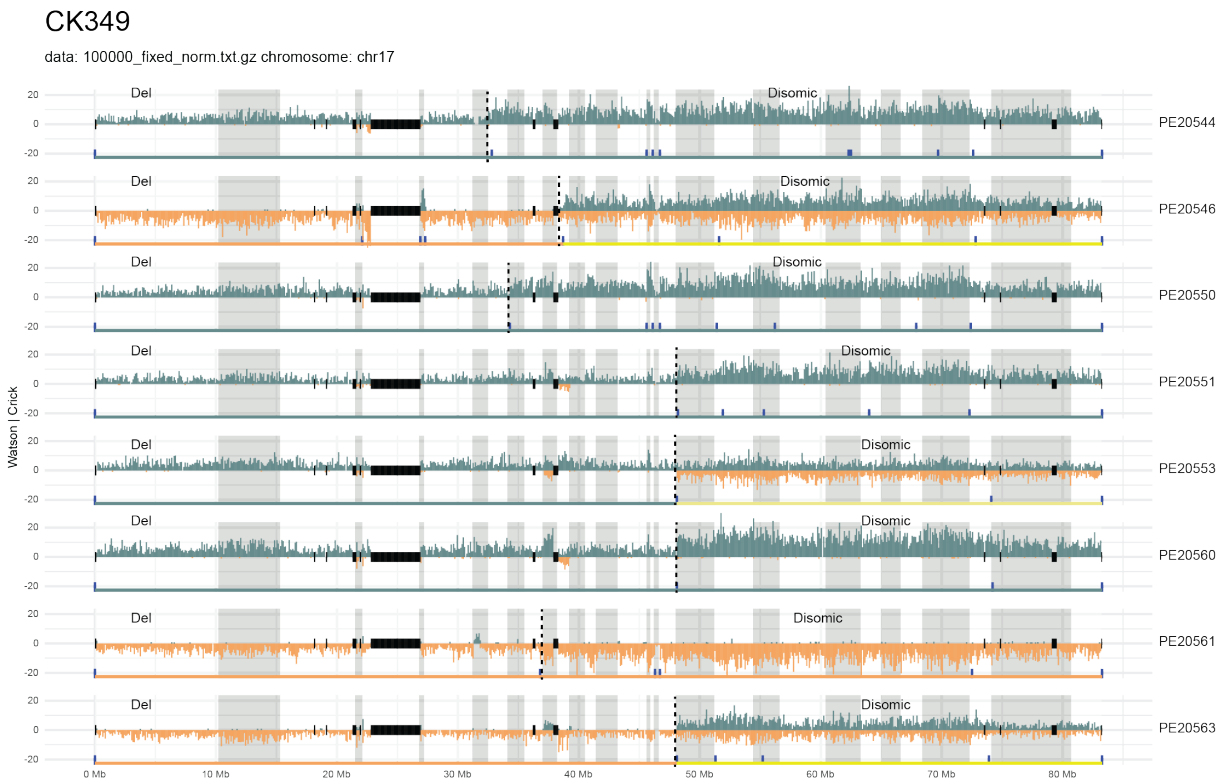
Appendix Figure 1 Strand-seq statistics. A-B Violin plots showing the **A** mapped non-duplicate read pairs and **B** read coverage of the Strand-seq libraries in each cell of the four CK-AML patient samples. Experiments performed jointly by Dr. Karen Grimes and myself.



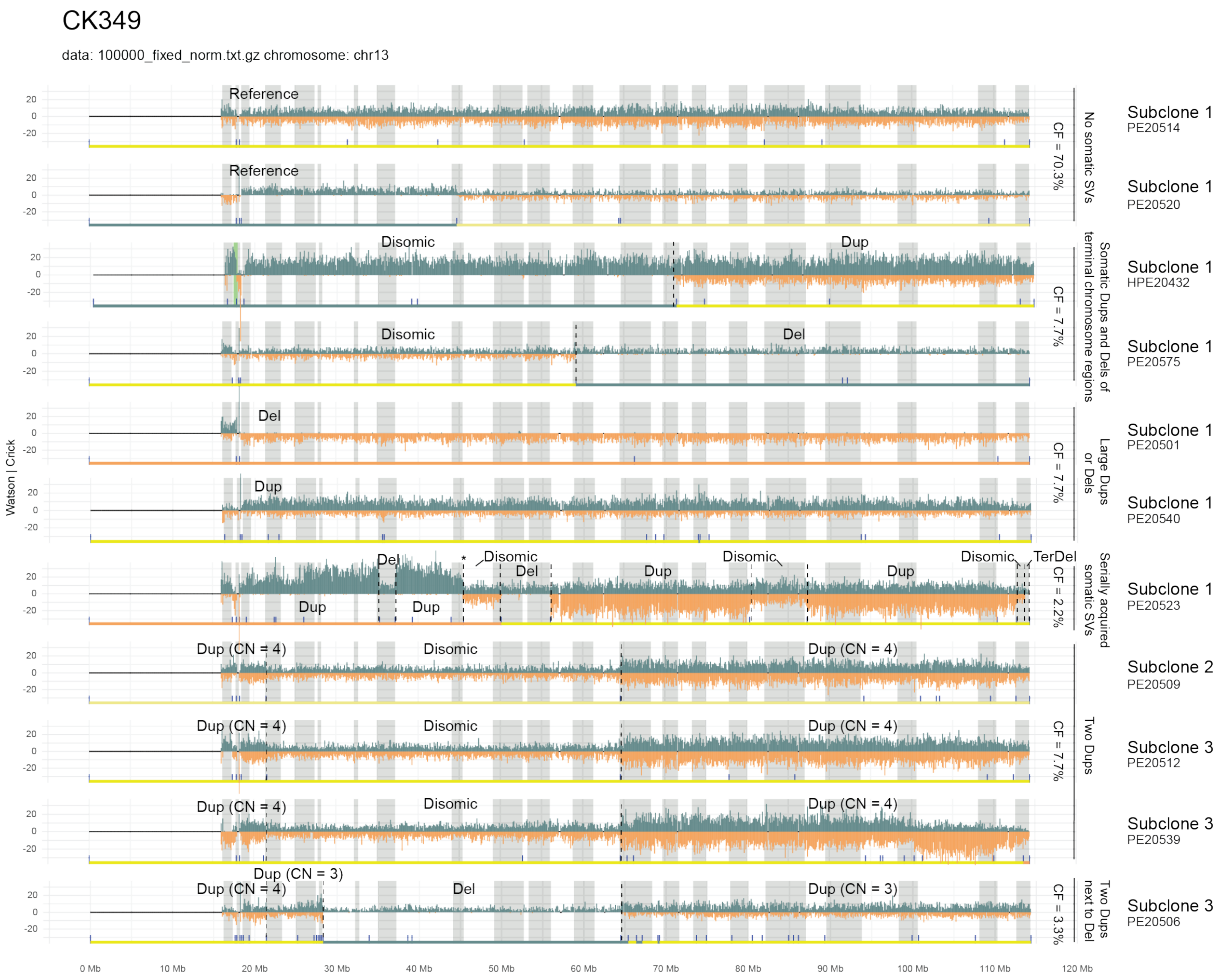
Appendix Figure 2 Chromothripsis and chromothripsis-like rearrangements in CK-AML. A Strand-specific read depth of representative single cells from CK282 showing clustered deletions, inverted duplications and inversions along a single homolog at chromosome 12, resulting from clonal chromothripsis. Reads denoting somatic SVs, discovered using scTRIP, mapped to the Watson (orange) or Crick (green) strand. Grey: single cell IDs. **B** Strand-specific read depth of two representative single cells from CK349 with chromothripsis-like rearrangements at chromosome 13. (*) indicates a switch in the strand orientation. **C-D** M-FISH of an engrafted cell from **C** PDX of CK282 and **D** PDX CK349. Abbreviations: chr, Chromosome; Del, Deletion; Inv, Inversion; Dup, Duplication; H, Haplotype; Ter, Terminal. Experiments performed jointly by Dr. Karen Grimes and myself.



Appendix Figure 3 CITE-seq statistics. A Violin plots of the number of unique molecular identified (UMI) transcripts per cell passing filter (number of distinct genes > 200 and < 8,000, number of unique molecular identifiers (UMI) > 1,000 and < 15% mitochondrial reads) per CK-AML sample. **B** Violin plots of the number of distinct genes detected per single cell passing filter per CK-AML sample. **C** Violin plots of the number of unique molecular identified (UMI) antibody-derived tags per cell passing filter per CK-AML sample. **D** Violin plots of the number of distinct antibody-derived tags detected per single cell passing filter per CK-AML sample. Abbreviations: ADT, antibody-derived tag.

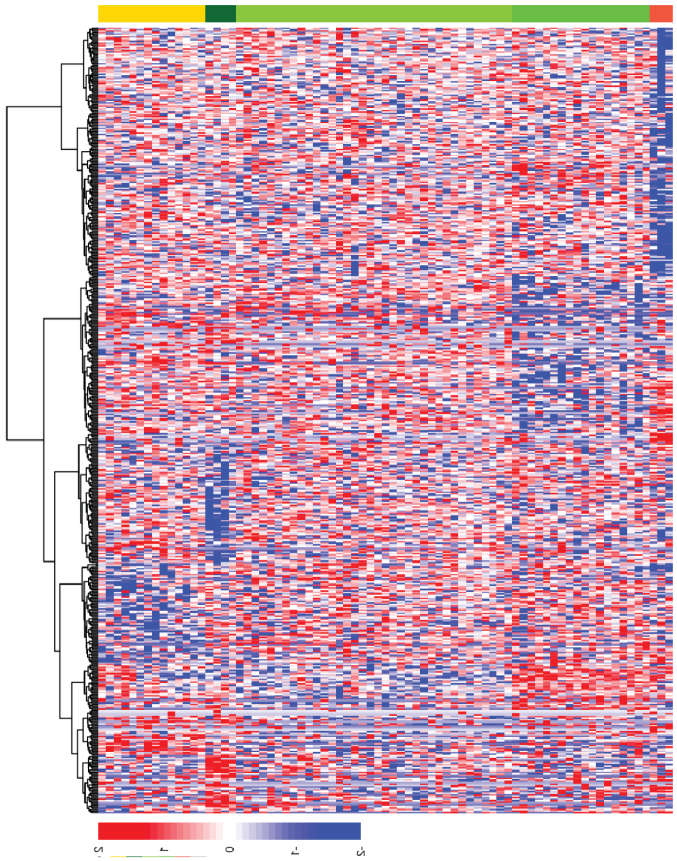


Appendix Figure 4 Ongoing chromosome instability. Strand-specific read depth of representative single cells from CK349 showing signs of active mutational processes at chromosome 17 displayed by varying breakpoints of the deletion (Del) at chromosome 17 spanning from p arm into q arm. Reads denoting somatic SVs, discovered using scTRIP, mapped to Watson (orange) or Crick (green) strand. Grey: single cell IDs. Experiments performed jointly by Dr. Karen Grimes and myself.

