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
Linda Christina Schuster, M.Sc.
born in Linz, Austria
Oral examination: 17.10.2022

Dissecting heterogeneity, clonal evolution, and epigenetic changes in distinct and molecularly defined AML subsets by multi omics single-cell sequencing

Referees: Prof. Dr. Karsten Rippe
Dr. Sevin Turcan

This work was performed from September 2018 to July 2022 under the supervision of Prof. Dr. Karsten Rippe in the Division Chromatin Networks at the German Cancer Research Center (DKFZ) and the BioQuant Center in Heidelberg, Germany.

Declaration

I hereby declare that I have written the submitted dissertation “Dissecting heterogeneity, clonal evolution, and epigenetic changes in distinct and molecularly defined AML subsets by multi omics single-cell sequencing” myself and in this process, have used no other sources or materials than those explicitly indicated. I hereby declare that I have not applied to be examined at any other institution, nor have I used the dissertation in this or any other form at any other institution as an examination paper, nor submitted it to any other faculty as a dissertation.

_____ _____

(Place, Date) Linda Christina Schuster

Table of contents

Summary

In acute myeloid leukemia (AML), initiation of tumorigenesis via multiple oncogenic mutations occurs throughout various stages of hematopoiesis that disrupt the corresponding transcriptomic and epigenetic profiles. The cancer cells that emerge are referred to as blasts and share biological features from these disease-specific alterations and patterns associated with differentiation and the tumor cell of origin. The resulting blasts show a large inter- and intra-tumor heterogeneity within molecularly defined AML subgroups that are highly relevant for risk stratification and personalized treatment strategies. Sequencing methods that analyze the transcriptome (scRNA-seq) and epigenome (scATAC-seq) are ideally suited to resolve tumor cell heterogeneity as well as non-malignant cell types in the microenvironment. Additionally, scATAC-seq allows to map the binding of transcription factors (TF) and infer cell-specific regulatory networks.

Here, I dissected inter- and intra-tumor heterogeneity in patients with different genetic aberrations representing major subgroups in AML, namely *MLL* fusions, *IDH* mutated, and *FLT3-ITD* rearranged AMLs. I established and adjusted the experimental and bioinformatic procedures to generate reproducible and scalable data by scRNA/ATAC-seq of peripheral blood and bone marrow biopsies from AML patients. I could demonstrate that leukemic cells could be successfully distinguished from the microenvironment based on marker gene annotation from the human cell atlas and ploidy inference. Furthermore, I used the experimental and data analysis framework to analyze specific molecular features of the three AML subgroups.

First, I characterized changes in the transcriptome and classified developmental stages of leukemic cells carrying *MLL-EDC4* fusions along the hematopoietic stem cell to the myeloid trajectory compared to other *MLL* fusions. Cell type prediction revealed extensive malignant cell diversity and a phenotype skewed towards stem- and progenitor-like populations in *MLL-EDC4* leukemic cells. To further elucidate transcriptomic properties of *MLL-EDC4* cells, TF activity was inferred. The results agreed with differential gene expression highlighting many TFs that play a critical role in hematopoiesis, endothelial-to-hematopoietic transition, or leukemic stem cell activation. Second, I developed an approach to resolve the subclone-specific response during *FLT3* inhibition with midostaurin. Analysis from scRNA and scATAC

showed different *FLT3* activity/chromatin signatures within clusters of leukemic cells in the relapse that could be explained by midostaurin resistance and the emergence of distinct subclones as detected by scDNA-seq. Third, I characterized how the chromatin accessibility landscape was influenced by *IDH1* mutated cells treated *ex vivo* with targeted therapy compared to *IDH1* wild-type cells. Treatment with the *IDH1* inhibitor revealed a partially reversible pattern of accessibility while other mutation-induced epigenetic modifications could not be reverted.

The scRNA-seq data acquired for the three different AML subgroups were then exploited to perform a cell type prediction analysis. The relative abundance of different malignant cell types discovered varied amongst tumors, with some having just two identities and others having a wide range of malignant cells. *MLL* fusions, except for *MLL-EDC4*, generally conferred a more differentiated phenotype predominantly consisting of monocytes/macrophage CD14-like and promonocyte-like cells. Both tumor entities harboring *FLT3*-ITDs or *IDH1* mutations showed a more complex composition of cell types along the myeloid differentiation trajectory than *MLL* fusions. The composition of cell types was generally more skewed to early progenitors at the point of diagnosis when compared to their matching relapse sample. This indicates a partial differentiation of AML cells that treatment might induce.

In summary, this thesis provides novel insights into the tumorigenesis process in AML by using a systematic and functional analysis approach of the transcriptome and open chromatin in single cells for three major genetically defined AML subgroups. A better comprehension of cellular hierarchies, epigenetic effects, clonal evolution, and their impact on gene regulation might help to understand disease progression, stratify patient risk, and help to improve the treatment of hematopoietic malignancies in the future.

Zusammenfassung

Bei der akuten myeloischen Leukämie (AML) kann die Tumorgenese durch multiple onkogene Mutationen in verschiedenen Stadien der Hämatopoese entstehen. Obwohl maligne Transformationen die entsprechenden transkriptomischen und epigenetischen Profile beeinflussen, weisen Krebszellen biologische Merkmale auf, die sowohl auf krankheitsspezifischen Veränderungen als auch auf Mustern beruhen, die dem Zelltyp der Ursprungszelle und der damit verbundenen Differenzierung ähneln. Die sich daraus ergebende Zusammensetzung der Blasten und ihrer Ursprungszellen können sowohl die Heterogenität zwischen als auch innerhalb des Tumors einer molekular definierten AML-Untergruppe erklären. Dies ist für die Risikostratifizierung aber auch für personalisierte Behandlungsstrategien von großer Bedeutung. Sequenziermethoden, die das Transkriptom und Epigenom einzelner Zellen analysieren, wie scRNA-seq und scATAC-seq, sind ideal geeignet, um die Heterogenität sowohl von Tumorzellen als auch von nicht-malignen Zelltypen in der Mikroumgebung zu untersuchen. Darüber hinaus kann mit scATAC-seq die Bindung von Transkriptionsfaktoren (TF) erörtert und auf zellspezifische Netzwerke geschlossen werden.

In dieser Arbeit untersuchte ich die inter- und intratumorale Heterogenität unter Einbeziehung subklonaler Merkmale bei Patienten mit verschiedenen genetischen Aberrationen, die wichtige Untergruppen der AML darstellen, nämlich *MLL*-Fusionen, *IDH*-mutierte und *FLT3*-ITDs. Dazu habe ich experimentelle und bioinformatischen Abläufe entwickelt und angepasst, um reproduzierbare und skalierbare Daten durch scRNA-seq in primären menschlichen Zellen zu erzeugen. Durch die Anwendung dieser Methode konnte ich zeigen, dass leukämische Zellen erfolgreich von der Mikroumgebung unterschieden werden können, basierend auf der Annotation von Markergenen aus dem „Human Cell Atlas“ und Ploidie-Inferenz. Darüber hinaus nutzte ich den experimentellen und datenanalytischen Rahmen, um spezifische Merkmale der drei AML-Untergruppen zu analysieren.

Zunächst charakterisierte ich die Veränderungen im Transkriptom und klassifizierte die Entwicklungsstadien leukämischer Zellen, die *MLL-EDC4*-Fusionen tragen, entlang der Entwicklung von hämatopoetischen Stammzellen zu myeloischen Zellen im Vergleich zu

anderen *MLL*-Fusionen. Die Vorhersage des Zelltyps ergab eine große Vielfalt maligner Zellen und einen Phänotyp, der auf stamm- und vorläuferartige Populationen in *MLL-EDC4*-Leukämiezellen ausgerichtet ist. Zur weiteren Charakterisierung der transkriptomischen Eigenschaften von *MLL-EDC4*-Zellen wurde die Aktivität von TFs abgeleitet. Die Ergebnisse stimmten mit der differentiellen Genexpression vieler TFs überein, die bei der Hämatopoese, dem endothelialen zum hämatopoetischen Übergang oder der Aktivierung leukämischer Stammzellen eine entscheidende Rolle spielen. Zweitens habe ich die molekularen Merkmale der subklonspezifischen Reaktion während der FLT3-Hemmung mit Midostaurin aufgeklärt. Die Analyse von scRNA und scATAC zeigte eine unterschiedliche FLT3-Aktivität/Chromatinsignatur innerhalb von Clustern leukämischer Zellen im Rezidiv, die durch Midostaurin-Resistenz und die Entstehung verschiedener Subklone, wie durch scDNA-seq nachgewiesen, erklärt werden konnte. Drittens habe ich untersucht, wie die Landschaft der Chromatinzugänglichkeit von IDH1-mutierten Zellen, die ex vivo mit einer gezielten Therapie behandelt wurden, im Vergleich zu IDH1-Wildtyp-Zellen beeinflusst wurde. Die Behandlung mit dem IDH1-Inhibitor ergab ein teilweise reversibles Muster der Zugänglichkeit, während andere mutationsbedingte epigenetische Veränderungen nicht rückgängig gemacht werden konnten.

Die für die drei verschiedenen AML-Untergruppen gewonnenen scRNA-seq-Daten wurden anschließend zur Klassifizierung des Zelltyps genutzt. Die relative Häufigkeit der malignen Zelltypen variierte von Tumor zu Tumor, wobei einige nur zwei Identitäten aufwiesen und andere ein breites Spektrum verschiedener Zelltypen verzeichneten. *MLL*-Fusionen, mit Ausnahme von *MLL-EDC4*, führten im Allgemeinen zu einem differenzierteren Phänotyp, der überwiegend aus Monozyten/Makrophagen CD14-ähnlichen und Promonozyten-ähnlichen Zellen bestand. Beide Tumorentitäten, die *FLT3*-ITDs oder *IDH1*-Mutationen aufwiesen, zeigten im Vergleich zu *MLL*-Fusionen eine komplexere Zusammensetzung der Zelltypen entlang der myeloischen Differenzierungstrajektorie. Die Komposition der Zelltypen war zum Zeitpunkt der Diagnose im Vergleich zur entsprechenden Rezidivprobe generell stärker auf unreife Vorläuferzellen ausgerichtet. Dies deutet auf eine partielle Differenzierung der AML-Zellen hin, die durch die Behandlung induziert werden, könnte.

Zusammenfassend lässt sich sagen, dass diese Arbeit neue Einblicke in den Tumorentstehungsprozess bei AML liefert, indem sie einen systematischen und funktionellen Analyseansatz des Transkriptoms und des offenen Chromatins in einzelnen

Zellen für drei wichtige genetisch definierte AML-Untergruppen verwendet. Ein besseres Verständnis der zellulären Hierarchien, der epigenetischen Effekte, der klonalen Evolution und ihrer Auswirkungen auf die Genregulation könnte dazu beitragen, den Krankheitsverlauf zu verstehen, das Patientenrisiko zu stratifizieren und die Behandlung von hämatopoetischen Erkrankungen in Zukunft zu verbessern.

Abbreviations

α-KG	α-ketoglutarate
ALL	Acute lymphoid leukemia
AML	Acute myeloid leukemia
Bp	Base pair
CH	Clonal hematopoiesis
CLL	Chronic lymphoid leukemia
CML	Chronic myeloid leukemia
CMP	Common myeloid progenitor
COL	Cell of origin of leukemia
CR	Complete remission
CRh	Complete remission with partial recovery of peripheral blood counts
CT	Chemotherapy
CTCF	CCCTC-binding factor
DNMT1	Methyltransferase 1
EFS	Event-free survival
FLT3	FMS-like tyrosine kinase 3
FOX	Forkhead box protein
FRiP	Fraction of reads in peaks
GMP	Granulocyte-macrophage progenitor
GRN	Gene regulatory network
HIF1A	Hypoxia inducible factor 1α
HSC	Hematopoietic stem cells
HSCT	Hematopoietic stem-cell transplantation
IDH-DS	IDH inhibitor differentiation syndrome
Ig	Immunoglobulin
ITD	Internal tandem duplication
ITH	Intra tumoral heterogeneity
JM	Juxtamembrane
KDM	Lysine demethylases
KID	Kinase insert domain
KLF	Krüppel-like factor
LEDGF	Lens epithelium-derived growth factor
LMPP	Lymphoid-primed multipotential progenitor
LSC	Leukemia stem cell
MDSC	Myeloid-derived suppressor cells
MeCP2	Methyl-CpG binding protein 2
MEN1	Menin
MEP	Megakaryocyte erythroid progenitor
MLL	Mixed lineage leukemia
MLL-r	MLL-rearranged
MPP	Multipotent progenitor
MRD	Measurable residual disease

Mut	Mutation
NGS	Next-generation sequencing
ORR	Objective response rate
OS	Overall survival
PDGFR	Platelet-derived growth factor receptor
PHD	Plan homeodomain
PKC	Protein kinase C
PTD	Partial tandem duplication
QC	Quality control
R-2-HG	R-2-hydroxy glutarate
R/R	Relapsed or refractory
RCT	Radio chemo therapy
RD	Repression domain
RPCA	Reciprocal principal component analysis
Sc	Single-cell
SET	(Su(var)3-9, enhancer of zeste, trithorax)
SNL	Speckled nuclear localization domain
T-AML	Therapy induced AML
TF	Transcription factor
TKD	Tyrosine kinase domain
TKI	Tyrosine kinase inhibitor
TM	Transmembrane
VAF	Variant allele frequency
VEGFR	Vascular endothelial growth factor receptor
WT	WildtypeWild-type

Introduction

1 Deregulation of hematopoiesis

The hematopoietic hierarchy originates from hematopoietic stem cells (HSCs) that primarily reside within the bone marrow. HSCs can produce all blood cellular components through hematopoietic precursors that ultimately differentiate into mature cells [1] (Figure 1). Furthermore, HSCs have the potential to divide asymmetrically, which generates two daughter cells with different developmental fates. While one daughter cell will keep the possibility of self-renewal and remain in the bone marrow, the other cell will differentiate along a specific lineage and enter the bloodstream [2]. Initiation of tumorigenesis can happen throughout all stages of maturation and lead to various forms of leukemia. Commonly, acute leukemias, such as myeloid and lymphoid leukemia (AML and ALL, respectively) confer a more undifferentiated phenotype, while chronic leukemias present with high levels of differentiated cells in the blood [3]. This thesis focusses on the myeloid lineage disorder AML.

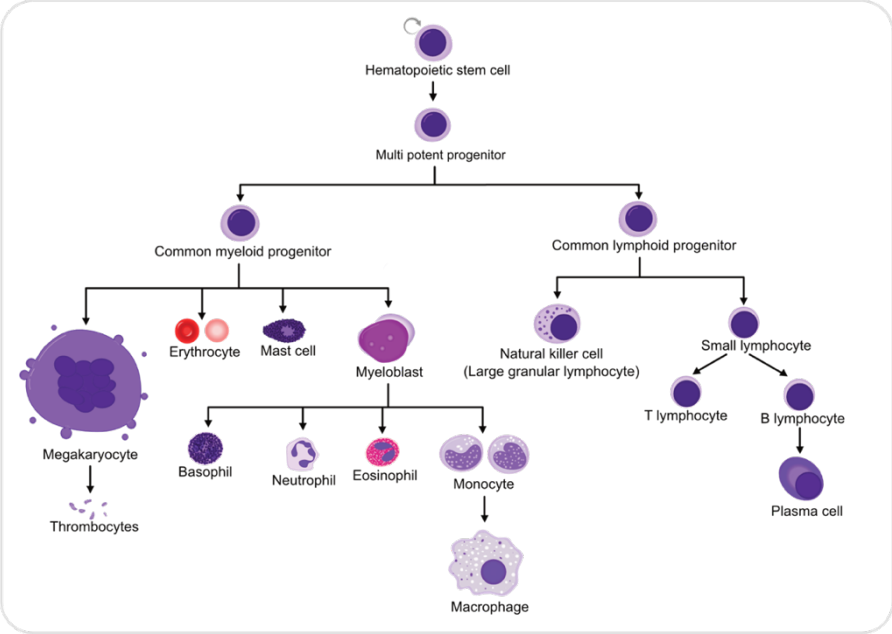


Figure 1

Diagram of normal hematopoiesis. HSCs have broad self-renewal capacity and give rise to distinct hematopoietic progenitors. The progenitor cells possess the ability to proliferate but cannot self-renew and are committed to one or more cell lineages. These progenitors generate numerous precursor cells which mature into differentiated hematopoietic cell types, as shown. Figure adapted from [4].

1.1 Pathogenesis of acute myeloid leukemia (AML)

In AML, the growth and differentiation of HSCs or progenitors become abnormal, leading to an accumulation of immature myeloid progenitors in the bone marrow and peripheral blood. This expansion of myeloblasts (Figure 1) happens at the expense of normal formation of their terminally differentiated counterparts, such as erythrocytes, leukocytes, or platelets [5]. These biological features are caused by several genetic and epigenetic alterations in AML cells. Structural changes such as balanced translocations and chromosomal gains or losses occur in around half of all AMLs [6].

AML is one of the most common types of leukemia and is three to four times more frequent in adults than ALL, where differentiation of the lymphoid lineage is disordered.

The clinical presentation of the disease is very heterogeneous and can range from an incidental finding through routine blood work to a severe condition that demands instant treatment. Apart from nonspecific symptoms such as decreased appetite and fatigue, bone marrow failure is fundamental to the disease. Other characteristic symptoms are shortness of breath on execution caused by anemia, a higher tendency of bruising and bleedings due to thrombocytopenia, and recurring infections because of neutropenia.

With a median age of 70 years in the western world, it can be termed a disease of the elderly [7]. For patients below the age of 60, overall survival (OS) has improved considerably in the past three decades, but treatment for older age groups has not improved significantly [8]. The ongoing improvement of OS in younger individuals but not in older patients can be partly explained by better tolerance of therapy and less accumulation of multiple poor risk factors. Prior to the 1960s, AML was thought to be an incurable disease. In the 1970s, a small percentage of patients – primarily those below the age of 60 years— survived thanks to advancements in chemotherapy, allogeneic hematopoietic stem-cell transplantation (HSCT), and supportive care [9]. In addition to the factors mentioned above, treatment innovation for older patients is restrained by rigorous inclusion and exclusion criteria in clinical trials. This results in a skewed selection of individuals based on age, frequently leading to the exclusion of older patients with more varied performance statuses and co-morbidities. Consequently, the outcomes of clinical studies might not be indicative of AML in the common population [8].

AML cells have far fewer coding sequence mutations than solid tumors. More than one cooperative aberration is necessary to develop AML, according to experimental findings from mouse models [10]. These cooperative mutations can be divided into functional groups based on how they affect different cellular functions (Table 1) [11]. Genome sequencing enabled the classification of these cooperating mutations into nine distinct functional groups, replacing old models with only three broad categories [11].

Table 1 Functional groups in AML based on genetic aberrations

Functional group	**Mutated genes**	**Frequency**
Signaling pathways	*FLT3*, *KIT*, *KRAS*, *NRAS* and serine/threonine kinases	59%
DNA methylation	*DNMT3A, IDH1, IDH2* and *TET2*	44%
Chromatin modifiers	*ASXL1, EZH2* and *MLL* fusions	30%
Nucleophosmin	*NPM1*	27%
Myeloid transcription factors	*CEBPA* and *RUNX1*	22%
Transcription factors	*MYH11–CBFB, RUNX1–RUNX1T1* and *PML–RARA*	18%
Tumor suppressors	*PHF6, TP53* and *WT1*	16%
Spliceosome complex	*U2AF1* and *SRSF2*	14%
Cohesin complex	*RAD21, SMC1, SMC3* and *STAG2*	13%

1.2 Genetic subtypes in AML

1.2.1 Aberrations in mixed lineage leukemia (*MLL*) gene

The *MLL* gene is located at 11q23 and encodes approximately 4000 amino acids long protein[12]. The N-terminus associates with the nuclear factors Menin (MEN1) and lens epithelium-derived growth factor (LEDGF) and allows to bind *MLL* target genes. While MEN1 serves as a link between MLL and LEDGF, LEDGF is a binder of dimethylated H3K36 [12,13]. The N-terminal region also consists of AT-hook motifs (DNA binding domains), two repression domains (RD1, RD2), with the first harboring an additional CxxC domain, and a speckled nuclear localization domain 1 and 2 (SNL-1, SNL-2). The CxxC domain shows homology to DNA methyltransferase 1 (DNMT1) and binds non-methylated CpG DNA[14]. In the middle, *MLL* harbors four plan homeodomain (PHD) fingers that facilitates protein-protein interactions and a bromodomain for binding acetylated lysine residues. A transcriptional activation- and a SET domain are found at the C-terminus (Figure 2A). Mono-, di-, and trimethylation of H3K4 are catalyzed by the SET domain *in vitro* [15,16]. The third PHD finger

helps MLL interact with the cyclophilin CYP33, which is critical for negative *MLL* target gene regulation [17].

In most *MLL* fusion proteins, the PHD finger, bromo-, activation-, and SET domains are all lost [18] (Figure 2B). The endopeptidase taspase-1 proteolytically cleaves wild-type *MLL* after translation. This results in two protein fragments, namely the N-terminal (MLL-N) and the C-terminal (MLL-C) fragment, that are assembled into a large protein complex that can read and write H3K4me3 chromatin signatures, allowing for marking active promoter regions within the genome [19,20]. RbBP5, Ash2L, and WDR5 are some of the other vital proteins that make up the MLL core complex [21]. These proteins assemble into a complex which can bind to distinct H3K4 methyltransferases that have SET domains, such as MLL [22] (Figure 2C). MLL then recruits these proteins along with the histone acetyltransferases CBP/p300 and hMOF, to individual target genes [21,23]. It was recently demonstrated that recruitment of hMOF and other histone modifying proteins is critical for *MLL* target gene expression, whereas H3K4 methyltransferase activity of MLL is not required.

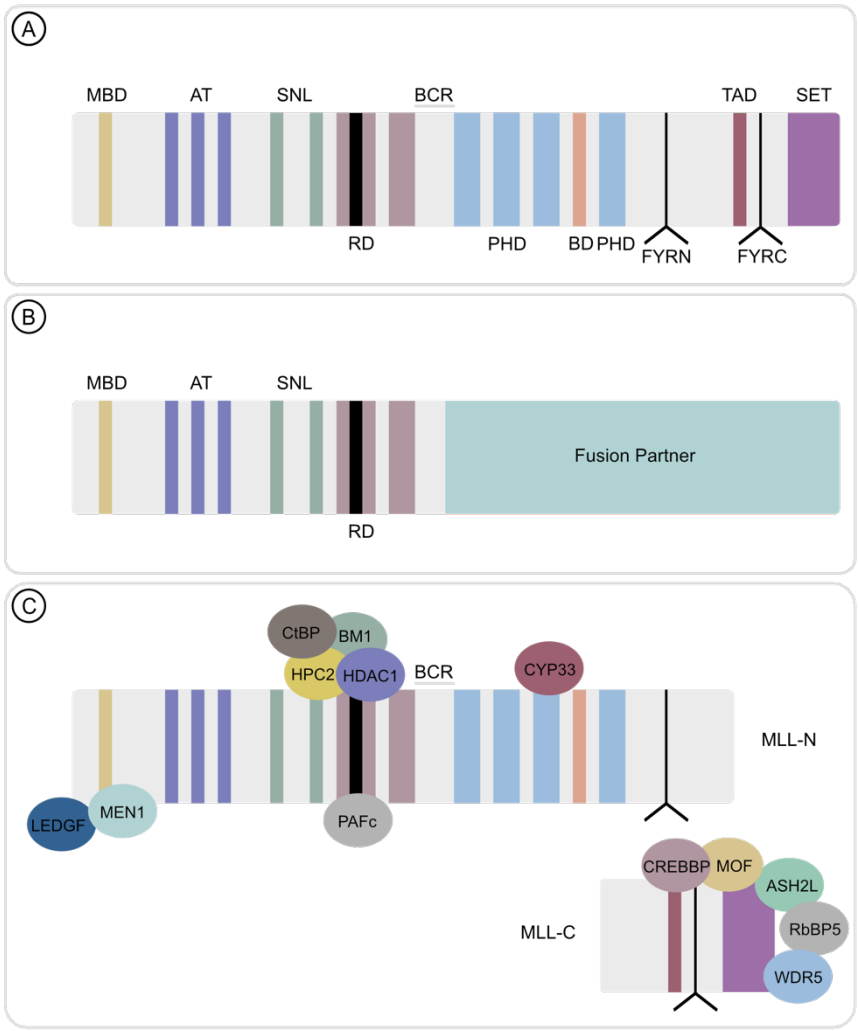


Figure 2

Structure of the MLL protein and wild-type versus aberrated MLL complexes. A Structure of wild-type MLL protein and its functional domains. AT, AT hooks; BCR, breakpoint cluster region; BD, bromodomain; CS1/ CS2, cleavage sites of taspase-1; FYRN/FYRC interacting domains of MLL-N and MLL-C after cleavage, MBD, Menin-binding domain; PHD, PHD fingers; RD, repression domains (black box in the first RD represents the CXXC domain); SET, H3K4 histone methyltransferase domain; SNL, speckled nuclear localization domains; TAD, transactivation domain. B Chromosomal rearrangements lead to in-frame fusion between the N-terminus in MLL and a fusion partner. PHD -, transactivation - and SET domains are lost. C Proteins interact with the two MLL-N and MLL-C fragments. Proteins conferring repressive functions are depicted above the *MLL* protein, whereas activating proteins are grouped below [24].