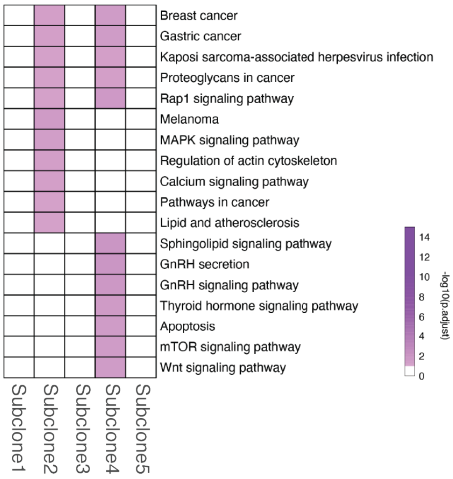


Appendix Figure 5 Subclonal heterogeneity and ongoing chromosome instability. Strand-specific read depth of representative single cells from CK349 showing different rearrangements detected at chromosome 13 in different subclones. Reads denoting somatic SVs, discovered using scTRIP, mapped to the Watson (orange) or Crick (green) strand. Abbreviations: chr, Chromosome; Del, Deletion; Dup, Duplication; Ter, Terminal. CF: Cell Fraction; CN, Copy-number. Experiments performed jointly by Dr. Karen Grimes and myself.

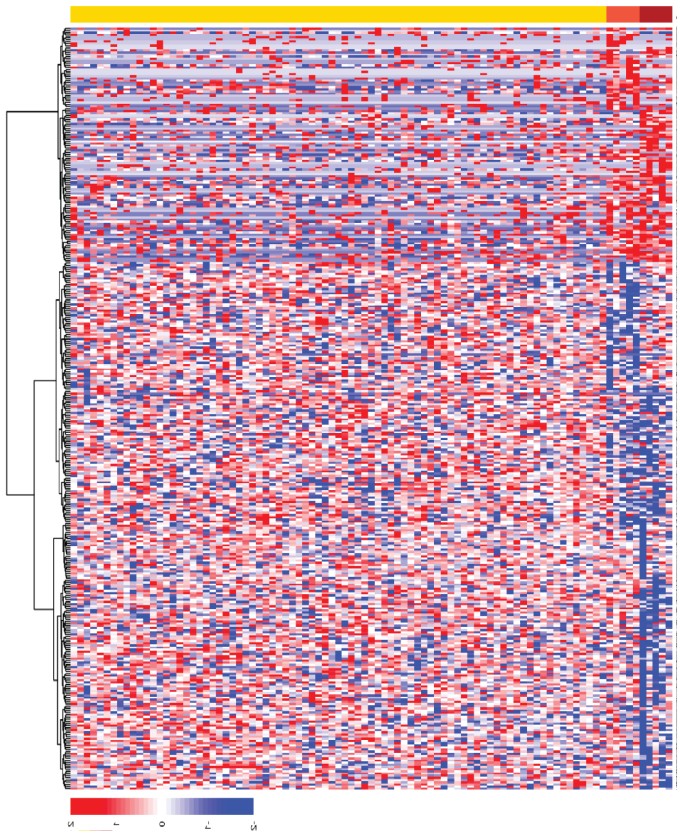
A CK282



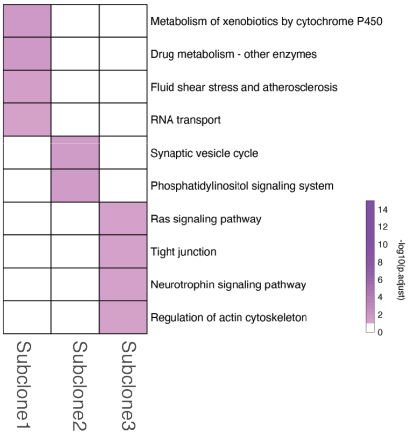
C CK282
Pathway enrichment on NO level (10% FDR)



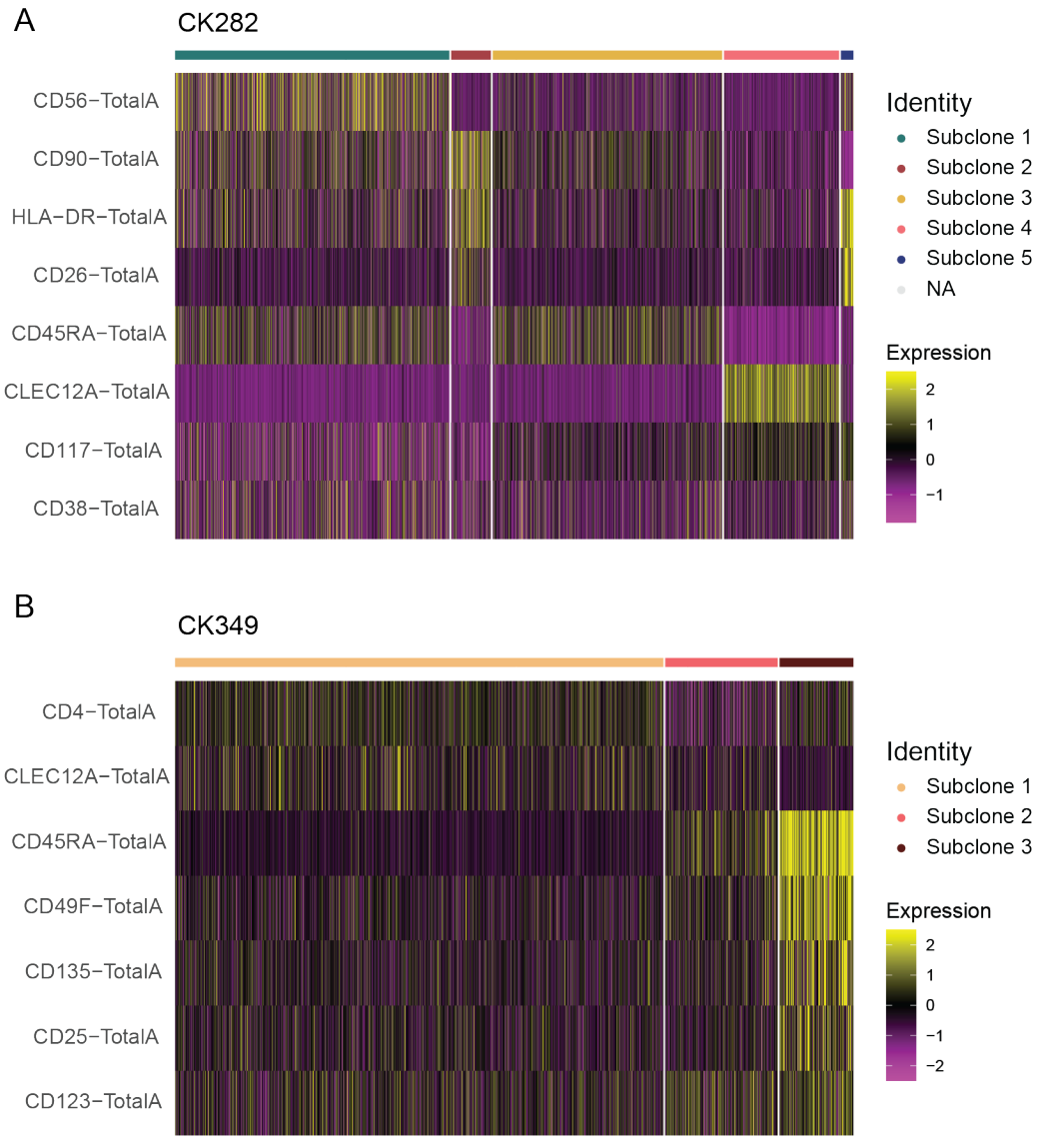
B CK349



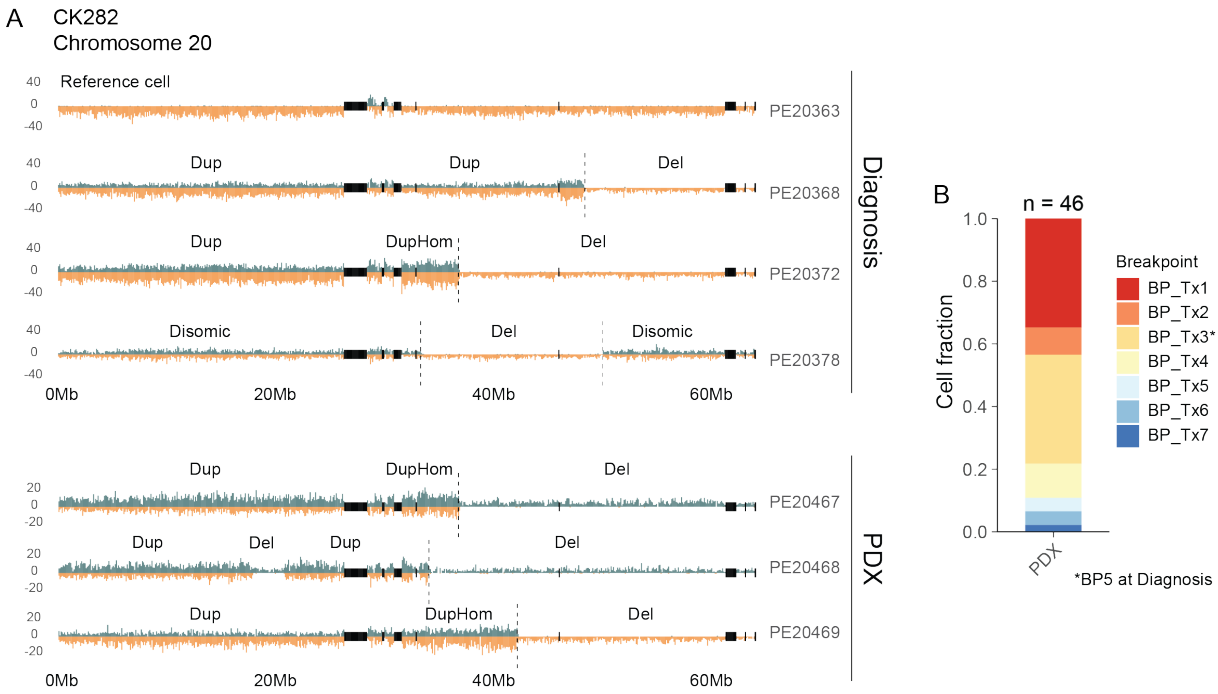
D CK349
Pathway enrichment on NO level (10% FDR)



Appendix Figure 6 Subclonal nucleosome occupancy profiles. A Heatmap of differentially dysregulated genes between subclones based on nucleosome occupancy in CK282. FDR < 0.1 was considered significant. Each row represents a dysregulated gene and each column a single cell with subclone information colored on top. **B** Same as panel (A) but in CK349. **C** Heatmap showing enriched pathway modules (FDR < 0.1) in the different subclones in CK282 based on nucleosome occupancy. Enrichment analysis was performed against the KEGG gene sets collection from MSigDB. D Same as panel (B) but in CK349. Abbreviations: NO, Nucleosome occupancy. Analysis of nucleosome occupancy data performed by Dr. Hyobin Jeong.