These epigenetic functions are exerted in a cell type-specific fashion by the MLL complex and are critical for the maintenance of stem cells and developmental processes, including mature cells. By retaining transcriptional memory, MLL complex ensures tissue identity [25].

To sustain the activation of essential regulatory genes in embryonic development [26] or hematopoiesis[27], *MLL* epigenetically maintains gene activation partly by H3K4 methylation [15]. For example, *MLL* regulates the expression of *HOX* genes throughout hematopoiesis.

While stem and early progenitors show high expression of *HOX* genes, there is a gradual decrease during cell maturation[28]. Thus, aberrations in *MLL* and subsequent changes in *HOX* expression have a robust oncogenic potential in leukemia [29].

AML gene rearrangements, usually balanced chromosomal translocations, have all been linked to the development of ALL or AML[30]. This is not a rare occurrence since chromosomal translocations are regularly linked to the development of human cancers. More than 350 recurring chromosomal translocations have been linked to various forms of human cancer. *MLL* translocations, on the other hand, are notable for their vast frequency; more than 80 distinct direct and 130 reciprocal *MLL* fusions have been detected in acute leukemias. Thus, the *MLL* gene is critical, as fusion genes of this gene appear to be sufficient for developing hemato-malignancies. However, most of these specific leukemias are caused by a small number of *MLL* fusions [30]. In more than 90% of all fusion partners the chromosomal translocations t(4;11) [*AFF1/AF4*], t(9,11) [*MLLT3/AF9*], t(10;11) [*MLLT10/AF10*], and t(11;19) [*MLLT1/ENL*] occurred in ALL while 48% of cases were caused by t(9;11) [*MLLT3/AF9*], t(10;11) [*MLLT10/AF10*], t(11;19) [*MLLT1/ENL*], and t(11;19) [*MLL/ELL*] in AML. As a result, the majority of diagnosed *MLL* rearranged leukemias are driven by a small fraction of *MLL* translocation partner genes. The cooperation of these most common fusion partners in a super elongation regulatory complex led to the hypothesis that translocations involving AF4, *AF5, AF9, AF10, AF17, ELL, ENL, pTEFb, or DOT1L* could induce deregulated transcriptional elongation or comparable phenotypes. However, mutually exclusive binding of numerous members was discovered, implying the existence of various smaller, instead of one large complex [31-35] (Figure 3).

Furthermore, this explanation failed to account for the various clinical phenotypes in relation to the fusion partner. Additionally, *MLL* fusions with partners such as septins, *MLL* partial tandem duplications (PTDs), or cytosolic coiled coil domain protein were not taken into consideration for this hypothesis. Until now, *MLL* rearrangements lack a simple and unified understanding of how they cause leukemia [24].

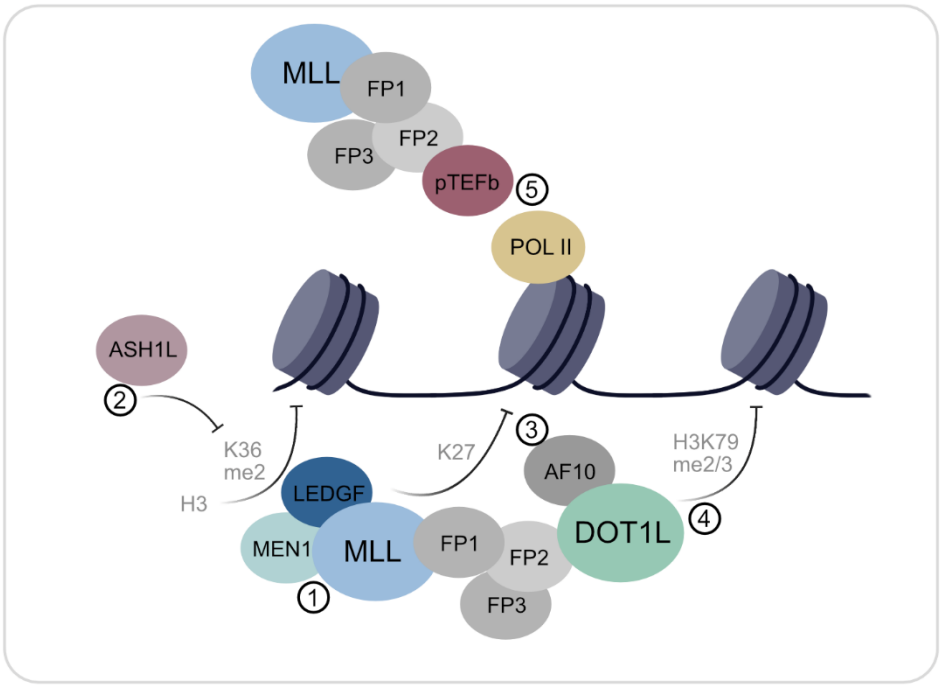


Figure 3

Protein complexes and nuclear proteins involved in transcriptional elongation and histone modifications. AF4, AF5, AF10, AF17, ENL and ELL have all been reported to fusion not only with MLL but also with each other. pTEFb and DOT1L probably do not assemble within one complex. Numbers display occasions for targeted inhibition. (**1**) *Protein-protein*: *MEN1-MLL* interaction. (**2**) *Chromatin*: LEDGF-K3K36me2 (inhibiting reader domain or ASH1L) (**3**) *AF10* – unmodified H3K27 interaction (inhibiting reader domain or demethylases) (**4**) *DOT1L* – placement of H3K79me2/me3 (inhibiting methyltransferase domain). *Pol II phosphorylation*: (**5**) Blocking of *pTEFb* [24].

However, a comparison of genome-wide gene expression analyses in *MLL*-rearranged (*MLL-r*) versus *MLL* wild-type leukemias has shown that this subtype of leukemia exerts a distinct gene expression signature regardless of translocation partner and lineage differentiation [36-38]. *HOX* cluster genes, especially *HOXA7-HOXA10* and the *HOX* cofactor *MEIS1,* are the most frequently upregulated genes in *MLL-r* leukemias [39,40]. The transcription factors (TF) encoded by *HOX* genes regulate developmental processes such as segmentation and hematopoiesis [39,41-43]. Expression of *HOX* genes and *MEIS1* is uppermost in hematopoietic stem cells and early progenitors and decreases with differentiation [41,44]. The developmental regulators seem to play a pivotal role in *MLL-r* leukemias and maintain or confer cell growth, self-renewal abilities, and advantages in survival. These properties might cause high levels of treatment resistance and an overall poor prognosis [45-47].

1.2.2 Aberrations in FMS-like tyrosine kinase 3 (*FLT3*)

FLT3 is a transmembrane tyrosine kinase receptor of the class III family and is encoded by the *FLT3* gene, which is located on chromosome 13q12. FLT3 consists of an extracellular region harboring five immunoglobulin(Ig)-like domains, following a transmembrane domain (TM) and an intracellular domain containing the juxtamembrane (JM) and two kinase domains, which are separated by a kinase insert domain [48] (Figure 4).

Figure 4
Domain structure of FLT3. The extracellular domain consists of five IG-like domains. Following the TM the intracellular region contains the JM and two kinase domains, separated by the KID [49]. IG, immunoglobulin; TM, transmembrane; JM, juxtamembrane; KID, kinase insert domain.

FLT3 plays a crucial role in hematopoietic cell survival and proliferation [50]. The ubiquitously expressed FLT3 ligand binds to FLT3, which operates as a cytokine receptor. FLT3 dimerizes and changes conformation after binding to its ligand, allowing auto-phosphorylation of the ATP-binding pocket. Through the PI3K/AKT, MAPK/ERK, and STAT signaling cascades, FLT3 receptor activation results in decreased apoptosis, enhanced cell proliferation, and impaired hematopoietic cell differentiation [51] (Figure 5). *FLT3* mutations are found in about 30% of AML patients, making it the most prevalent molecular abnormality in the disease [52]. These molecular mutations can be divided into tyrosine kinase domain (TKD) mutations and internal tandem duplications (ITDs). *FLT3*-ITDs are found in around 25% of individuals with *de novo* AML and occur in the auto-inhibition regulating the juxtamembrane domain of the FLT3 receptor in exons 14 and 15 [53]. The length of the duplication can range from only three to up to more than 400 base pairs (bp). At the same time, the transcriptomic reading frame is preserved, either by nucleotide insertion at the ITD junction to retain the original reading frame or by in-frame duplication [54]. Thus, the FLT3 receptor becomes more sensitive to its ligand, resulting in enhanced leukemic cell proliferation. Point mutations usually occur at

aspartic acid residue D835 in the activating loop of the tyrosine kinase domain (*FLT3*-TKD) in approximately seven percent of *de novo* AML cases [54].

Both *FLT3* TKDs and ITDs produce aberrant signaling via the *ERK, PI3-kinase*, while the latter alters *STAT5* signaling, leading to stem cell transformation [55,56]. While the prognostic relevance of *FLT3*-TKD is unclear, *FLT3*-ITD mutations are associated with an unfavorable outcome [57].

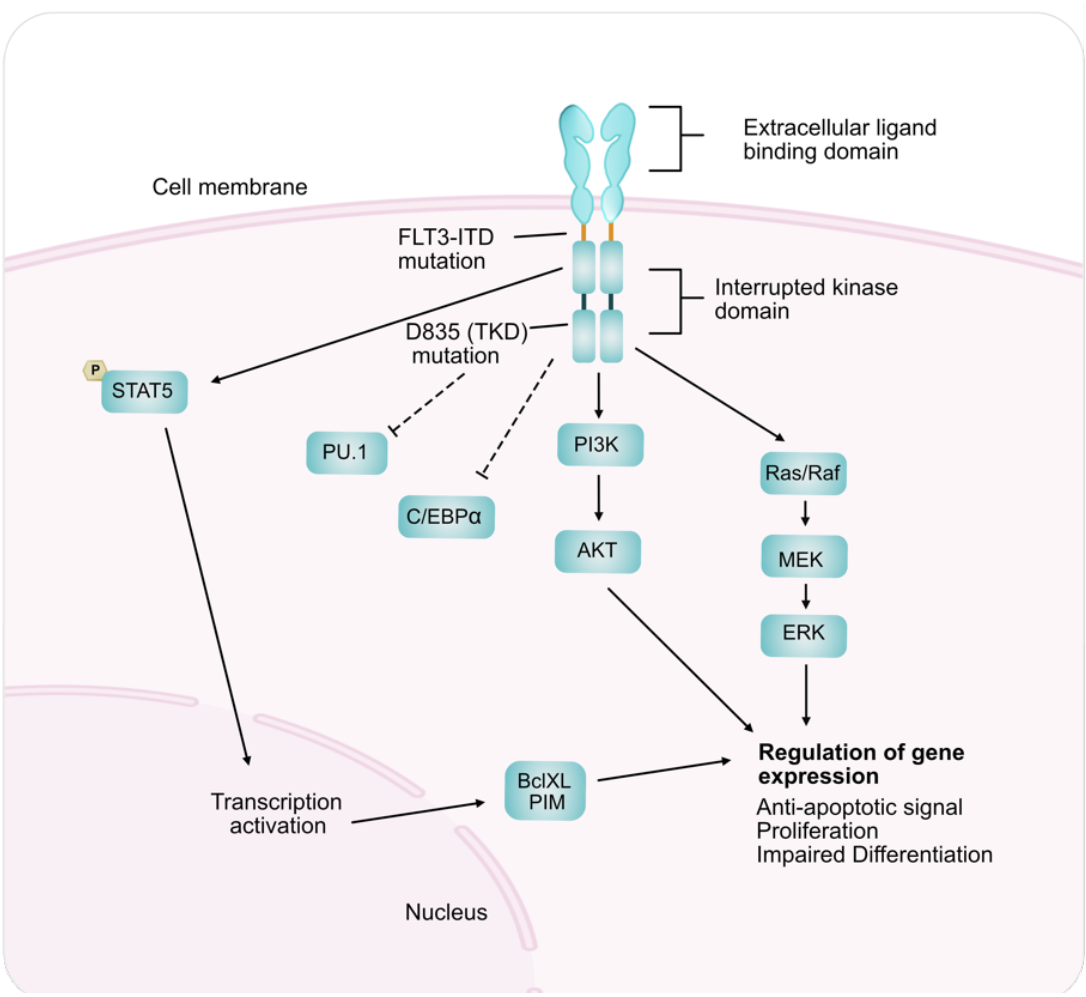


Figure 5

Schematic overview of tyrosine-protein kinase and its interactions. Activating mutations of *FLT3* caused ITDs or point mutations in TKD lead to a constitutive signaling of tyrosine kinase. FLT3 signaling recruits pathways, such as Ras/Raf, PI3K, and STAT5, promoting cellular proliferation and chemoresistance [58]. ITD, internal tandem duplication; TKD, tyrosine kinase domain; P3K, PI3-kinase.

1.2.3 Aberrations in Isocitrate dehydrogenase (*IDH*)

The *IDH1* gene and its mitochondrial homolog *IDH2* are located on chromosomes 2q33 and 16q26 and encode catalytic enzymes IDHs 1 and 2, respectively [59]. IDHs belong to the family of homodimeric enzymes with roles in hypoxia adaption and cellular metabolism [60-63]. Additionally, they are one of the epigenetic regulators that have been found to be functionally disrupted in leukemia. The IDH enzymes are involved in the citric acid cycle and normally function to catalyze the oxidative decarboxylation of isocitrate, resulting in α-ketoglutarate (α-KG) to generate NADPH from $NADP^+$. For the proper action of numerous dioxygenases involved in metabolic and epigenetic control, appropriate cellular levels of α-KG are required [64-67] (Figure 6A).

Aberrations of those enzymes promote the conversion of α-KG to the oncometabolite (R)-2-hydroxyglutarate (R-2HG) (Figure 6B). This product shift is caused by a decreased affinity for isocitrate and increased affinity for α-KG, probably due to conformational and biochemical changes [68-70]. The structural similarity of R-2-HG to α-KG, except for the oxidation state on the C-2 carbon position, leads to competitive inhibition of α-KG-dependent dioxygenases regulating a variety of cellular mechanisms like DNA demethylation, hydroxylation and degradation of hypoxia-inducible factor 1α (HIF1A) or histone to protein hydroxylation [71-73]. Especially dioxygenases such as the TET family of DNA hydroxylases and the JmjC-domain-containing histone demethylases are most crucial. TET2 promotes DNA demethylation by converting 5-methylcytosine to 5-hydroxymethylcytosine [67,74], and its hydroxylase activity has been demonstrated to affect other epigenetic processes besides methylation alterations [75,76]. The mechanism for TET2-mediated leukemogenesis, such as enhanced proliferation and self-renewal properties that might be associated with a deregulation of *HOXA* genes, indicates tumor-suppressing properties in wild-type *TET2* [74,77-80].

Besides functional inhibition of TET2 creating a hypermethylation profile, other R-2-HG-dependent epigenetic mechanisms have been demonstrated. These effects include histone modifications, transcriptional deregulation caused by impairment of histone lysine demethylases (KDMs) [64,75,76,81] , and altered splicing involving modulation of RNAPII by CCCTC-binding factor (CTCF) and methyl-CpG binding protein 2 (MeCP2) due to changes in DNA methylation [82].

IDH1 and *IDH2* somatic mutations occur in around 20% of AML patients and arise as early clonal events in disease progression [83]. Additionally, the mutations are relatively stable throughout disease progression, probably because these mutations represent early genomic events [84]. *IDH2* mutations have a higher incidence in older patients and intermediate risk groups and around 8–19% of AML patients are affected [11,83,85]. Patients with *IDH* mutations generally have a higher blast and platelet count, and more often neutropenia [86-88]. The mutation usually occurs in one of two arginine hotspots inside the enzymatic active site, either at residue R140 or R172. R140 is altered in about 80% of cases while the latter one is mutated in about 20% of instances. *IDH1* mutations are seen in 7–14% of AML patients, and the most prevalent mutation is a cysteine or histidine substitution for arginine, R132C and R132H, respectively [83,85,89].

Numerous studies have investigated the role of *IDH1* and *IDH2* mutations on prognosis in multiple myeloid malignancies, such as AML, with varied outcomes [90]. A recent meta-study indicates worse overall survival for AML patients with mutations in *IDH1,* whereas *IDH2* mutations had a favorable prognostic impact [91]. Another study focusing on *IDH* mutation ancestry could link ancestral *IDH1* aberrations to a worse prognosis than subclonal *IDH1* mutations. In contrast, survival of patients with *IDH2* mutations could not be associated with clonal hierarchy [84]. These findings show that clonal dominance of mutations in *IDH*, which is at least applicable for *IDH1*, might play a role in these individuals' outcomes.

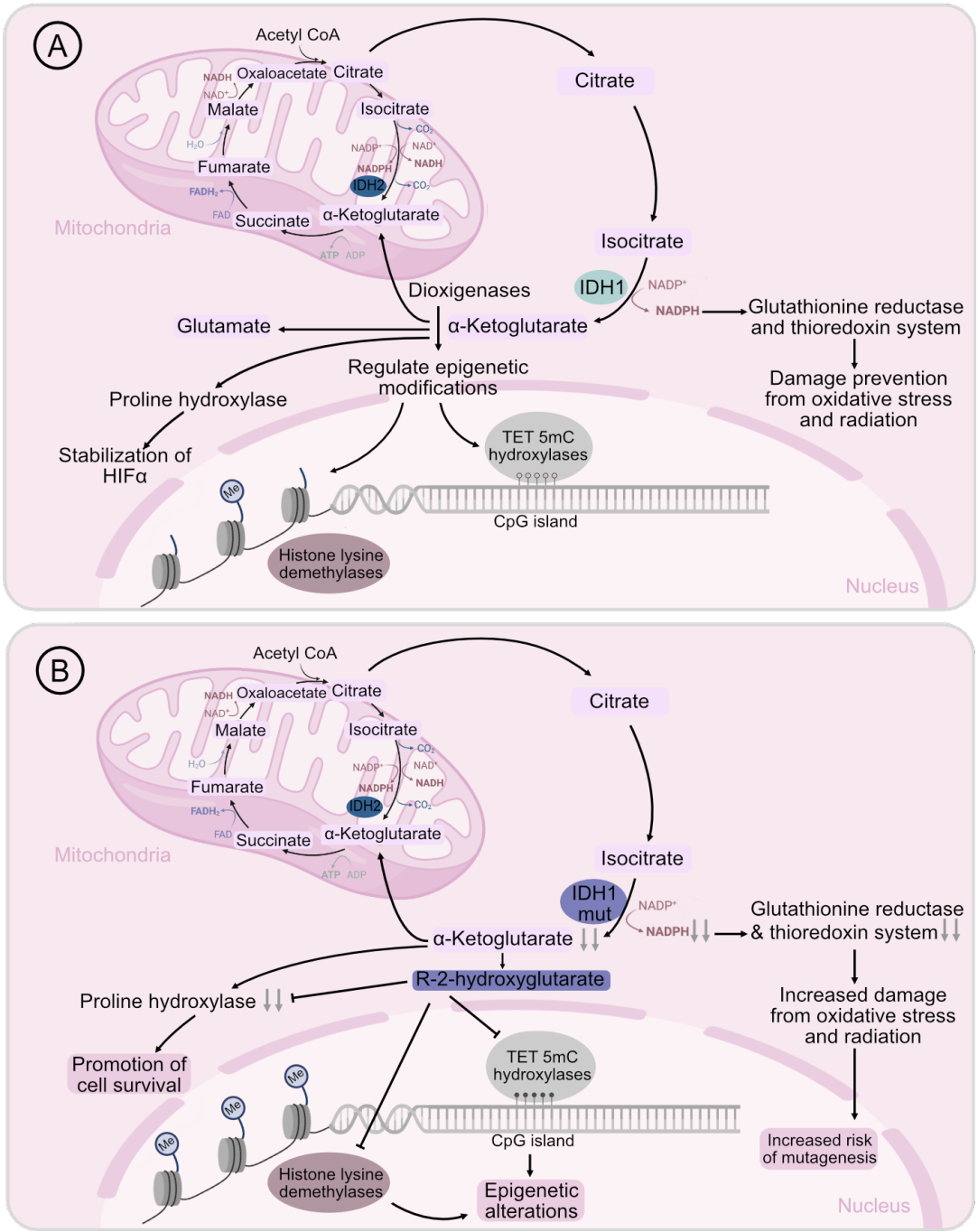


Figure 6

The role of *IDH* in metabolism. (**A**) Normal function of wild-type *IDH* in the tricarboxylic acid cycle (TCA). While IDH1 is in the cytoplasm and peroxisomes, IDH2 resides in the mitochondria. IDH catalyzes the oxidative decarboxylation of isocitrate to α-ketoglutarate, producing NADPH in the process. A proline hydroxylase-mediated mechanism controls the degradation of hypoxia-inducible factor alpha (HIF1A) by α-ketoglutarate. Dioxygenases, such as histone lysine demethylase and the TET enzymes, control epigenetic modifications and use α-ketoglutarate as a co-substrate. NADPH contributes to the glutathione reductase and thioredoxin systems, which protect cells from oxidative and radiation-induced damage. (**B**) The presence of mutant *IDH1* causes the generation of a neomorphic oncometabolite R-2-hydroxyglutarate which is produced from α-ketoglutarate. This causes inhibition of histone lysine methylase and TET enzymes, which are required for proper epigenetic control. Simultaneously, the production of α-ketoglutarate and NADPH is reduced. This prevents HIF1A from degrading properly and increases cell survival. Reduced NADPH production raises the risk of oxidative stress and mutagenesis [92].

1.3 Therapy of AML

1.3.1 Standard therapy

AML treatment is usually divided into induction, consolidation, and maintenance. During induction, cytotoxic chemotherapy is used to induce complete remission, favorably without the measurable residual disease (MRD). In most cases, consolidation therapy is less intense than induction therapy unless HSCT is used. Maintenance therapy doses are frequently considerably lower than those used during induction [93]. The amount of cytotoxic chemotherapy that can be administered strongly depends on the patient's overall health, which is directly tied to age. Patients with AML require extensive supportive care, including blood and platelet transfusions, as well as an antimicrobial medication due to treatment and immunocompromisation. Since the 1960s, a chemotherapeutic drug called cytarabine has been used for routine therapies of patients with lymphoma or leukemia [94]. Cytarabine is an S-phase specific antimetabolite that incorporates into the DNA, subsequently impairing DNA polymerase and synthesis[95].

In patients <60 years, standard remission induction therapy usually comprises 7–10 days of cytarabine in combination with three days of an anthracycline, a DNA intercalating drug that interferes with DNA metabolism and binds topoisomerase II [96]. Consolidation therapy usually entails numerous sessions of high-dose cytarabine [97]. In younger individuals with AML, this regime results in long-term cures in 30-40% [98] whereas 5-year survival of patients >60 years is less than 15% [99]. Dissecting the heterogeneity of AML at the cytogenetic, molecular, and clinical levels resulted in enhanced predictive and prognostic capacities, as well as the development of targeted therapies for specific AML subsets [100].

1.3.2 Midostaurin as targeted therapy for patients with *FLT3*-ITDs

Although genetic factors influence AML prognosis [93,101], therapeutic targeting of recurrently altered genes is difficult [102]. Aberrations in epigenetic regulator genes like *ASXL1* or *DNMT3A* are frequently found in AML [103] and are associated with a poor prognosis [83,104]; the same alterations are also seen in hematopoietic stem cells of healthy older adults with clonal hematopoiesis [105,106]. Although mutations associated with age-related clonal hematopoiesis appear to be essential drivers of clonal development toward AML [103,106], their presence does

not automatically make them therapeutic targets [107]. Mutant oncoproteins only qualify as therapeutic targets if their inhibition causes growth restriction, differentiation, or a reduction in the viability of AML blasts. The constitutively active FLT3 receptor tyrosine kinase, caused by mutations in the *FLT3* gene, is a perfect target oncoprotein in AML. *FLT3* mutations usually appear late in the course of AML pathogenesis [83,108]. *FLT3* oncoproteins hijack the signal transduction machinery of leukemic cells, resulting in a significant reliance on *FLT3*-signaling pathways for survival. When *FLT3* signaling is disrupted by a tyrosine kinase inhibitor (TKI), apoptosis is induced both *in vitro* and *in vivo* [109,110]. Due to the dependence of *FLT3* mutated AML cells on FLT3 oncoproteins, clinical development of FLT3 inhibitors represents an immense potential to fight this subtype of AML.