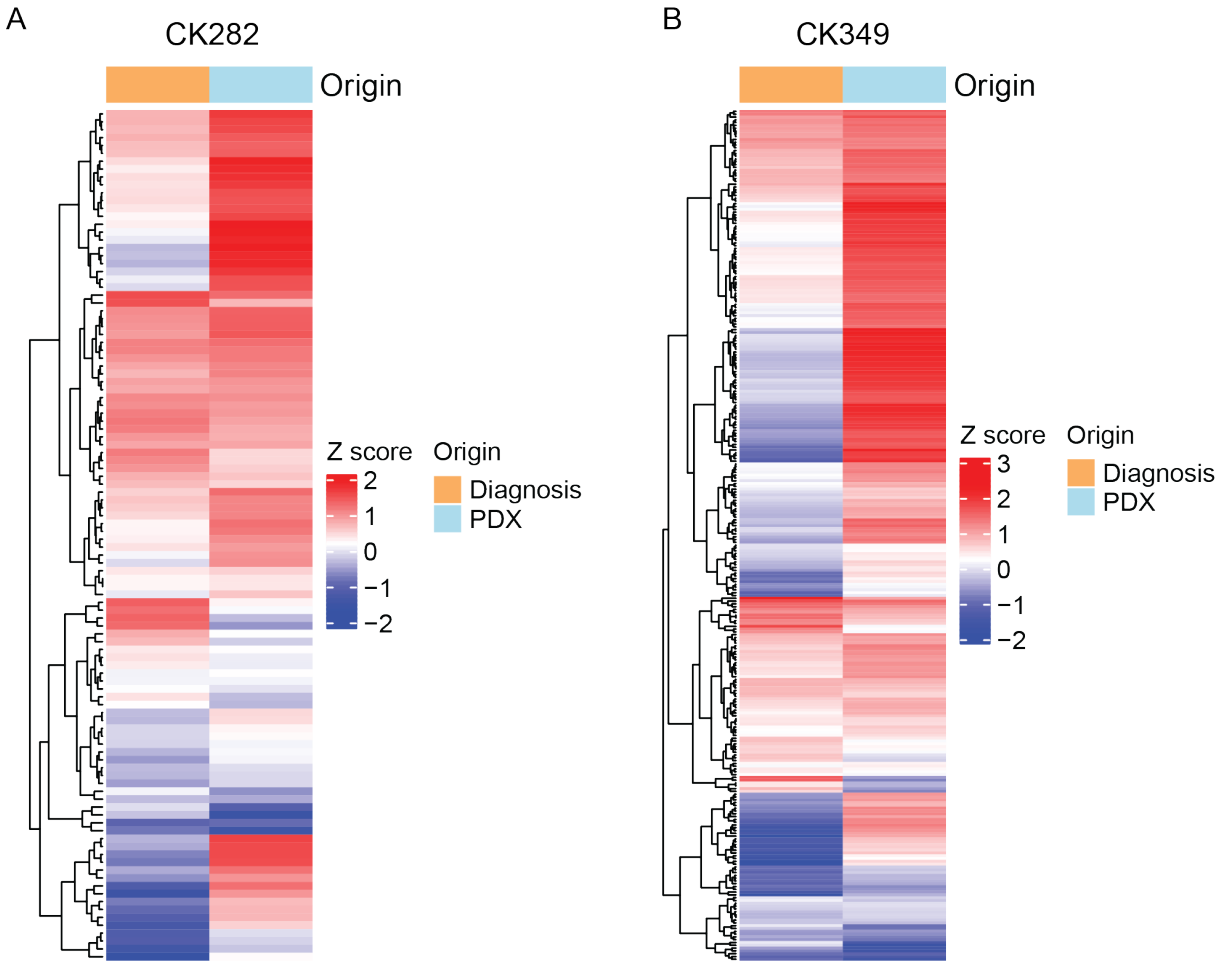


Appendix Figure 7 Subclonal cell surface protein expression. A-B Heatmap showing differentially expressed cell surface markers for each subclone in **A** CK282 and **B** CK349.

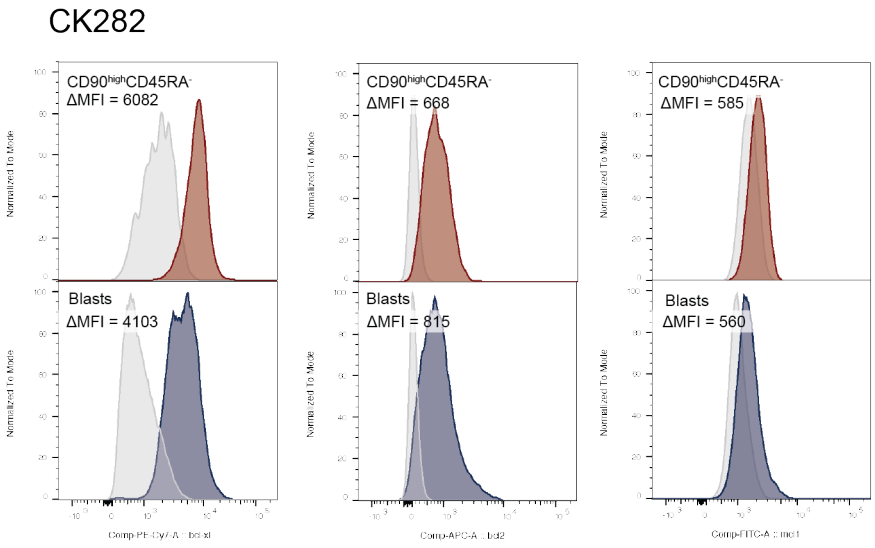


Appendix Figure 8 Ongoing chromosome instability during disease evolution. A Strand-specific read depth of representative single cells from CK282 showing signs of active mutational processes at chromosome 20 at diagnosis (top) and in the engrafted cells of the patient-derived xenograft (PDX) (bottom) displayed by varying breakpoints of the deletion (Del). Reads denoting somatic SVs, discovered using scTRIP, mapped to the Watson (orange) or Crick (green) strand. **B** Stacked bar plot showing the fraction of indicated breakpoints at chromosome 20 in the patient-derived xenograft. Number of cells analyzed is shown above the bar. Abbreviations: Dup, Duplication; Del, Deletion; Hom, Homologous; PDX, Patient-derived-xenograft; BP, Break point. Experiments performed jointly by Dr. Karen Grimes and myself.

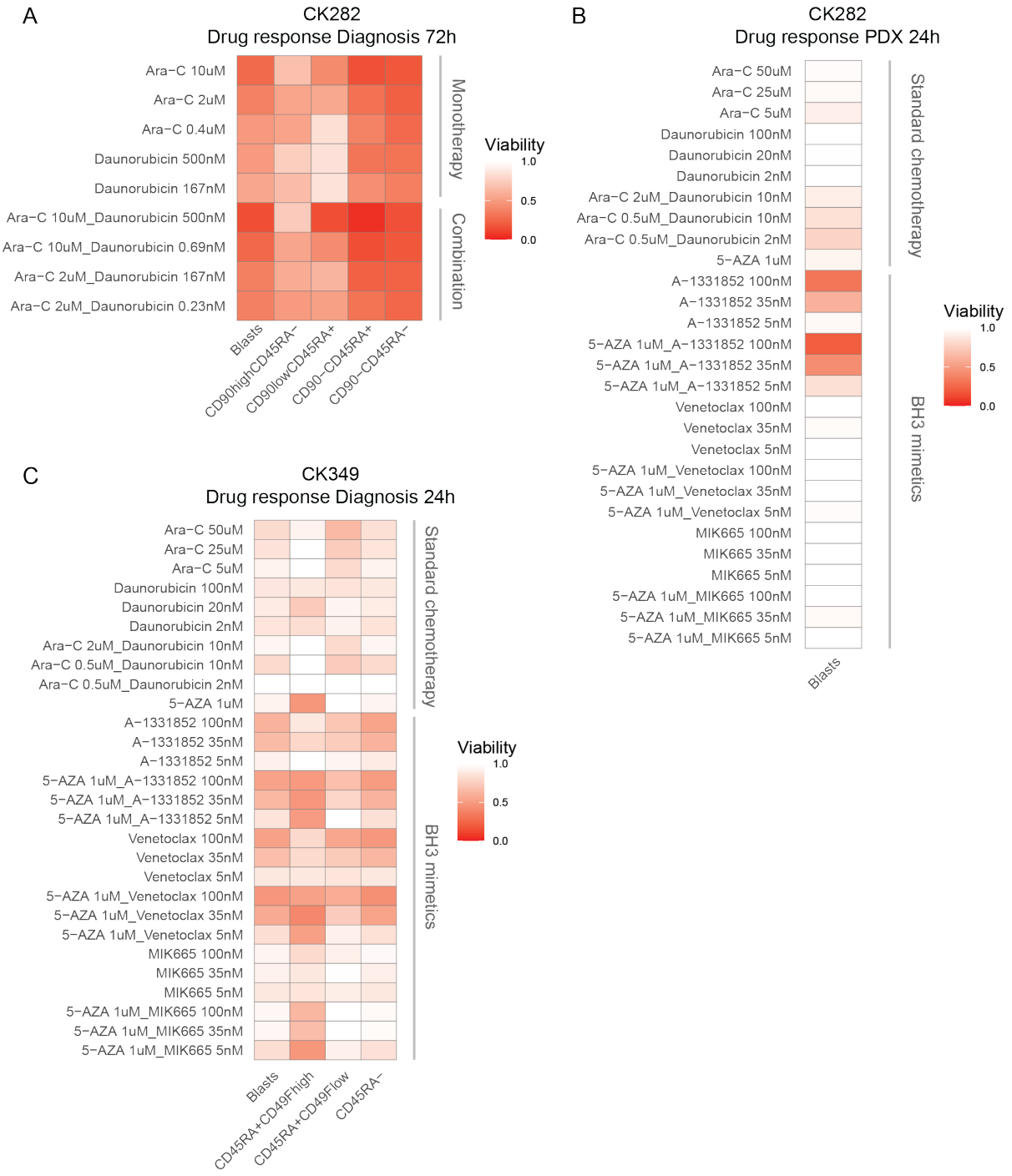
Appendix Figure 9 Consistent engraftment of CK-AML cells. A Circos plots of three patient-derived xenografts (PDXs) from CK282 showing similarities in the complex rearrangements and translocations involving multiple chromosomes between the PDXs, as assessed by optical genome mapping. Chromosomes (outside of the circular plot) and chromosomal rearrangements are shown as arcs connecting the two relevant genomic regions in the middle. ID of the PDX is shown on top of the circos plot. **B** Copy number (CN) variation tracks from two PDXs from CK349 showing similarities in the CN profiles between the PDXs as assessed by EPIC methylation arrays. ID of the PDX is shown on top of the genome tracks. Abbreviations: PDX, Patient-derived xenograft.