Midostaurin is a multikinase inhibitor that targets FLT3 and was authorized by the FDA as the first targeted therapy for AML patients harboring *FLT3* mutations [111]. This first-generation inhibitor is relatively nonspecific (in contrast to second generation inhibitors) with additional activity against receptor targets, including platelet-derived growth factor receptor (PDGFR), protein kinase C (PKC), vascular endothelial growth factor receptor (VEGFR), CDK1, c-Syk,c-Src, c-Kit and c-Fgr [112,113] (

Figure 7). FLT3 inhibitors are further divided into two groups based on their receptor-interaction mechanism. All FLT3 inhibitors block FLT3 receptor phosphorylation and activation by preventing ATP binding to the tyrosine kinase domain (TKD) [114]. Type I inhibitors such as midostaurin or gilteritinib can interact with FLT3 both in the active or inactive conformation. In contrast, type II inhibitors only can bind to FLT3's hydrophobic region, that is accessible in the inactive conformation only. *FLT3* D835 mutations at the TDK consequences an active conformation, preventing type II inhibitor-FLT3-TKD interaction. Inhibitors of the type II might select for D835 mutations over time, resulting in acquired resistance, but inhibitors of type I can sustain efficacy at both *FLT3*-TKD and *FLT3*-ITD mutations [115]. In the RATIFY placebo-controlled phase III trial, administration of midostaurin versus placebo combined with 3+7 induction, consolidation, and maintenance for previously untreated *FLT3*$^{+}$ AML were studied. Median OS and event-free survival (EFS) were significantly improved regardless of *FLT3* mutation or tumor burden. However, the midostaurin and placebo groups´ complete remission (CR) rates were nearly identical. The most common adverse effects of midostaurin treatment were anemia, nausea, and rash [111].

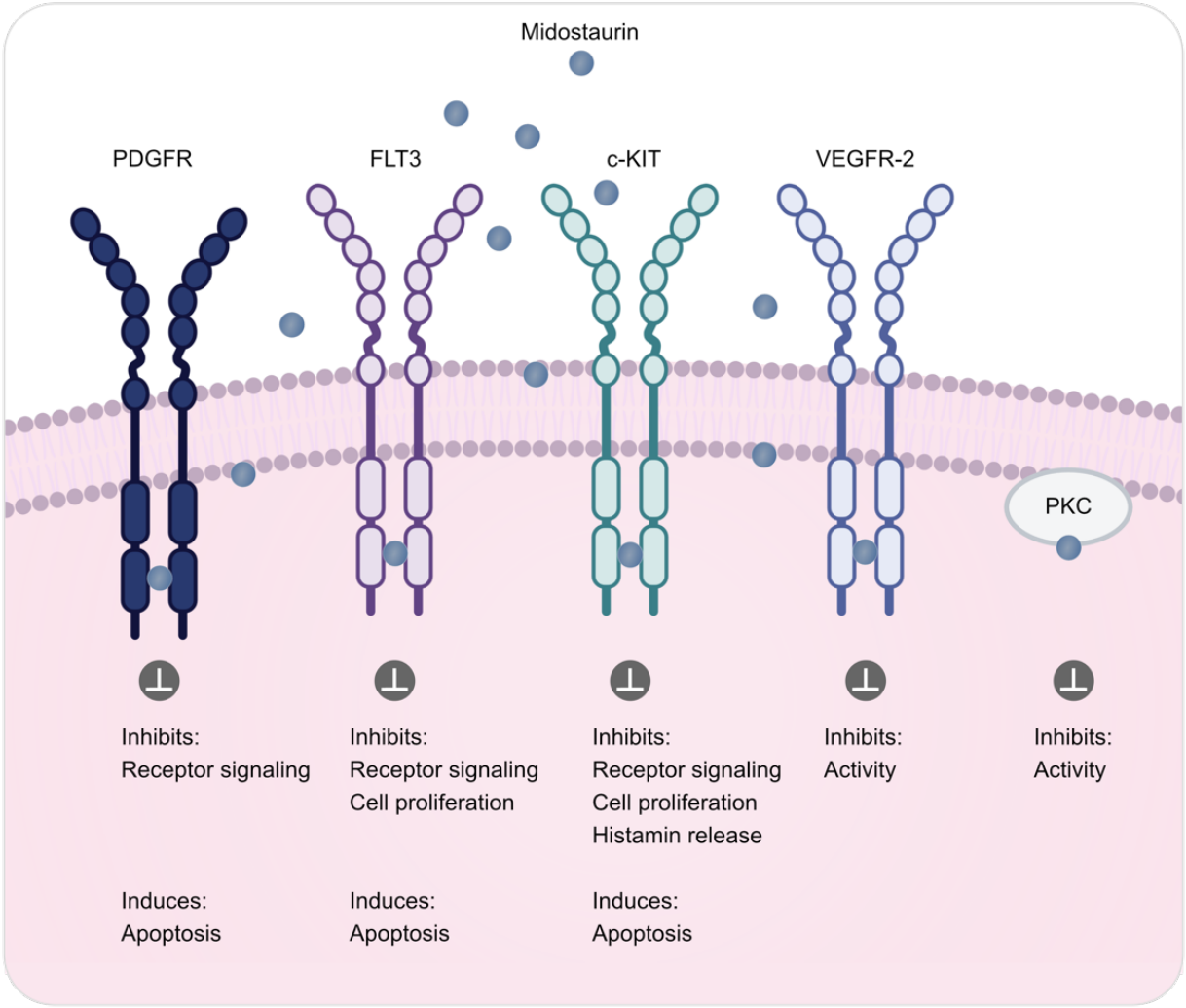


Figure 7

Multitarget tyrosine kinase-inhibitor midostaurin. Midostaurin binds to the catalytic domains of the proliferation-promoting tyrosine kinases FLT3, c-KIT, PDGFR, VEGFR-2, and serine/threonine kinase family PKC. Leukemia cells with *FLT3*-ITD and TKD mutations are susceptible to midostaurin. As a result, there is a reduction in the function, proliferation, and survival of these cells, as well as decreased angiogenesis [116].

1.3.3 Targeted therapy for patients with *IDH* mutations

Mutant IDH inhibitors embody a new class of metabolic cancer therapy that induces tumor cell differentiation. Among these, the two orally available and selective small molecular inhibitors, ivosidenib (AG-120) and enasidenib (AG-221), target recurrent mutations in the isocitrate dehydrogenase genes *IDH1* and *IDH2*, respectively [117]. Both IDH inhibitors have been linked to IDH inhibitor differentiation syndrome (IDH-DS), an uncommon yet fatal adverse side effect. The onset of IDH-DS can range from 7 to up to >120 days after initiation, and symptoms include fever, pulmonary infiltrates, pericardial effusion, weight gain, acute kidney injury, or skin rash. To counteract these symptoms, treatment interruption alone might not be sufficient, necessitating corticosteroid therapy [118].

Enasidenib is FDA approved for treating patients with $IDH2^{mut}$ relapsed or refractory (R/R) AML, that are unfit for intensive chemotherapy. This agent inhibits the formation of the mutant enzyme by binding to its allosteric site and maintaining a homodimer conformation, preventing a conformation change required for its catalysis and subsequent generation of R-2-HG [119]. Even though enasidenib monotherapy is not licensed for this use, there has been increased interest in using it as a first-line treatment for older patients with newly diagnosed $IDH2^{mut}$ AML who are not suitable for intensive chemotherapy. However, the exceptionally high overall response rates seen with azacitidine and venetoclax argue against monotherapy with enasidenib. However, new trials [120] investigating the effects of IDH inhibitors combined with venetoclax and/or azacytidine will show whether there is an overall survival benefit for newly diagnosed *IDH1* mutant AML.

Ivosidenib (AG-120) is an *IDH1* inhibitor that has been approved by the FDA for treatment of individuals with untreated and R/R $IDH1^{mut}$ AML who are not candidates for intensive chemotherapy. The FDA first approved this drug due the findings of a phase I trial which showed ivosidenib treatment resulted in a 41.6 percent overall response rate in patients with *IDH1*-mutant R/ R AML, with a third of patients reaching complete remission/ complete remission with partial recovery of peripheral blood counts (CR/CRh) [121]. Even though this was a single-arm study, a significant clinical benefit associated with ivosidenib therapy was shown, including fewer infections in responders and extinction of detectable $IDH1^{mut}$ in a

quarter of patients. Most adverse effects included diarrhea, nausea, fatigue, leukocytosis, decreased appetite, thrombocytopenia, anemia, peripheral edema and IDH-DS was reported in 18%. Because of the superior response rates achieved with upfront venetoclax and azacytidine therapy in elderly patients with newly diagnosed *IDH1*^{mut} AML, numerous clinicians are preceding monotherapy with AG-120 in the frontline context [120]. In October 2020, the company Agios withdrew its European marketing authorization application after the European Medicinal Agency's (EMA) Committee for Medicinal Products for Human Use determined that results from Agios' single-arm, uncontrolled phase I study did not not provide enough evidence to establish a favorable benefit-risk balance for the suggested indication. The company stated to advance their ongoing phase III randomized controlled trials evaluating ivosidenib combinations in newly diagnosed AML with the goal of successful approval both in Europe and in the United States [122].

BAY1436032 is an oral small-molecule inhibitor of mutant *IDH1* that also interacts with an allosteric site on the mutant enzyme [123]. Preclinical studies showed that BAY1436032 suppresses R-2HG synthesis and colony formation *in vitro* while increasing myeloid differentiation, clearance of leukemic blasts, and survival in mouse models [124]. Based on these promising results, a clinical phase I study was initiated to evaluate four different dose levels of BAY1436032 in individuals harboring a mutation in *IDH1*. BAY1436032 scored an OR rate of 15% and a substantially lower OS than other mutant *IDH1* inhibitors in R/R AML [121]. The low response rate could be caused by incomplete target inhibition and a shorter half-life compared to ivosidenib. The extent to which differences in target inhibition, as seen by lower R-2-HG levels and clinical response rates attained by BAY1436032 compared to ivosidenib, are attributed to pharmacokinetic properties is unclear [125]. Some oncogenic qualities of mutant *IDH* are independent of R-2H-G [72], and it is unknown how effective BAY1436032 and ivosidenib are in inhibiting R-2HG-independent oncogenic effects. Currently, there is another non-randomized, open-label, phase I study to define the maximum tolerated and/or recommended phase II dose to characterize its safety, tolerability, pharmacokinetics, and pharmacodynamics [126].

2 Mechanisms of cancer progression

2.1 Developmental hierarchies in AML

AML pathogenesis has been studied extensively to determine how closely AML populations resemble the hierarchical developmental process seen in normal hematopoiesis [127-129]. Studies have shown that leukemic cells have a hierarchical structure resembling the differentiation hierarchy of non-malignant hematopoiesis [127-129]. The subclonal environment can contain both cells that hold stem-cell properties and highly developed cells without self-renewal capacity. The former ones are termed leukemic stem cells (LSCs) and are characterized to initiate, propagate, and sustain leukemia after transplantation [127-129]. LSCs possess infinite self-renewal capacity and constantly create immature blood cells. Naive leukemic blasts accumulate due to the block of the hematopoietic differentiation mechanism, one of the hallmarks of AML[127-129]. Although leukemic blasts lack the ability to self-renew, they may drive pathologic effects on hematopoietic function or tumor microenvironment [130]. Non-malignant cells of the microenvironment also have an impact on disease progression. The immune system is able to defend expansion of cancer cells until subclones emerge that decrease or evade host immunity. Extrinsic microenvironmental alterations and intrinsic features of leukemic cells, such as the production of immunomodulatory proteins, might cause accumulation of suppressive T regulatory cells and hinder cytotoxic T lymphocyte activation [130].

The existence of cancer stem cells was first experimentally proven in AML, and it is now widely accepted that AML relies on stem cells with self-renewing ability for maintenance of AML clones, primarily poorly differentiated blasts [1].

Early research revealed that a fraction of $CD34^+$ $CD38^-$ show LSC characteristics, suggesting a similar immunophenotype to non-malignant HSCs [131]. These cells evolved into AML blasts after being transplanted into NOD/SCID mice, illustrating a hierarchical structure of a misdirected hematopoiesis originating from the stem cell level. This seemed true for most AML subtypes at the time, except for acute promyelocytic leukemia, where $CD34^+$ $CD38^-$ cells could not confer leukemia in xenograft models.[132]. The expression pattern of several surface markers was then used to characterize the phenotype of human AML LSCs better. Markers

such as CD123 (IL3RA) [133], CD96 [134,135], C-type lectin-like molecule-1 (CLEC12A, CLL-1) [136-138] , and TIM-3 [20] are thought to be unique for LSCs and are being explored as targeted therapy for leukemia[139]. Recent research has discovered that AML LSCs are significantly more heterogeneous than previously thought. LSCs can also be found in the $CD34^-$ and $CD38^+$ compartments [140,141] or in populations expressing markers of committed progenitors such as CD38 or CD45RA [142].