Appendix Figure 10 Bulk RNA-seq of diagnosis and engrafted cells. A Heatmap showing the expression of the engraftment-driving genes in cells at diagnosis and in patient-derived xenograft using bulk RNA-seq data in **A** CK282 and **B** CK349. Genes shown are the differentially expressed genes identified from the engraftment-driving using CITE-seq. Abbreviations: PDX, Patient-derived xenograft.



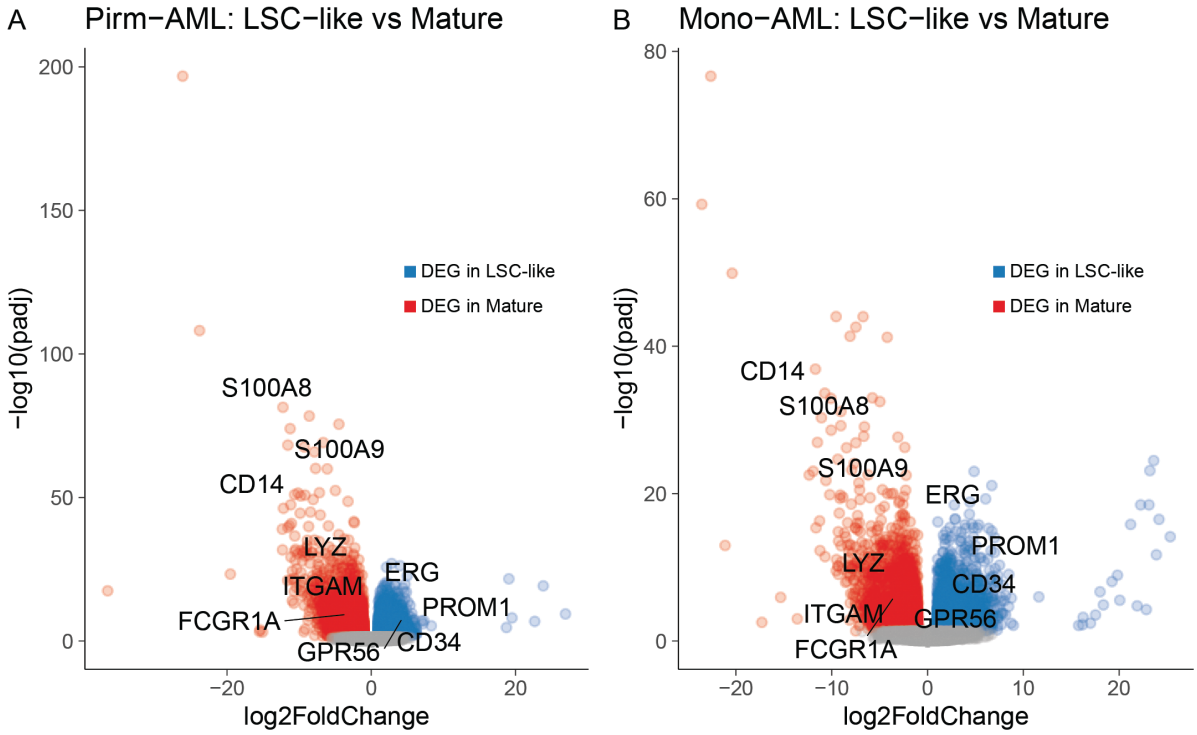
Appendix Figure 11 Expression of BCL-2 family members. A Fluorescence intensity of BCL-xL (left), BCL-2 (middle) and MCL-1 (right) protein expression in CD90^{high}CD45RA^{-} cells (red) compared to all blasts (blue). Delta mean fluorescence intensity (MFI) shown at the top of the plots was calculated as the difference in MFI between the specific protein expression (colored histogram) and its IgG control (grey histogram) in the assessed population. Abbreviations: MFI, Mean fluorescence intensity.



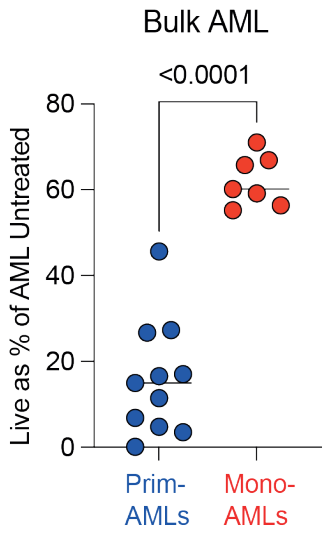
Appendix Figure 12 Single cell drug sensitivity profiling of CK-AML patient samples. A Viabilities (fraction of viable cells compared to untreated control) of different CK282 populations after 72h *ex vivo* exposure with indicated concentrations of standard chemotherapy regimens. **B** Viabilities of human blasts after 24h *ex vivo* exposure with indicated concentrations of standard chemotherapy regimens as well as BH3 mimetics in the patient-derived xenograft of CK282. **C** Viabilities of different CK349 populations after 24h *ex vivo* exposure with

indicated concentrations of standard chemotherapy regimens as well as BH3 mimetics. Abbreviations: PDX, Patient-derived xenograft.

Supplementary items related to Chapter 4



Appendix Figure 13 Differential expression between *LSC-like* and *Mature* cells. A-B Volcano plots showing differentially expressed genes (adjusted p-value < 0.05 and $|\log_2$ fold-change$| > 1$) between *LSC-like* (blue) and *Mature* (red) cells in **A** Prim-AML and **B** Mono-AML. Examples of differentially expressed gene are labelled in the plot. Abbreviations: DEG, Differentially Expressed Gene.



Appendix Figure 14 *Ex vivo* drug response in un-fractioned AML cells to venetoclax and azacytidine. Relative viability of unfractioned bulk leukemic cells from Prim-AML (n = 11) or Mono-AML (n = 7) patient samples after 24h venetoclax and azacytidine treatment *ex vivo*. Viabilities were normalized to untreated controls. Wilcoxon test was used to compare the groups.

References

1. Kantarjian H, Kadia T, DiNardo C, et al. Acute myeloid leukemia: current progress and future directions. Blood cancer journal 2021;11:41.
2. Dohner H, Weisdorf DJ, Bloomfield CD. Acute Myeloid Leukemia. N Engl J Med 2015;373:1136-52.
3. Löwenberg B, Downing JR, Burnett A. Acute myeloid leukemia. N Engl J Med 1999;341:1051-62.
4. Corces-Zimmerman MR, Hong WJ, Weissman IL, Medeiros BC, Majeti R. Preleukemic mutations in human acute myeloid leukemia affect epigenetic regulators and persist in remission. Proc Natl Acad Sci U S A 2014;111:2548-53.
5. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2019. CA: a cancer journal for clinicians 2019;69:7-34.
6. Kampen KR. The discovery and early understanding of leukemia. Leukemia research 2012;36:6-13.
7. Velpeau A. Sur la résorption du pus et sur l'altération du sang dans les maladies. Rev Med 1827;2:216-8.
8. Bennett J. Leucocythemia (1845), clinical lectures on the principles and practice of medicine. Last Edinburgh edn.,(S. Samuel & W. Wood, New York, 1860); 1860.
9. Neumann E. Über die Bedeutung des Knochenmarks für die Blutbildung. Zentralblatt für die medizinischen Wissenschaften 1868;44:122.
10. Ehrlich P. Beitrage zur Kenntnis der granulierten. Bindegewebszellen und der eosinophilen Leukozyten. Arch Anat Physiol 1879;3:166-9.
11. Liesveld JL, Lichtman MA. Acute Myelogenous Leukemia. In: Kaushansky K, Lichtman MA, Prchal JT, et al., eds. Williams Hematology, 9e. New York, NY: McGraw-Hill Education; 2015.
12. Kay AB. Paul Ehrlich and the Early History of Granulocytes. Microbiology spectrum 2016;4.
13. Piller G. Leukaemia - a brief historical review from ancient times to 1950. Br J Haematol 2001;112:282-92.
14. Ebstein W. Ueber die acute Leukämie und Pseudoleukämie. Deut Arch Klin Med 1889;44:343.
15. Naegeli O. Ueber rothes knochenmark und myeloblasten. DMW-Deutsche Medizinische Wochenschrift 1900;26:287-90.
16. Khoury JD, Solary E, Abla O, et al. The 5th edition of the World Health Organization Classification of Haematolymphoid Tumours: Myeloid and Histiocytic/Dendritic Neoplasms. Leukemia 2022;36:1703-19.
17. Deschler B, Lübbert M. Acute myeloid leukemia: epidemiology and etiology. Cancer 2006;107:2099-107.
18. Shallis RM, Wang R, Davidoff A, Ma X, Zeidan AM. Epidemiology of acute myeloid leukemia: Recent progress and enduring challenges. Blood reviews 2019;36:70-87.
19. Jaiswal S, Fontanillas P, Flannick J, et al. Age-related clonal hematopoiesis associated with adverse outcomes. N Engl J Med 2014;371:2488-98.
20. Tyner JW, Tognon CE, Bottomly D, et al. Functional genomic landscape of acute myeloid leukaemia. Nature 2018;562:526-31.
21. Ley TJ, Miller C, Ding L, et al. Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. N Engl J Med 2013;368:2059-74.
22. Morita K, Wang F, Jahn K, et al. Clonal evolution of acute myeloid leukemia revealed by high-throughput single-cell genomics. Nat Commun 2020;11:5327.

23. Döhner H, Estey E, Grimwade D, et al. Diagnosis and management of AML in adults: 2017 ELN recommendations from an international expert panel. Blood 2017;129:424-47.
24. Döhner H, Wei AH, Appelbaum FR, et al. Diagnosis and management of AML in adults: 2022 recommendations from an international expert panel on behalf of the ELN. Blood, The Journal of the American Society of Hematology 2022;140:1345-77.
25. Papaemmanuil E, Gerstung M, Bullinger L, et al. Genomic Classification and Prognosis in Acute Myeloid Leukemia. N Engl J Med 2016;374:2209-21.
26. Middeke JM, Herold S, Rücker-Braun E, et al. TP53 mutation in patients with high-risk acute myeloid leukaemia treated with allogeneic haematopoietic stem cell transplantation. Br J Haematol 2016;172:914-22.
27. Short NJ, Montalban-Bravo G, Hwang H, et al. Prognostic and therapeutic impacts of mutant TP53 variant allelic frequency in newly diagnosed acute myeloid leukemia. Blood Adv 2020;4:5681-9.
28. Mrozek K. Cytogenetic, molecular genetic, and clinical characteristics of acute myeloid leukemia with a complex karyotype. Semin Oncol 2008;35:365-77.
29. Grimwade D, Walker H, Oliver F, et al. The importance of diagnostic cytogenetics on outcome in AML: analysis of 1,612 patients entered into the MRC AML 10 trial. The Medical Research Council Adult and Children's Leukaemia Working Parties. Blood 1998;92:2322-33.
30. Rausch T, Jones DT, Zapatka M, et al. Genome sequencing of pediatric medulloblastoma links catastrophic DNA rearrangements with TP53 mutations. Cell 2012;148:59-71.
31. Bochtler T, Stolzel F, Heilig CE, et al. Clonal heterogeneity as detected by metaphase karyotyping is an indicator of poor prognosis in acute myeloid leukemia. J Clin Oncol 2013;31:3898-905.
32. Mrozek K, Eisfeld AK, Kohlschmidt J, et al. Complex karyotype in de novo acute myeloid leukemia: typical and atypical subtypes differ molecularly and clinically. Leukemia 2019;33:1620-34.
33. Grimwade D, Walker H, Harrison G, et al. The predictive value of hierarchical cytogenetic classification in older adults with acute myeloid leukemia (AML): analysis of 1065 patients entered into the United Kingdom Medical Research Council AML11 trial. Blood 2001;98:1312-20.
34. Schoch C, Kern W, Kohlmann A, Hiddemann W, Schnittger S, Haferlach T. Acute myeloid leukemia with a complex aberrant karyotype is a distinct biological entity characterized by genomic imbalances and a specific gene expression profile. Genes Chromosomes Cancer 2005;43:227-38.
35. Schoch C, Haferlach T, Bursch S, et al. Loss of genetic material is more common than gain in acute myeloid leukemia with complex aberrant karyotype: A detailed analysis of 125 cases using conventional chromosome analysis and fluorescence in situ hybridization including 24-color FISH. Genes, Chromosomes and Cancer 2002;35:20-9.
36. Till JE, McCulloch EA. Hemopoietic stem cell differentiation. Biochimica et biophysica acta 1980;605:431-59.
37. Notta F, Zandi S, Takayama N, et al. Distinct routes of lineage development reshape the human blood hierarchy across ontogeny. Science 2016;351:aab2116.
38. Bonnet D, Dick JE. Human acute myeloid leukemia is organized as a hierarchy that originates from a primitive hematopoietic cell. Nature Medicine 1997;3:730-7.
39. Lapidot T, Sirard C, Vormoor J, et al. A cell initiating human acute myeloid leukaemia after transplantation into SCID mice. Nature 1994;367:645-8.
40. Clarke MF, Dick JE, Dirks PB, et al. Cancer stem cells--perspectives on current status and future directions: AACR Workshop on cancer stem cells. Cancer Res 2006;66:9339-44.

41. Guan Y, Gerhard B, Hogge DE. Detection, isolation, and stimulation of quiescent primitive leukemic progenitor cells from patients with acute myeloid leukemia (AML). Blood 2003;101:3142-9.
42. Hope KJ, Jin L, Dick JE. Acute myeloid leukemia originates from a hierarchy of leukemic stem cell classes that differ in self-renewal capacity. Nat Immunol 2004;5:738-43.
43. Ishikawa F, Yoshida S, Saito Y, et al. Chemotherapy-resistant human AML stem cells home to and engraft within the bone-marrow endosteal region. Nat Biotechnol 2007;25:1315-21.
44. Saito Y, Uchida N, Tanaka S, et al. Induction of cell cycle entry eliminates human leukemia stem cells in a mouse model of AML. Nat Biotechnol 2010;28:275-80.
45. Ng SW, Mitchell A, Kennedy JA, et al. A 17-gene stemness score for rapid determination of risk in acute leukaemia. Nature 2016;540:433-7.
46. Eppert K, Takenaka K, Lechman ER, et al. Stem cell gene expression programs influence clinical outcome in human leukemia. Nat Med 2011;17:1086-93.
47. Gupta PB, Pastushenko I, Skibinski A, Blanpain C, Kuperwasser C. Phenotypic Plasticity: Driver of Cancer Initiation, Progression, and Therapy Resistance. Cell Stem Cell 2019;24:65-78.
48. Jones CL, Stevens BM, D'Alessandro A, et al. Inhibition of Amino Acid Metabolism Selectively Targets Human Leukemia Stem Cells. Cancer Cell 2018;34:724-40.e4.
49. Stevens BM, Jones CL, Pollyea DA, et al. Fatty acid metabolism underlies venetoclax resistance in acute myeloid leukemia stem cells. Nat Cancer 2020;1:1176-87.
50. Raffel S, Falcone M, Kneisel N, et al. BCAT1 restricts alphaKG levels in AML stem cells leading to IDHmut-like DNA hypermethylation. Nature 2017;551:384-8.
51. Duy C, Li M, Teater M, et al. Chemotherapy Induces Senescence-Like Resilient Cells Capable of Initiating AML Recurrence. Cancer Discov 2021;11:1542-61.
52. Li S, Garrett-Bakelman F, Perl AE, et al. Dynamic evolution of clonal epialleles revealed by methclone. Genome Biol 2014;15:472.
53. Thomas D, Majeti R. Biology and relevance of human acute myeloid leukemia stem cells. Blood 2017;129:1577-85.
54. Trumpp A, Haas S. Cancer stem cells: The adventurous journey from hematopoietic to leukemic stem cells. Cell 2022;185:1266-70.
55. Dohner H, Wei AH, Lowenberg B. Towards precision medicine for AML. Nat Rev Clin Oncol 2021;18:577-90.
56. Pollyea DA, Tallman MS, de Botton S, et al. Enasidenib, an inhibitor of mutant IDH2 proteins, induces durable remissions in older patients with newly diagnosed acute myeloid leukemia. Leukemia 2019;33:2575-84.
57. Roboz GJ, DiNardo CD, Stein EM, et al. Ivosidenib induces deep durable remissions in patients with newly diagnosed IDH1-mutant acute myeloid leukemia. Blood 2020;135:463-71.
58. Diepstraten ST, Anderson MA, Czabotar PE, Lessene G, Strasser A, Kelly GL. The manipulation of apoptosis for cancer therapy using BH3-mimetic drugs. Nat Rev Cancer 2022;22:45-64.
59. DiNardo CD, Jonas BA, Pullarkat V, et al. Azacitidine and Venetoclax in Previously Untreated Acute Myeloid Leukemia. N Engl J Med 2020;383:617-29.
60. Maiti A, Carter BZ, Andreeff M, Konopleva MY. Beyond BCL-2 Inhibition in Acute Myloid Leukemia: Other Approaches to Leverage the Apoptotic Pathway. Clinical lymphoma, myeloma & leukemia 2022;22:652-8.
61. Lindahl T, Barnes DE. Repair of endogenous DNA damage. Cold Spring Harbor symposia on quantitative biology 2000;65:127-33.
62. Jackson SP, Bartek J. The DNA-damage response in human biology and disease. Nature 2009;461:1071-8.