Additionally, it was shown that in AML samples with an increased fraction of $CD34^+$ cells, the activity of LSCs could also be determined in expanded subgroups with an immuno-phenotypical progenitor state. These findings revealed that LSCs immunophenotype could be similar to those of granulocyte-macrophage progenitors (GMP) (Figure1) and lymphoid-primed multipotential progenitors (LMPP), indicating a wide phenotypic variety [143]. In addition to possessing infinite self-renewal capacity, LSCs have been thought to share many biological traits with non-malignant HSCs, including being predominantly in the quiescent G0 phase of the cell cycle, and higher drug efflux pump activity, and only accounting for a small percentage of total leukemic cells [144]. It is hypothesized that while leukemic blasts are eradicated during conventional chemotherapy, LSCs survive this regimen due to their quiescent cell cycle state leading to relapse. Because driver mutations in AML are so heterogeneous, the biological features such as LSC frequency, immunophenotype, and cell cycle vary substantially amongst leukemias [145]. It was also discovered that the frequency of LSCs varies substantially between leukemias [132,146]. Recent studies highlighted that the LSC frequency could change over time, with up to two logs greater LSC frequency detected in particular relapses compared to their matching sample at diagnosis [145]. LSC frequency is vital for the clinical course of AML, with studies indicating that a greater LSC frequency confers poor prognosis and shorter relapse-free survival [147,148].

Despite these challenges, the benefits of LSC-targeted therapeutics are appealing enough that extensive research is being conducted to find features that discriminates LSCs from HSCs and can be used to synthesize LSC-targeted agents [144]. LSC-specific antigens like CD123 or metabolic vulnerabilities such as the reliance on enhanced oxidative phosphorylation, more significant oxidative stress, increased activation of stress response pathways such as NF-kB signaling, or dependency on unfolded protein response pathways are promising features that could be exploited for therapeutic intervention [139].

However, Van Galen *et al.* showed that undifferentiated cells with self-renewal properties could be of broad interest. Also, differentiated malignant populations drive the clinical presentation of AML [149]. Monocyte-like leukemic cells found in varying abundances in AML patients effectively suppressed T-cell activation *in vitro*. This indicates a potential role in altering T-cell phenotypes and creating an immunosuppressive microenvironment [130]. Thus, characterization of mechanisms and functions of immunomodulatory subcompartments and assessing their connection with myeloid-derived suppressor cells could be used to alter their activity for therapeutic adjustments [150,151].

2.2 The cell of origin in AML

The normal hematopoietic cell from which the LSC develops through a series of changing events is known as the cell of origin of leukemia (COL). LSCs can either arise from a non-malignant cell by a gain of proliferation of an HSC or a gain of self-renewing abilities of a committed progenitor (Figure 8).

Usually, the phenotype of LSCs will reflect the consequence of numerous subsequent oncogenic mutations, which can result in a phenotype that does not necessarily resemble the COL. Since most AMLs harbor four to eight genetic abnormalities that can be identified as driver mutations, this malignancy is genetically extremely heterogeneous [11]. Understanding the features of the COL is vital for understanding the origins of leukemia and for fathoming the underlying LSC biology. This insight might help to stratify risk factors for disease development and to improve LSC-targeted therapies [139].

The human hematopoietic system is not very receptive to experimental manipulations; hence progress in finding the COL in human AML has been gradual over many years. A body of research has shown that the expression of selected *MLL-X* fusions is potent inducing a leukemic phenotype in mice [152,153]. *MLL-ENL* and *MLL-AF9* fusions expressed in lineage marker-depleted (Sca-1 and c-Kit-positive) bone marrow stem and progenitor (LSK) cells, CMPs, or GMPS, were associated with a similar phenotype [47,154,155]. However, HSC-derived clones were more likely to initiate AML than GMP-derived clones in mice. Moreover, leukemias with stem cell-derived COL were less responsive to chemotherapy than the GMP-derived leukemias. These findings show the impact of the origin of a cell on gene expression,

epigenetic state, and treatment responsiveness in AML. The differences in COL might explain clinical heterogeneity within molecularly characterized AML subgroups [47].

Better insight into the COL in human AML arose with the introduction of next-generation sequencing (NGS) tools and their broad use to analyze AML genetics in detail. The identification of *DNMT3A* aberrations associated with AML found in lymphocytes, collected at remission, indicates the existence of a preleukemic HSC clone [156]. Since most hematopoietic cells are short-lived, the acquisition of numerous mutations in a particular clonal lineage demonstrates a gradual accumulation, beginning with HSCs that have self-renewal abilities [157].

Without self-renewal properties, any obtained mutation will be lost in a short period because the progenitor cells exhaust their replicative potential after a finite number of divisions and daughter cells differentiate. AML, like most malignancies, becomes more common as people get older. After finding somatic mutations in HSCs and differentiated blood cells of older adults without hematological impairment, a precursor condition called age-related clonal hematopoiesis (CH) was defined. *DNMT3A*, *TET2,* and *ASXL2* were the most prevalent CH-associated alterations, and patients with these mutations had a considerably greater probability of developing myeloid malignancies, including AML. As a result, myeloid cells in the CH compartment might have a clonal advantage over their regular counterparts in terms of survival and the ability to acquire further mutations, resulting in the formation of LSCs. Since preleukemic clones arise long periods before the onset of disease, only a small number of people with CH progress to AML, strengthening the assumption that HSCs are the COL [158,159]. Nevertheless, patterns of sequential mutation accumulation from self-renewing CH stem cells to LSCs, and then to progenitors without self-renewal properties should be characterized for better understanding of clonal expansion mechanisms and future intervention of developing AML. The low frequency of LSCs hinders proper identification and genetic characterization.

Similarly, following AML treatment, a small yet persistent cell population, known as measurable residual disease (MRD), is frequently enriched with hard to track LSCs . Surveillance is vital because MRD can arise from the therapy itself, and specialized targeted therapy could counteract this phenomenon. Most MRD monitoring and clinical DNA sequencing approaches such as NGS or flow cytometry rely on bulk tumor cells and do not

allow for easy clonal evolution tracking in rare tumor cells [160]. However, single-cell multi-omic approaches have demonstrated their superiority in characterizing AML stem cells or hierarchies [149,156,161]. Gene expression distinguishes between stem cells and other leukemic cells, whereas genetic status can discriminate between cancer and healthy populations.

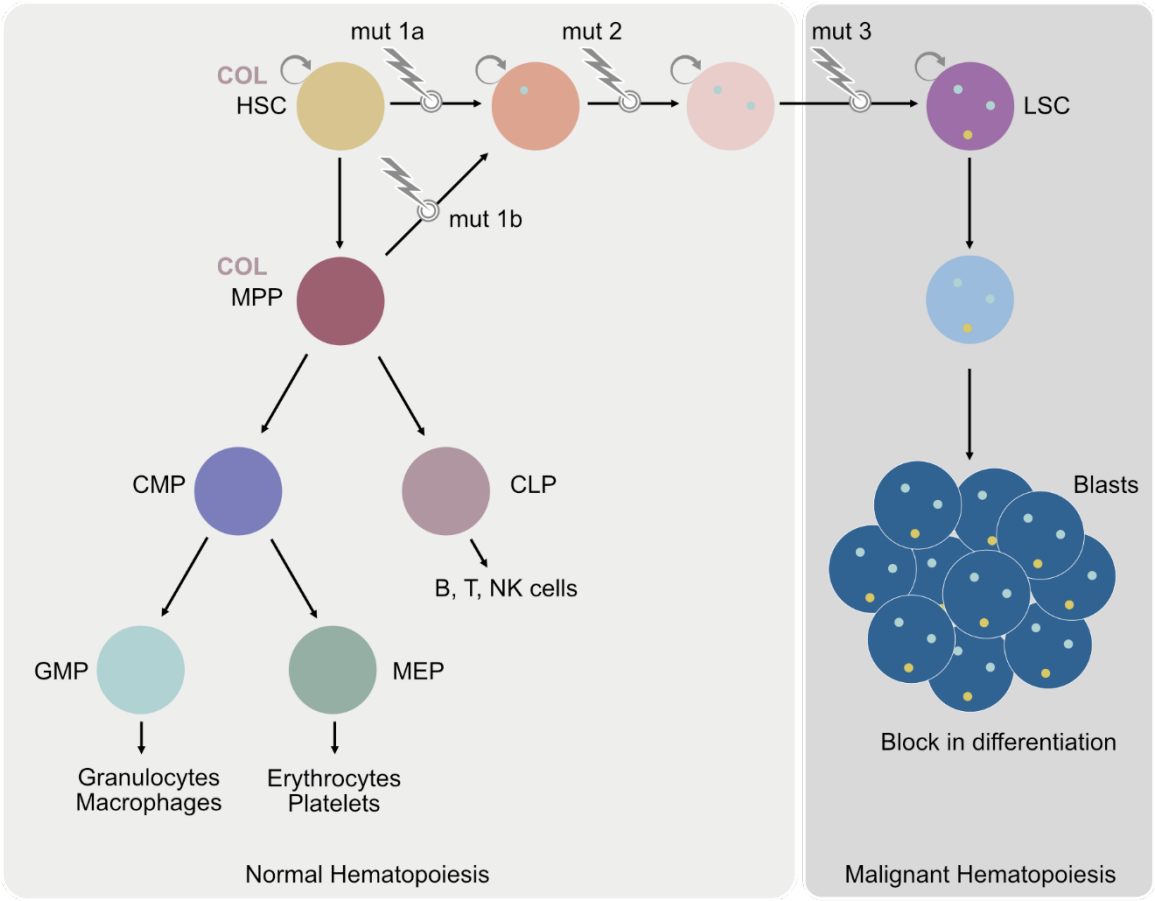


Figure 8

Comparison of normal versus malignant hematopoiesis in AML. Two different scenarios are depicted for the cell of origin. In the first, HSC acquires mutation (mut) 1a, mut 2, and mut 3, leading to a leukemic stem cell with self-renewal properties. LSCs generate blasts without the ability of self-renewal and terminal differentiation. Thus, cells in the right rectangle only mimic normal hematopoiesis but do not produce functional blood cells [139]. CLP, common lymphoid progenitor; COL, cell of origin for leukemia; GMP, granulocyte macrophage progenitor; HSC, hematopoietic stem cell; CMP, common myeloid progenitor; LSC, leukemic stem cell; MEP, megakaryocyte erythroid progenitor; mut, mutation; MPP, multipotent progenitor

3 Single-cell (sc) approaches to define inter and intra-tumoral heterogeneity

Traditional sequencing methods can only retrieve the average of numerous cells, making it impossible to evaluate a small number of cells and resulting in the loss of cellular heterogeneity information. Single-cell methods (Figure 9) offer the benefits of identifying variability among individual cells, detecting rare cell populations, and outlining cell maps compared to classical sequencing technique [162].

The expression of a limited group of uniquely expressed marker genes has traditionally been used to define cell types, including the cell of origin. On the other hand, single-cell genomic analyses reveal a vast range of cell states based on more prominent regulatory traits found within generally well-defined cell identities. Various cellular programs, including inflammatory signaling, stochastic fluctuations in gene regulation inside a cell and the cell cycle, are reflected in variations in the state of a cell [163]. Cell states may be identified by single-cell transcriptomic and epigenomic techniques, and transcription factor activity can be further used to define them.