63. Cosenza MR, Rodriguez-Martin B, Korbel JO. Structural Variation in Cancer: Role, Prevalence, and Mechanisms. Annual review of genomics and human genetics 2022;23:123-52.
64. Consortium ITP-CAoWG. Pan-cancer analysis of whole genomes. Nature 2020;578:82-93.
65. Li Y, Roberts ND, Wala JA, et al. Patterns of somatic structural variation in human cancer genomes. Nature 2020;578:112-21.
66. Rahman N. Realizing the promise of cancer predisposition genes. Nature 2014;505:302-8.
67. Nik-Zainal S, Davies H, Staaf J, et al. Landscape of somatic mutations in 560 breast cancer whole-genome sequences. Nature 2016;534:47-54.
68. Behjati S, Gundem G, Wedge DC, et al. Mutational signatures of ionizing radiation in second malignancies. Nat Commun 2016;7:12605.
69. Liehr T, Claussen U. Current developments in human molecular cytogenetic techniques. Current molecular medicine 2002;2:283-97.
70. Korbel JO, Urban AE, Affourtit JP, et al. Paired-end mapping reveals extensive structural variation in the human genome. Science 2007;318:420-6.
71. Rausch T, Zichner T, Schlattl A, Stütz AM, Benes V, Korbel JO. DELLY: structural variant discovery by integrated paired-end and split-read analysis. Bioinformatics 2012;28:i333-i9.
72. Chiang DY, Getz G, Jaffe DB, et al. High-resolution mapping of copy-number alterations with massively parallel sequencing. Nat Methods 2009;6:99-103.
73. Method of the Year 2022: long-read sequencing. Nat Methods 2023;20:1.
74. Beyter D, Ingimundardottir H, Oddsson A, et al. Long-read sequencing of 3,622 Icelanders provides insight into the role of structural variants in human diseases and other traits. Nat Genet 2021;53:779-86.
75. Ebert P, Audano PA, Zhu Q, et al. Haplotype-resolved diverse human genomes and integrated analysis of structural variation. Science 2021;372.
76. Nurk S, Koren S, Rhie A, et al. The complete sequence of a human genome. Science 2022;376:44-53.
77. Schwartz DC, Li X, Hernandez LI, Ramnarain SP, Huff EJ, Wang YK. Ordered restriction maps of Saccharomyces cerevisiae chromosomes constructed by optical mapping. Science 1993;262:110-4.
78. Yang H, Garcia-Manero G, Sasaki K, et al. High-resolution structural variant profiling of myelodysplastic syndromes by optical genome mapping uncovers cryptic aberrations of prognostic and therapeutic significance. Leukemia 2022;36:2306-16.
79. Luebeck J, Coruh C, Dehkordi SR, et al. AmpliconReconstructor integrates NGS and optical mapping to resolve the complex structures of focal amplifications. Nat Commun 2020;11:4374.
80. Hadi K, Yao X, Behr JM, et al. Distinct Classes of Complex Structural Variation Uncovered across Thousands of Cancer Genome Graphs. Cell 2020;183:197-210.e32.
81. Korbel JO, Campbell PJ. Criteria for inference of chromothripsis in cancer genomes. Cell 2013;152:1226-36.
82. Stephens PJ, Greenman CD, Fu B, et al. Massive genomic rearrangement acquired in a single catastrophic event during cancer development. Cell 2011;144:27-40.
83. Maciejowski J, de Lange T. Telomeres in cancer: tumour suppression and genome instability. Nature reviews Molecular cell biology 2017;18:175-86.
84. Holland AJ, Cleveland DW. Chromoanagenesis and cancer: mechanisms and consequences of localized, complex chromosomal rearrangements. Nat Med 2012;18:1630-8.
85. Rosswog C, Bartenhagen C, Welte A, et al. Chromothripsis followed by circular recombination drives oncogene amplification in human cancer. Nat Genet 2021;53:1673-85.

86. Li Y, Schwab C, Ryan S, et al. Constitutional and somatic rearrangement of chromosome 21 in acute lymphoblastic leukaemia. Nature 2014;508:98-102.
87. Koche RP, Rodriguez-Fos E, Helmsauer K, et al. Extrachromosomal circular DNA drives oncogenic genome remodeling in neuroblastoma. Nat Genet 2020;52:29-34.
88. Gawad C, Koh W, Quake SR. Single-cell genome sequencing: current state of the science. Nat Rev Genet 2016;17:175-88.
89. Navin N, Kendall J, Troge J, et al. Tumour evolution inferred by single-cell sequencing. Nature 2011;472:90-4.
90. Sanders AD, Falconer E, Hills M, Spierings DCJ, Lansdorp PM. Single-cell template strand sequencing by Strand-seq enables the characterization of individual homologs. Nat Protoc 2017;12:1151-76.
91. Sanders AD, Hills M, Porubský D, Guryev V, Falconer E, Lansdorp PM. Characterizing polymorphic inversions in human genomes by single-cell sequencing. Genome Res 2016;26:1575-87.
92. Falconer E, Hills M, Naumann U, et al. DNA template strand sequencing of single-cells maps genomic rearrangements at high resolution. Nat Methods 2012;9:1107-12.
93. Sanders AD, Meiers S, Ghareghani M, et al. Single-cell analysis of structural variations and complex rearrangements with tri-channel processing. Nat Biotechnol 2020;38:343-54.
94. Chaisson MJP, Sanders AD, Zhao X, et al. Multi-platform discovery of haplotype-resolved structural variation in human genomes. Nat Commun 2019;10:1784.
95. van Wietmarschen N, Lansdorp PM. Bromodeoxyuridine does not contribute to sister chromatid exchange events in normal or Bloom syndrome cells. Nucleic Acids Res 2016;44:6787-93.
96. Porubsky D, Sanders AD, van Wietmarschen N, et al. Direct chromosome-length haplotyping by single-cell sequencing. Genome Res 2016;26:1565-74.
97. Porubsky D, Garg S, Sanders AD, et al. Dense and accurate whole-chromosome haplotyping of individual genomes. Nat Commun 2017;8:1293.
98. Jeong H, Grimes K, Rauwolf KK, et al. Functional analysis of structural variants in single cells using Strand-seq. Nat Biotechnol 2022.
99. Schones DE, Cui K, Cuddapah S, et al. Dynamic regulation of nucleosome positioning in the human genome. Cell 2008;132:887-98.
100. Teif VB, Vainshtein Y, Caudron-Herger M, et al. Genome-wide nucleosome positioning during embryonic stem cell development. Nat Struct Mol Biol 2012;19:1185-92.
101. Struhl K, Segal E. Determinants of nucleosome positioning. Nat Struct Mol Biol 2013;20:267-73.
102. Lai B, Gao W, Cui K, et al. Principles of nucleosome organization revealed by single-cell micrococcal nuclease sequencing. Nature 2018;562:281-5.
103. Klein AM, Mazutis L, Akartuna I, et al. Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. Cell 2015;161:1187-201.
104. Macosko EZ, Basu A, Satija R, et al. Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. Cell 2015;161:1202-14.
105. Stoeckius M, Hafemeister C, Stephenson W, et al. Simultaneous epitope and transcriptome measurement in single cells. Nat Methods 2017;14:865-8.
106. Zhu C, Preissl S, Ren B. Single-cell multimodal omics: the power of many. Nat Methods 2020;17:11-4.
107. Hao Y, Hao S, Andersen-Nissen E, et al. Integrated analysis of multimodal single-cell data. Cell 2021;184:3573-87 e29.
108. Granja JM, Klemm S, McGinnis LM, et al. Single-cell multiomic analysis identifies regulatory programs in mixed-phenotype acute leukemia. Nat Biotechnol 2019;37:1458-65.
109. Gorin G, Svensson V, Pachter L. Protein velocity and acceleration from single-cell multiomics experiments. Genome Biol 2020;21:39.

110. Frangieh CJ, Melms JC, Thakore PI, et al. Multimodal pooled Perturb-CITE-seq screens in patient models define mechanisms of cancer immune evasion. Nat Genet 2021;53:332-41.
111. Mimitou EP, Cheng A, Montalbano A, et al. Multiplexed detection of proteins, transcriptomes, clonotypes and CRISPR perturbations in single cells. Nat Methods 2019;16:409-12.
112. Jacobson MD, Weil M, Raff MC. Programmed cell death in animal development. Cell 1997;88:347-54.
113. Singh R, Letai A, Sarosiek K. Regulation of apoptosis in health and disease: the balancing act of BCL-2 family proteins. Nature reviews Molecular cell biology 2019;20:175-93.
114. Ashkenazi A, Dixit VM. Death receptors: signaling and modulation. Science 1998;281:1305-8.
115. Galluzzi L, Vitale I, Aaronson SA, et al. Molecular mechanisms of cell death: recommendations of the Nomenclature Committee on Cell Death 2018. Cell death and differentiation 2018;25:486-541.
116. Smith CA, Farrah T, Goodwin RG. The TNF receptor superfamily of cellular and viral proteins: activation, costimulation, and death. Cell 1994;76:959-62.
117. Kischkel FC, Hellbardt S, Behrmann I, et al. Cytotoxicity-dependent APO-1 (Fas/CD95)-associated proteins form a death-inducing signaling complex (DISC) with the receptor. Embo j 1995;14:5579-88.
118. Tait SW, Green DR. Mitochondria and cell death: outer membrane permeabilization and beyond. Nature reviews Molecular cell biology 2010;11:621-32.
119. Chipuk JE, Bouchier-Hayes L, Green DR. Mitochondrial outer membrane permeabilization during apoptosis: the innocent bystander scenario. Cell death and differentiation 2006;13:1396-402.
120. Mihara M, Erster S, Zaika A, et al. p53 has a direct apoptogenic role at the mitochondria. Mol Cell 2003;11:577-90.
121. Chipuk JE, Kuwana T, Bouchier-Hayes L, et al. Direct activation of Bax by p53 mediates mitochondrial membrane permeabilization and apoptosis. Science 2004;303:1010-4.
122. Soderquist RS, Crawford L, Liu E, et al. Systematic mapping of BCL-2 gene dependencies in cancer reveals molecular determinants of BH3 mimetic sensitivity. Nat Commun 2018;9:3513.
123. Kaufmann SH, Karp JE, Svingen PA, et al. Elevated expression of the apoptotic regulator Mcl-1 at the time of leukemic relapse. Blood 1998;91:991-1000.
124. Wuillème-Toumi S, Robillard N, Gomez P, et al. Mcl-1 is overexpressed in multiple myeloma and associated with relapse and shorter survival. Leukemia 2005;19:1248-52.
125. Bailey C, Black JRM, Reading JL, et al. Tracking Cancer Evolution through the Disease Course. Cancer Discovery 2021;11:916.
126. Turajlic S, Sottoriva A, Graham T, Swanton C. Resolving genetic heterogeneity in cancer. Nat Rev Genet 2019;20:404-16.
127. Baca SC, Prandi D, Lawrence MS, et al. Punctuated evolution of prostate cancer genomes. Cell 2013;153:666-77.
128. Field MG, Durante MA, Anbunathan H, et al. Punctuated evolution of canonical genomic aberrations in uveal melanoma. Nat Commun 2018;9:116.
129. Gerlinger M, McGranahan N, Dewhurst SM, Burrell RA, Tomlinson I, Swanton C. Cancer: evolution within a lifetime. Annu Rev Genet 2014;48:215-36.
130. Markowetz F. A saltationist theory of cancer evolution. Nat Genet 2016;48:1102-3.
131. Vendramin R, Litchfield K, Swanton C. Cancer evolution: Darwin and beyond. EMBO J 2021;40:e108389.
132. Shorokhova M, Nikolsky N, Grinchuk T. Chromothripsis-Explosion in Genetic Science. Cells 2021;10.

133. Kolb T, Ernst A. Cell-based model systems for genome instability: Dissecting the mechanistic basis of chromothripsis in cancer. International journal of cancer 2021.
134. Rücker FG, Schlenk RF, Bullinger L, et al. TP53 alterations in acute myeloid leukemia with complex karyotype correlate with specific copy number alterations, monosomal karyotype, and dismal outcome. Blood 2012;119:2114-21.
135. Zheng GX, Terry JM, Belgrader P, et al. Massively parallel digital transcriptional profiling of single cells. Nat Commun 2017;8:14049.
136. Korsunsky I, Millard N, Fan J, et al. Fast, sensitive and accurate integration of single-cell data with Harmony. Nat Methods 2019;16:1289-96.
137. Yang L, Rau R, Goodell MA. DNMT3A in haematological malignancies. Nat Rev Cancer 2015;15:152-65.
138. Paik J-H, Kollipara R, Chu G, et al. FoxOs Are Lineage-Restricted Redundant Tumor Suppressors and Regulate Endothelial Cell Homeostasis. Cell 2007;128:309-23.
139. Micci F, Thorsen J, Haugom L, Zeller B, Tierens A, Heim S. Translocation t(1;16)(p31;q24) rearranging CBFA2T3 is specific for acute erythroid leukemia. Leukemia 2011;25:1510-2.
140. Ottema S, Mulet-Lazaro R, Beverloo HB, et al. Atypical 3q26/MECOM rearrangements genocopy inv(3)/t(3;3) in acute myeloid leukemia. Blood 2020;136:224-34.
141. Groschel S, Sanders MA, Hoogenboezem R, et al. A single oncogenic enhancer rearrangement causes concomitant EVI1 and GATA2 deregulation in leukemia. Cell 2014;157:369-81.
142. Yamazaki H, Suzuki M, Otsuki A, et al. A remote GATA2 hematopoietic enhancer drives leukemogenesis in inv(3)(q21;q26) by activating EVI1 expression. Cancer Cell 2014;25:415-27.
143. Fonatsch C, Gudat H, Lengfelder E, et al. Correlation of cytogenetic findings with clinical features in 18 patients with inv (3)(q21q26) or t (3; 3)(q21; q26). Leukemia 1994;8:1318-26.
144. Yamaoka A, Suzuki M, Katayama S, Orihara D, Engel JD, Yamamoto M. EVI1 and GATA2 misexpression induced by inv(3)(q21q26) contribute to megakaryocyte-lineage skewing and leukemogenesis. Blood Adv 2020;4:1722-36.
145. Grigg A, Gascoyne R, Phillips G, Horsman D. Clinical, haematological and cytogenetic features in 24 patients with structural rearrangements of the Q arm of chromosome 3. British journal of haematology 1993;83:158-65.
146. Liberzon A, Birger C, Thorvaldsdottir H, Ghandi M, Mesirov JP, Tamayo P. The Molecular Signatures Database (MSigDB) hallmark gene set collection. Cell Syst 2015;1:417-25.
147. Liberzon A, Subramanian A, Pinchback R, Thorvaldsdottir H, Tamayo P, Mesirov JP. Molecular signatures database (MSigDB) 3.0. Bioinformatics 2011;27:1739-40.
148. Muller S, Cho A, Liu SJ, Lim DA, Diaz A. CONICS integrates scRNA-seq with DNA sequencing to map gene expression to tumor sub-clones. Bioinformatics 2018;34:3217-9.
149. Kim Y, Lin Q, Zelterman D, Yun Z. Hypoxia-regulated delta-like 1 homologue enhances cancer cell stemness and tumorigenicity. Cancer Res 2009;69:9271-80.
150. Lin CC, Hsu YC, Li YH, et al. Higher HOPX expression is associated with distinct clinical and biological features and predicts poor prognosis in de novo acute myeloid leukemia. Haematologica 2017;102:1044-53.
151. Liu J, Li W, Wu L. Pan-cancer analysis suggests histocompatibility minor 13 is an unfavorable prognostic biomarker promoting cell proliferation, migration, and invasion in hepatocellular carcinoma. Frontiers in pharmacology 2022;13:950156.
152. Zhou JK, Fan X, Cheng J, Liu W, Peng Y. PDLIM1: Structure, function and implication in cancer. Cell stress 2021;5:119-27.

153. Grinat J, Heuberger J, Vidal RO, et al. The epigenetic regulator Mll1 is required for Wnt-driven intestinal tumorigenesis and cancer stemness. Nat Commun 2020;11:6422.
154. Raspadori D, Damiani D, Lenoci M, et al. CD56 antigenic expression in acute myeloid leukemia identifies patients with poor clinical prognosis. Leukemia 2001;15:1161-4.
155. Sasca D, Szybinski J, Schüler A, et al. NCAM1 (CD56) promotes leukemogenesis and confers drug resistance in AML. Blood 2019;133:2305-19.
156. Herrmann H, Sadovnik I, Cerny-Reiterer S, et al. Dipeptidylpeptidase IV (CD26) defines leukemic stem cells (LSC) in chronic myeloid leukemia. Blood 2014;123:3951-62.
157. Warfvinge R, Geironson L, Sommarin MNE, et al. Single-cell molecular analysis defines therapy response and immunophenotype of stem cell subpopulations in CML. Blood 2017;129:2384-94.
158. Notta F, Doulatov S, Laurenti E, Poeppl A, Jurisica I, Dick JE. Isolation of single human hematopoietic stem cells capable of long-term multilineage engraftment. Science 2011;333:218-21.
159. Li J, Ran Q, Xu B, et al. Role of CD25 expression on prognosis of acute myeloid leukemia: A literature review and meta-analysis. PloS one 2020;15:e0236124.
160. Saxton RA, Sabatini DM. mTOR Signaling in Growth, Metabolism, and Disease. Cell 2017;168:960-76.
161. Pereira O, Teixeira A, Sampaio-Marques B, Castro I, Girão H, Ludovico P. Signalling mechanisms that regulate metabolic profile and autophagy of acute myeloid leukaemia cells. Journal of cellular and molecular medicine 2018;22:4807-17.
162. Lu J, Cannizzaro E, Meier-Abt F, et al. Multi-omics reveals clinically relevant proliferative drive associated with mTOR-MYC-OXPHOS activity in chronic lymphocytic leukemia. Nat Cancer 2021;2:853-64.
163. Aibar S, Gonzalez-Blas CB, Moerman T, et al. SCENIC: single-cell regulatory network inference and clustering. Nat Methods 2017;14:1083-6.
164. Mertens F, Johansson B, Mitelman F. Isochromosomes in neoplasia. Genes Chromosomes Cancer 1994;10:221-30.
165. Wong KF, Kwong YL. Isochromosome 8q is a Marker of Secondary Acute Myeloid Leukemia. Cancer Genetics and Cytogenetics 2000;120:171-3.
166. Palin K, Pitkänen E, Turunen M, et al. Contribution of allelic imbalance to colorectal cancer. Nat Commun 2018;9:3664.
167. Tuna M, Amos CI, Mills GB. Whole-chromosome arm acquired uniparental disomy in cancer development is a consequence of isochromosome formation. Neoplasia (New York, NY) 2022;25:9-17.
168. Buccisano F, Rossi FM, Venditti A, et al. CD90/Thy-1 is preferentially expressed on blast cells of high risk acute myeloid leukaemias. Br J Haematol 2004;125:203-12.
169. Nilsson L, Åstrand-Grundström I, Anderson K, et al. Involvement and functional impairment of the CD34+ CD38− Thy-1+ hematopoietic stem cell pool in myelodysplastic syndromes with trisomy 8. Blood, The Journal of the American Society of Hematology 2002;100:259-67.
170. Rhodes DR, Yu J, Shanker K, et al. Large-scale meta-analysis of cancer microarray data identifies common transcriptional profiles of neoplastic transformation and progression. Proc Natl Acad Sci U S A 2004;101:9309-14.
171. Deb S, Xu H, Tuynman J, et al. RAD21 cohesin overexpression is a prognostic and predictive marker exacerbating poor prognosis in KRAS mutant colorectal carcinomas. British journal of cancer 2014;110:1606-13.
172. Ginestier C, Hur MH, Charafe-Jauffret E, et al. ALDH1 is a marker of normal and malignant human mammary stem cells and a predictor of poor clinical outcome. Cell Stem Cell 2007;1:555-67.