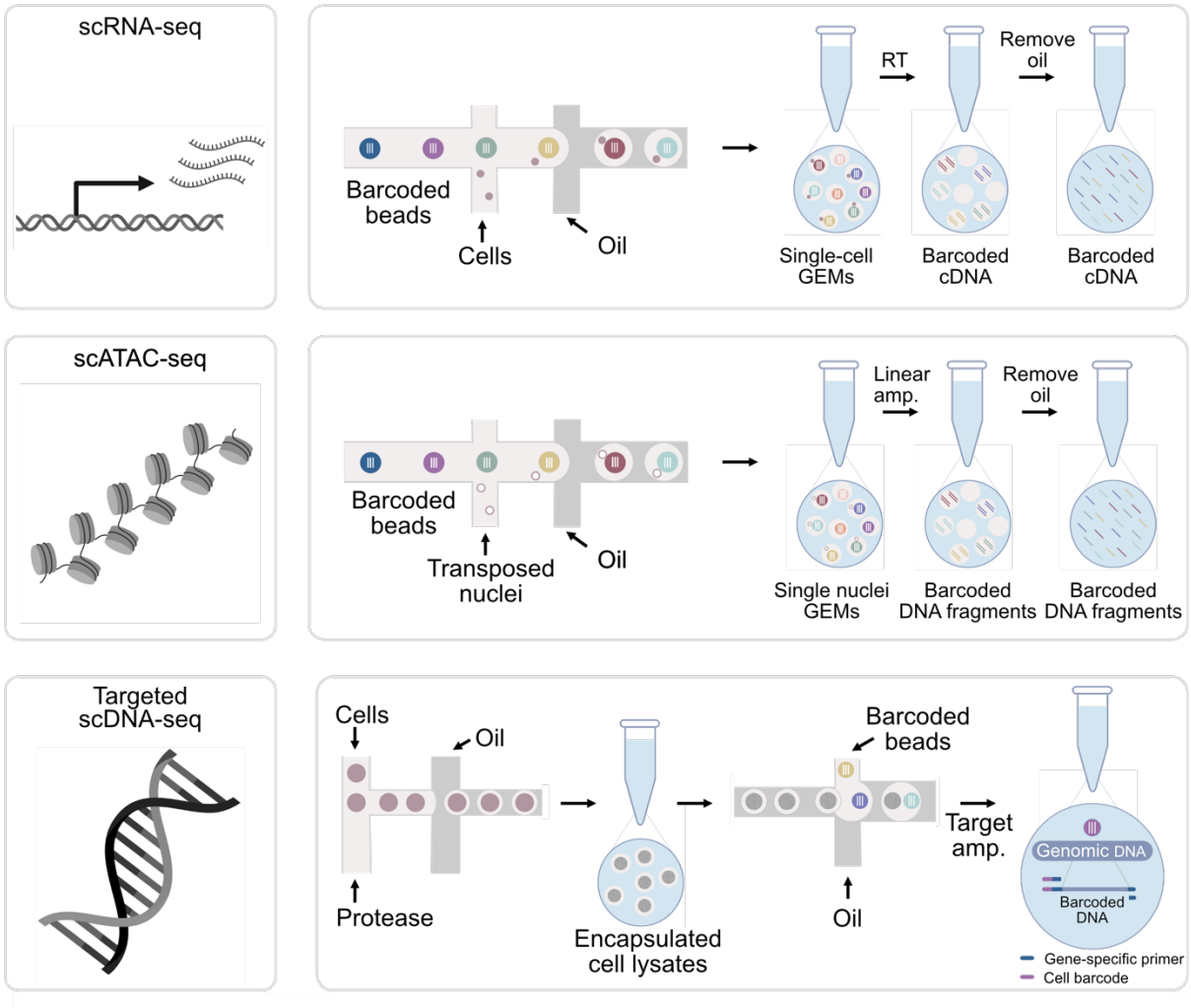


Figure 9

Schematic overview of used single-cell methods. For scRNA-seq, cells are loaded and partitioned on a microfluidic chip in the presence of oil and barcoded gel bead to create GEMs. Amplification and library preparation is performed in bulk after GEMS are broken. For scATAC, nuclei incubated with a transposase enzyme pre-loaded with sequencing adapters, are loaded onto a microfluidic chip in the presence of oil and barcoded gel bead to create GEMs. DNA fragments are then isolated and amplified by PCR. For targeted scDNA-seq, cells are loaded and partitioned on a microfluidic cartridge along with a protease enzyme mix. Cells are lysed, and protease is digested, enabling access to DNA. In a second encapsulation step, cell lysates are combined with primers, barcoded beads, and reagents. The specific regions are then amplified with a unique cell barcode. Amp, amplified; GEM, gel beads in emulsion; RT, reverse transcription; sc, single-cell.

3.1 Single cell RNA sequencing

Transcriptome analysis is a powerful methodology for solving the problem of mapping genes to phenotypes, which has been a long-standing barrier in biology and medicine. Although all cells within an organism have almost identical genomes, transcriptomic information of any cell only represents the activity of a selection of genes. Classical bulk sequencing only produces an average signal of expression from all cells, although different cell types express

a distinct transcriptome. Increasing data reveals that gene expression varies even within cell types [164-166], and that stochastic expression reflects the compositionof cell types and influences cell fate [167,168]. The bulk of transcriptomic analysis continues to be conducted on the premise that cells from a particular tissue are homogenous, meaning that considerable cell-to-cell heterogeneity is likely to be missed. A more accurate knowledge of individual cells' transcriptome will be vital for evaluating their involvement in cellular activities and how gene expression might drive positive or detrimental states to comprehend stochastic biological processes better. Since the first single-cell transcriptome analysis [169] there has been a global boom in interest in acquiring high-resolution pictures of single-cell heterogeneity. Researchers have recently deconvoluted distinct immune cell populations in healthy and disease conditions through significant developments in accessible experimental methods and bioinformatics pipelines [170]. scRNA-seq (Figure 9) is also used to characterize cell lineage relationships in early development [171], cell differentiation [172], or cell fate determination [173]. Additionally, scRNA-seq enables inference of gene regulatory networks (GRNs) that can dissect complex biological processes by revealing regulatory interactions between proteins and genes [174,175]. Especially multimodal studies that combine single-cell epigenomic and transcriptome data will enable the creation of GRNs that provide more insight into regulatory programs and dynamics in cancer progression [176].

Tumor heterogeneity is a specific event that may occur within and between tumors. Unknown tumor characteristics, that have been missed by bulk transcriptome investigations, can be revealed using scRNA-seq [177]. Furthermore, it can be used to examine expression patterns of individual pathways or assess gene programs during drug treatment that might give insights into therapy resistance in cancer cells [178,179]. Individual cells are constantly going through dynamic processes and reacting to a variety of external stimuli. Some of these reactions are quick, while others are more gradual and might take years to manifest. The molecular profile of a cell, including its protein and RNA composition, reflects this dynamic process [180]. Single-cell approaches allow the collection of distinct instantaneous timepoints along a trajectory. Bioinformatic algorithms are then applied to reconstruct dynamic cellular trajectories based on cell cycle and differentiation [172]. Several studies have shown that this pseudo-time assessment can infer cell fate dynamics in cancer [181,182]. However, pseudo-time models may not be well suited to truly depict complex cell trajectories by which cells migrate

between cell states. Actual trajectories can include dead ends that are incompatible with a metastatic progressions, such as quiescent, or senescent cell populations, reversible routes that are consistent with trans- and dedifferentiation[183] or a combination of these dynamics [181,182]. Thus, pseudotime and its accompanying bioinformatic techniques might not be applicable to all cancer lineage trajectories.

Van Galen *et al.* demonstrated that scRNA-seq data is consistent with clinical parameters such as cell-surface markers by exploring transcriptomes of AML patients and a machine learning classifier to distinguish malignant from normal cells [149]. Cells from AML samples were projected onto a trajectory of healthy bone marrow along the HSC to myeloid differentiation axis based on their similarity. Their classifier identified six malignant cell types that resembled one of six normal cell types along the HSC to myeloid trajectory. They were labeled as HSC-like, progenitor-like, GMP-like, promonocyte-like, monocyte-like or, cDC-like. This method has the advantage of an unbiased transcriptional classification that provides more complex information on AML cell types and hematopoietic differentiation and is not restricted to a limited number of predefined markers. Investigation of subclonal transcriptomic expression revealed immunomodulatory properties of monocyte-like blasts, whereas HSC-like cells showed dysregulated transcriptomic patterns. The combination of single-cell genetic and transcriptomic data not only allowed the characterization of the tumor ecosystems in AML, the distinguishment between healthy and malignant cells, but also for definition of AML hierarchies. This study also depicts the potential for scRNA-seq to elucidate LSC evolution or characterization of pre-malignant clones in the future.

3.2 Single cell ATAC sequencing

The human epigenome involves various molecular processes that control the accessibility of genomic information. While genes, that are crucial to maintain particular cell type states, are kept accessible, dispensable genes will be epigenetically silenced. Hence, epigenetic variation provides gene regulatory mechanisms relevant during physiological cell maturation [184,185]. Epigenetics, on the other hand, are not only involved in physiological development but also in disease etiology, including cancer. In disorders like cancer, these gene regulatory mechanisms are disrupted, resulting in altered gene expression that promotes cancer initiation, growth, and metastasis [186,187]. Oncogenic mutations in a putative cell of origin

cause regulatory program disruption, culminating in a cascade of cellular phenotypic modifications and abnormal cell function [188,189]. However, tumor cells maintain an epigenetic memory that reflects their cell-of-origin [190]. It is becoming evident that the epigenetic environment influences the mutational processes that lead to cancer. Thus, genetic and epigenetic variation are strongly intertwined and cannot be treated as independent occurrences [47,190,191].

The fast advancement of single-cell genomic technologies in recent years revealed the range of cell states that exist during cancer formation, maintenance, and therapy resistance [192]. Enhanced cellular plasticity mediates these cellular phenotypes, resulting in progressively evolving changes in chromatin landscape, similar to gene regulation during development, and increased intratumoral heterogeneity (ITH) [193]. Due to this heterogeneity, bulk assays fail to characterize distinct mechanisms of gene regulation that have an impact on tumor progression. Profiling of heterogeneous cancer cell states was improved by single-cell epigenomics [194-198]. Single-cell studies have shown substantial differences in cell identity within tumors, and tumor cells were reported to acquire developmental programs and distinct cell fates [181,199,200]. These losses or shifts of lineage identity contribute to ITH and seem continuous, with no distinct subpopulations in the tumor [200]. Single-cell epigenomic approaches, such as coupled single-cell transcriptomic and epigenomic readouts to identify regulatory dynamics of TFs, can can shed light into the molecular processes that modulate trans- and dedifferentiation fates [201-203]. By exploiting the association of chromatin accessibility with gene expression across single cells, computational tools employing multi-omic single-cell technologies, such as scATAC-seq combined with scRNA-seq, can connect distant enhancers to genes. These methods can detect regulatory chromatin domains that characterize chromatin areas with a high density of enhancer-gene interactions. Lineage-determining genes are usually enriched in domains of regulatory chromatin which highly overlap with regulatory regions termed super enhancers [203,204]. Accessibility at of regulatory chromatin domains precedes gene expression, implying that alterations may prime cells for lineage commitment. The development of single-cell methods that assess chromatin accessibility allows to discover deregulated TFs that enhance the activation of lineage-skewing regulatory elements, exposing processes that drive cell lineage infidelity.

3.3 Single cell DNA sequencing

scDNA-sequencing (Figure 9) overcomes the limits of bulk sequencing to capture DNA mutations that are only present in a small fraction of cells [205]. Sequencing error a fixed proportion of total sequencing coverage, limits bulk DNA sequencing. Although deeper sequencing coverage can improve sensitivity for low-frequency mosaic DNA characteristics, sequencing error further increases. This means that mosaic features lower than 0.5% mosaicism cannot be discriminated from sequencing errors and are not detected [205,206]. scDNA-seq, however, is not restricted by sequencing error because it is substantially less than the anticipated 50% score for heterozygous features. Nonetheless, single-cell DNA amplification errors can sometimes outperform actual genetic traits, necessitating the continual development of innovative technologies for high-fidelity single-cell genome amplification [207]. Additionally, scDNA-seq is capable of detecting co-presence of two different low-level mosaic genetic variants in the same cells or if they are mutually exclusive. Prior to sequencing, bulk procedures homogenize materials, obliterating information on mosaic DNA features within the same cell or other cellular subsets. scDNA-seq, on the other hand, keeps this information. Especially in cancer research, scDNA-seq has paved the way to analyze the evolutionary mechanisms that evolve due to selection forces throughout oncogenesis, proliferation, and therapy [208-210]. Cancer produces genetically heterogeneous lineages throughout time, and the resulting ITH is critical in tumor growth, metastasis, and response to therapy [211]. The ITH is determined by its driver mutations and the subclonal structure of its lineages which are both crucial for understanding its biology. Most cancers harbor more than one main subclonal lineage and many also show a higher number of low-frequency clones [212-215]. Subclonal diversity discovered by scDNA-seq has been linked to tumor subtype in some circumstances; for example, AML patients with *FLT3* mutations show higher clonal diversity than in *FLT3* wild-type AML [216]. One study also highlighted that the presence of four or more subclonal lineages correlates with poor prognosis, indicating subclonal diversity might function as a prognostic marker [217]. While bulk DNA sequencing data may infer ITH [218], scDNA-seq gives a more complete and high-resolved perspective of clonal development and ITH [208]. Subclonal patterns obtained from scDNA-seq and bulk DNA are mostly consistent. However, the first frequently discloses subclones not seen in bulk [215,216,219]. Additionally, scDNA-seq is less impaired by a small sample size or low tumor purity.