173. Ran D, Schubert M, Pietsch L, et al. Aldehyde dehydrogenase activity among primary leukemia cells is associated with stem cell features and correlates with adverse clinical outcomes. Experimental hematology 2009;37:1423-34.
174. Blume R, Rempel E, Manta L, et al. The molecular signature of AML with increased ALDH activity suggests a stem cell origin. Leuk Lymphoma 2018;59:2201-10.
175. Rasmussen ML, Kline LA, Park KP, et al. A Non-apoptotic Function of MCL-1 in Promoting Pluripotency and Modulating Mitochondrial Dynamics in Stem Cells. Stem Cell Reports 2018;10:684-92.
176. Veas-Perez de Tudela M, Delgado-Esteban M, Maestre C, et al. Regulation of Bcl-xL-ATP Synthase Interaction by Mitochondrial Cyclin B1-Cyclin-Dependent Kinase-1 Determines Neuronal Survival. J Neurosci 2015;35:9287-301.
177. Lucas CM, Milani M, Butterworth M, et al. High CIP2A levels correlate with an antiapoptotic phenotype that can be overcome by targeting BCL-XL in chronic myeloid leukemia. Leukemia 2016;30:1273-81.
178. Pollyea DA, Stevens BM, Jones CL, et al. Venetoclax with azacitidine disrupts energy metabolism and targets leukemia stem cells in patients with acute myeloid leukemia. Nat Med 2018;24:1859-66.
179. Waclawiczek A, Leppä AM, Renders S, et al. Combinatorial BCL-2 family expression in Acute Myeloid Leukemia Stem Cells predicts clinical response to Azacitidine/Venetoclax. Cancer Discovery 2023;Provisionally accepted.
180. Negrini S, Gorgoulis VG, Halazonetis TD. Genomic instability--an evolving hallmark of cancer. Nature reviews Molecular cell biology 2010;11:220-8.
181. Ediriwickrema A, Gentles AJ, Majeti R. Single-cell genomics in AML: extending the frontiers of AML research. Blood 2023;141:345-55.
182. Klco JM, Spencer DH, Miller CA, et al. Functional heterogeneity of genetically defined subclones in acute myeloid leukemia. Cancer Cell 2014;25:379-92.
183. Swanton C. Intratumor heterogeneity: evolution through space and time. Cancer Res 2012;72:4875-82.
184. Greaves M. Evolutionary determinants of cancer. Cancer Discov 2015;5:806-20.
185. Fontana MC, Marconi G, Feenstra JDM, et al. Chromothripsis in acute myeloid leukemia: biological features and impact on survival. Leukemia 2018;32:1609-20.
186. Watkins TBK, Lim EL, Petkovic M, et al. Pervasive chromosomal instability and karyotype order in tumour evolution. Nature 2020;587:126-32.
187. Garsed DW, Marshall OJ, Corbin VD, et al. The architecture and evolution of cancer neochromosomes. Cancer Cell 2014;26:653-67.
188. Melo JV, Barnes DJ. Chronic myeloid leukaemia as a model of disease evolution in human cancer. Nat Rev Cancer 2007;7:441-53.
189. Singh S, Brocker C, Koppaka V, et al. Aldehyde dehydrogenases in cellular responses to oxidative/electrophilic stress. Free radical biology & medicine 2013;56:89-101.
190. Januchowski R, Wojtowicz K, Sterzyńska K, et al. Inhibition of ALDH1A1 activity decreases expression of drug transporters and reduces chemotherapy resistance in ovarian cancer cell lines. The international journal of biochemistry & cell biology 2016;78:248-59.
191. Petiti J, Lo Iacono M, Rosso V, et al. Bcl-xL represents a therapeutic target in Philadelphia negative myeloproliferative neoplasms. Journal of cellular and molecular medicine 2020;24:10978-86.
192. Waibel M, Solomon VS, Knight DA, et al. Combined targeting of JAK2 and Bcl-2/Bcl-xL to cure mutant JAK2-driven malignancies and overcome acquired resistance to JAK2 inhibitors. Cell Rep 2013;5:1047-59.
193. Webster P. A golden era of cancer clinical trials. Nat Med 2022;28:602-5.

194. Lagadinou ED, Sach A, Callahan K, et al. BCL-2 inhibition targets oxidative phosphorylation and selectively eradicates quiescent human leukemia stem cells. Cell Stem Cell 2013;12:329-41.
195. Pan R, Hogdal LJ, Benito JM, et al. Selective BCL-2 inhibition by ABT-199 causes on-target cell death in acute myeloid leukemia. Cancer Discov 2014;4:362-75.
196. Vo TT, Ryan J, Carrasco R, et al. Relative mitochondrial priming of myeloblasts and normal HSCs determines chemotherapeutic success in AML. Cell 2012;151:344-55.
197. Malone ER, Oliva M, Sabatini PJB, Stockley TL, Siu LL. Molecular profiling for precision cancer therapies. Genome medicine 2020;12:8.
198. Cherry EM, Abbott D, Amaya M, et al. Venetoclax and azacitidine compared with induction chemotherapy for newly diagnosed patients with acute myeloid leukemia. Blood Adv 2021;5:5565-73.
199. Kuusanmäki H, Leppä AM, Pölönen P, et al. Phenotype-based drug screening reveals association between venetoclax response and differentiation stage in acute myeloid leukemia. Haematologica 2020;105:708-20.
200. Pei S, Pollyea DA, Gustafson A, et al. Monocytic Subclones Confer Resistance to Venetoclax-Based Therapy in Patients with Acute Myeloid Leukemia. Cancer Discov 2020;10:536-51.
201. Bhatt S, Pioso MS, Olesinski EA, et al. Reduced Mitochondrial Apoptotic Priming Drives Resistance to BH3 Mimetics in Acute Myeloid Leukemia. Cancer Cell 2020;38:872-90.e6.
202. Cai SF, Chu SH, Goldberg AD, et al. Leukemia Cell of Origin Influences Apoptotic Priming and Sensitivity to LSD1 Inhibition. Cancer Discov 2020;10:1500-13.
203. Gorczyca W. Flow cytometry immunophenotypic characteristics of monocytic population in acute monocytic leukemia (AML-M5), acute myelomonocytic leukemia (AML-M4), and chronic myelomonocytic leukemia (CMML). Methods in cell biology 2004;75:665-77.
204. Stahl M, Menghrajani K, Derkach A, et al. Clinical and molecular predictors of response and survival following venetoclax therapy in relapsed/refractory AML. Blood Adv 2021;5:1552-64.
205. Donato E, Correia NC, Andresen C, et al. Retained functional normal and preleukemic HSCs at diagnosis are associated to good prognosis in DNMT3Amut NPM1mut AMLs. Blood Adv 2022.
206. Pabst C, Bergeron A, Lavallee VP, et al. GPR56 identifies primary human acute myeloid leukemia cells with high repopulating potential in vivo. Blood 2016;127:2018-27.
207. Florek M, Haase M, Marzesco AM, et al. Prominin-1/CD133, a neural and hematopoietic stem cell marker, is expressed in adult human differentiated cells and certain types of kidney cancer. Cell Tissue Res 2005;319:15-26.
208. Tume L, Paco K, Ubidia-Incio R, Moya J. CD133 in breast cancer cells and in breast cancer stem cells as another target for immunotherapy. Gaceta Mexicana de Oncología 2016;15:22-30.
209. Loughran SJ, Kruse EA, Hacking DF, et al. The transcription factor Erg is essential for definitive hematopoiesis and the function of adult hematopoietic stem cells. Nat Immunol 2008;9:810-9.
210. Klonisch T, Wiechec E, Hombach-Klonisch S, et al. Cancer stem cell markers in common cancers - therapeutic implications. Trends Mol Med 2008;14:450-60.
211. Lagasse E, Clerc RG. Cloning and expression of two human genes encoding calcium-binding proteins that are regulated during myeloid differentiation. Mol Cell Biol 1988;8:2402-10.

212. Koike T, Harada N, Yoshida T, Morikawa M. Regulation of myeloid-specific calcium binding protein synthesis by cytosolic protein kinase C. Journal of biochemistry 1992;112:624-30.
213. Ziegler-Heitbrock HW, Ulevitch RJ. CD14: cell surface receptor and differentiation marker. Immunol Today 1993;14:121-5.
214. Mou T, Pawitan Y, Stahl M, et al. The transcriptome-wide landscape of molecular subtype-specific mRNA expression profiles in acute myeloid leukemia. American journal of hematology 2021;96:580-8.
215. Konopleva M, Pollyea DA, Potluri J, et al. Efficacy and Biological Correlates of Response in a Phase II Study of Venetoclax Monotherapy in Patients with Acute Myelogenous Leukemia. Cancer Discov 2016;6:1106-17.
216. Jones CL, Stevens BM, D'Alessandro A, et al. Cysteine depletion targets leukemia stem cells through inhibition of electron transport complex II. Blood 2019;134:389-94.
217. Kreso A, Dick JE. Evolution of the cancer stem cell model. Cell Stem Cell 2014;14:275-91.
218. McKenzie MD, Ghisi M, Oxley EP, et al. Interconversion between Tumorigenic and Differentiated States in Acute Myeloid Leukemia. Cell Stem Cell 2019;25:258-72.e9.
219. Gomez-Bougie P, Maiga S, Tessoulin B, et al. BH3-mimetic toolkit guides the respective use of BCL2 and MCL1 BH3-mimetics in myeloma treatment. Blood 2018;132:2656-69.
220. Tessoulin B, Papin A, Gomez-Bougie P, et al. BCL2-Family Dysregulation in B-Cell Malignancies: From Gene Expression Regulation to a Targeted Therapy Biomarker. Frontiers in oncology 2018;8:645.
221. Chiron D, Dousset C, Brosseau C, et al. Biological rational for sequential targeting of Bruton tyrosine kinase and Bcl-2 to overcome CD40-induced ABT-199 resistance in mantle cell lymphoma. Oncotarget 2015;6:8750-9.
222. Kuusanmäki H, Kytölä S, Vänttinen I, et al. Ex vivo venetoclax sensitivity testing predicts treatment response in acute myeloid leukemia. Haematologica 2022.
223. Pabst C, Krosl J, Fares I, et al. Identification of small molecules that support human leukemia stem cell activity ex vivo. Nat Methods 2014;11:436-42.
224. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 2009;25:1754-60.
225. Tischler G, Leonard S. biobambam: tools for read pair collation based algorithms on BAM files. Source Code for Biology and Medicine 2014;9:1-18.
226. Kamburov A, Herwig R. ConsensusPathDB 2022: molecular interactions update as a resource for network biology. Nucleic Acids Res 2022;50:D587-d95.
227. Butler A, Hoffman P, Smibert P, Papalexi E, Satija R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. Nature biotechnology 2018;36:411-20.
228. Hafemeister C, Satija R. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. Genome Biol 2019;20:296.
229. Stuart T, Butler A, Hoffman P, et al. Comprehensive Integration of Single-Cell Data. Cell 2019;177:1888-902 e21.
230. Barshan E, Ghodsi A, Azimifar Z, Zolghadri Jahromi M. Supervised principal component analysis: Visualization, classification and regression on subspaces and submanifolds. Pattern Recognition 2011;44:1357-71.
231. Uhrig S, Ellermann J, Walther T, et al. Accurate and efficient detection of gene fusions from RNA sequencing data. Genome Res 2021;31:448-60.
232. Veeneman BA, Shukla S, Dhanasekaran SM, Chinnaiyan AM, Nesvizhskii AI. Two-pass alignment improves novel splice junction quantification. Bioinformatics 2016;32:43-9.

233. Aryee MJ, Jaffe AE, Corrada-Bravo H, et al. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. Bioinformatics 2014;30:1363-9.
234. Picelli S, Faridani OR, Björklund AK, Winberg G, Sagasser S, Sandberg R. Full-length RNA-seq from single cells using Smart-seq2. Nat Protoc 2014;9:171-81.
235. Dobin A, Davis CA, Schlesinger F, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics 2013;29:15-21.
236. Tarasov A, Vilella AJ, Cuppen E, Nijman IJ, Prins P. Sambamba: fast processing of NGS alignment formats. Bioinformatics 2015;31:2032-4.
237. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics 2014;30:923-30.
238. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 2014;15:550.