Information on cancer onset and its mutational processes could be revealed using phylogenetic studies of ITH profiles acquired by scDNA-seq. There, the occurrences of mutation types and certain driver mutations that arise can be ordered and dissected. Many structural variants, such as copy number aberrations or aneuploidies, are acquired early within a short time frame during tumor evolution, a term called punctual clonal evolution. On the contrary, point mutations can occur continuously with a few later-accumulating focal copy number aberrations that involve specific driver genes [210,219-223]. This indicates that the origins of certain occurrences of many cancers are due to genome-wide, high-impact mutational bursts rather than a gradual accumulation of mutations. Surprisingly, some studies found cell populations within the same patient that could not be linked to the expected tumor lineage yet possessed a modest amount of copy number aberrations or just one driver mutation. This probably represents driver gene anomalies and baseline copy number aberrations in normal cells [216,223,224]. The association between these cells and the onset of cancer is currently unknown. Therefore, it is a significant area for future investigation. scDNA-seq may one day enable the determination of the cells of origin, resolving long-standing arguments over the presence and significance of tumor stem cell hierarchies [208,225]. scDNA-seq has also shown that ITH changes throughout clinical therapy and mutant blasts acquire resistance to treatment. While one study could detect therapy-resistant subclones prior to treatment [226], another study employed targeted scDNA-seq combined with immunological profiling to examine the phenotypic shifts that malignant cells were exposed to after being treated with a medication stimulating erythroid differentiation [227]. The ability of scDNA-seq to reveal different responses of tumor cells with individual genotypes to the same therapy is highlighted in this work. This method will be a crucial guide for converting subclonal genotypes into treatment response predictions if widely adopted. However, there are two drawbacks to using scDNA-seq for therapy monitoring. To begin with, tumor collection during therapy is typically only achievable at discrete, sporadic intervals. Second, ITH and treatment resistance are not solely caused by genetics; epigenetic heterogeneity also influences tumor cell biology [226,228]. Especially when combined with other multi-omics, spatial or high-throughput capabilities, scDNA-seq is a powerful tool to alter cancer research. Genotype-phenotype correlations may be built using multi-omic scDNA-seq to learn how certain subclonal genotypes are linked to cellular phenotypes, invasiveness, therapy responsiveness, and more [227,229-231].

4 Scope of the thesis

This thesis aimed to develop a framework that dissects AML intra-tumor heterogeneity in patients with different genetic aberrations that represent major AML subgroups (*MLL,* IDH^{mut}, and *FLT3*-ITD rearranged AMLs). I investigated AML pathogenesis from different perspectives (transcriptome, open chromatin, and genetic lesions) using a multi-omics approach.

The first aim of the thesis was to establish an experimental and bioinformatic workflow needed for the identification and classification of myeloid leukemic cells. Various tools for leukemic cell determination, integration, and pseudotime analysis were employed to generate a framework that could be used for all further AML patient samples.

The second aim was to dissect inter- and intra-tumoral heterogeneity and classify developmental stages of leukemic cells carrying *MLL-EDC4* fusions along the hematopoietic stem cell to the myeloid trajectory compared to other *MLL* fusions. I could identify the differential stages of AML blasts and uncover a unique phenotype of leukemic cells with an *MLL-EDC4* fusion.

The third aim was to gain mechanistic insight into why responsiveness to midostaurin treatment is lost in *FLT3-ITD* rearranged AMLs. This was addressed by performing a scRNA-seq, scATAC-seq, and targeted scDNA-seq analysis. It revealed links of aberrant *FLT3* activity to gene expression and chromatin accessibility changes of leukemic cells during relapse that the emergence of resistant subclones could explain.

The fourth aim was to assess the ability of mutant *IDH* inhibitors to revert chromatin accessibility changes and aberrant transcription factor binding induced by the mutation. Both experiments using cell lines and primary samples were conducted to explore this epigenetic scarring. A partially reversible pattern of accessibility after treatment of IDH1mut with the inhibitor was observed. Similarly, the scATAC-seq analysis of the TF1 cell line carrying $IDH2^{mut}$ treated with AG-221 could identify partial response to treatment as well.

The fifth aim was to stratify patient-specific AML blasts based on the scRNA-seq analysis along the HSC to myeloid differentiation axis across the three AML subgroups (*MLL* fusions, *FLT3* mutations, and *IDH* mutations) and to compare the influence of treatment on cell type abundance. The analysis revealed an extensive malignant cell diversity and provided detailed information on AML cell types and differentiation states.

By addressing these five aims, my thesis work introduces an integrative single-cell analysis of the transcriptome, open chromatin, and genome to dissect AML cell types. It provides detailed knowledge of cellular hierarchies, epigenetic changes, and their impact on gene regulation. It is concluded that extending this approach to a larger number of patient samples will provide valuable insight into disease heterogeneity and treatment response for clinical decision-making.

1 Materials

1.1 Custom buffers

Table 2 Buffer composition

Buffer	Composition
20× PBS	45 mM Na_2HPO_4; 181 mM NaH_2PO_4; 3 M NaCl (pH 7.4)
Elution buffer	50 mM Tris (pH 8.0), 1 mM EDTA, 1 % SDS, 50 mM NaHCO3
Erythrocyte lysing buffer (pH 7.2)	50 mM Tris (pH 8.0), NH_4Cl, ddH_2O, 1 M HCl
Freezing medium	90 % FBS, 10 % DMSO
TE buffer	10 mM Tris pH 8.0, 1 mM EDT

1.2 Commercial kits and reagents

Table 3 Commercial kits and reagents

Kit/reagent	Item number	Company
10% tween-20	1662404	Bio-Rad, USA
1x penicillin-streptomycin, 10,000 U/mL penicillin, 10 mg/mL Streptomycin	P06-07050	PAN-Biotech, Germany
Agilent high sensitivity DNA kit	5067-4626	Agilent Technologies, USA
Cell titer glo	G7571	Promega, USA
Chromium chip E single cell ATAC kit, 16 reactions	PN-1000086	10x Genomics, USA
Chromium i7 multiplex kit	PN-120262	10x Genomics, USA
Chromium next GEM nhip J single cell kit, 16 reactions	PN-1000230	10x Genomics, USA
Chromium next GEM single cell multiome ATAC kit A, 16 reactions	PN-1000280	10x Genomics, USA
Chromium next GEM single cell multiome reagent kit A, 16 reactions	PN-1000282	10x Genomics, USA
Chromium single cell 3' library & gel bead kit v2	PN-120237	10x Genomics, USA
Chromium single cell 3' GEM, library & gel bead kit v3,16 reactions	PN-1000075	10x Genomics, USA
Chromium single cell A chip kit	PN-1000009	10x Genomics, USA
Chromium single cell ATAC library & gel bead kit, 16 reactions	PN-1000110	10x Genomics, USA
Chromium single cell B chip kit, 16 reactions	PN-1000074	10x Genomics, USA
D1000 ladder	5067-5586	Agilent Technologies, USA
D1000 reagents	5067-5583	Agilent Technologies, USA
D1000 sample buffer	5067-5602	Agilent Technologies, USA
D5000 ladder	5067-5590	Agilent Technologies, USA

D5000 reagents	5067-5589	Agilent Technologies, USA
D5000 ScreenTape	5067-5588	Agilent Technologies, USA
Deoxynucleotide (dNTP) solution mix	N0447S	New England Biolabs, USA
Digitonin	D141-100MG	Merck Sigma-Aldrich, Germany
Dimethyl sulfoxide (DMSO)	D2650	Sigma-Aldrich, Germany
DTT	646563	Millipore Sigma, USA
Dual index kit TT set A, 96 reactions	PN-1000215	10x Genomics, USA
Enasidenib	21277	Cayman Chemical, USA
Ethanol, pure (200 proof, anhydrous)	E7023-500ML	Millipore Sigma, USA
Fast DNA ladder	N3238S	New England Biolabs, USA
Fast SYBR green master mix (2x)	4385612	Thermo Fisher Scientific, USA
Fetal bovine serum	P30-3602	PAN-Biotech, Germany
Gibco RPMI 1640 (ATCC Modification), 500mL	A1049101	Fisher Scientific, USA
Glycerin (glycerol), 50% (v/v) aqueous solution	PN-3290-32	Ricca Chemical Company, USA
High sensitivity D1000 ScreenTape	5067-5584	Agilent Technologies, USA
Human GM-CSF Recombinant Protein	PHC2011	Thermo Fisher Scientific, USA
Library construction kit, 16 reactions	PN-1000190	10x Genomics, USA
$MgCl_2$ (25 mM)	R0971	Thermo Fisher Scientific, USA
NEBnext HF 2x PCR master mix	M0541S	New England Biolabs, USA
Nuclei EZ Lysis Buffer	N3408	Sigma-Aldrich, Germany
Protease inhibitor cocktail (100X)	5871	Cell Signaling Technology, USA
Proteinase K solution (20 mg/mL)	M3037.0005	Genaxxon bioscience, USA
Q5® HF DNA polymerase	M0491S	New England Biolabs, USA
Qubit dsDNA HS assay kit	Q32851	Thermo Fisher Scientific, USA
Qubit RNA HS assay kit	Q32852	Thermo Fisher Scientific, USA
Recovery™ cell culture freezing medium	12648010	Thermo Fisher Scientific, USA
Sigma protector RNase inhibitor	3335399001	Millipore Sigma, USA
Single index kit N set A, 96 reactions	PN-1000212	10x Genomics, USA
SPRIselect reagent kit	B23318	Beckman Coulter, USA
ß-mercaptoethanol	63689	Sigma-Aldrich, Germany
Tagment DNA buffer (TD-buffer; 2x)	15027866	Illumina, USA
Tagment DNA enzyme (Tn5 transposase)	15028212	Illumina, USA
Tapestri single-cell cartridge kit	MB02-0001	Mission Bio, USA
Tapestri single-cell DNA kit v1	MB03-0017	Mission Bio, USA
Tapestri single-cell DNA myeloid kit v2	MB03-0017	Mission Bio, USA
Triton X-100	1.08643.1000	Merck Sigma/Aldrich, Germany
Venor GeM advance kit	11-7024	Minerva Biolabs, Germany
Wizard SV gel and PCR clean-up system	A9281	Promega, USA

1.3 Instruments

Table 4 Instruments

Instrument	Company
2100 Bioanalyzer instrument	Agilent Technologies, USA
ChemiDoc MP imaging system	Bio-Rad, Germany
E-Gel Safe Imager	Invitrogen, USA
HiSeq 4000 sequencing system	Illumina, USA
LUNA automated cell counter	Logos Biosystems, South Korea
Mission Bio Tapestri platform	Mission Bio, USA
NovaSeq 6000 system	Illumina, USA
Qubit 2.0 fluorometer	Thermo Fisher Scientific, USA
StepOnePlus real-time PCR system	Thermo Fisher Scientific, USA
T100 thermo cycler	Bio-Rad, Germany
Tapestation 4200	Agilent Technologies, USA
NanoDrop One	Thermo Fisher Scientific, USA

1.4 Software

Table 5 Software

Software	Version	Source
Affinity Designer	1.10.4	Serif (Europe) Ltd., UK
BED Tools	2.25.0	Quinlan & Hall, 2010 [232]
Biorender	last accessed: March 2022	Biorender, USA [233]
DiffBind	2.12.0	Ross-Innes et al., 2012 [234]
FastQC	0.11.9	Andrews, 2010 [235]
IGV Tools	2.3.23	Robinson et al., 2011 [236]
Integrative Genomics Viewer	2.6.2	Robinson et al., 2011 [236]
Macs2	2.1.2	Zhang et al., 2008 [237]
MultiQC	1.7	Ewels et al., 2016 [238]
RStudio	1.0.153	RStudio, USA
SAMtools	1.3	Li et al., 2009 [239]
Tapestri Insights	2.1	Mission Bio, USA
Trimmomatic	0.36	Bolger et al., 2014 [240]

1.5 R packages

Table 6 R and Python packages

Package	Source
Bioconductor	Gentleman RC et al., 2004 [241]
Complex heatmap	Gu et al., 2016 [242]
Copykat	Gao et al., 2021 [243]
Cytoscape	Gustavsen et al., 2019 [244]
Data.table	Dowle et al., 2021 [245]
Dorothea	Garcia-Alonso et al., 2019 [246]
Escape	Borcherding et al., 2021 [247]
Ggplot2	Hadley Wickham, 2016 [248]
Ggpointdensity	Kremer et al., 2019 [249]
Harmony	Korsunsky et al., 2019 [250]
Hyperr	Federico et al., 2020 [251]
Numpy	Harris et al., 2020 [252]
Paga	Wolf et al., 2019 [253]
Pandas	Pandas Development Team, 2020 [254]
Scanpy	Wolf et al., 2018 [255]
Scrublet	Wolock et al., 2019 [256]
Sctransform	Hafemeister et al., 2019 [257]
Scvelo	Bergen et al., 2020 [258]
Seurat	Satija et al., 2015[259]
Singler	Aran et al., 2019 [260]
Slingshot	Street et al., 2018 [261]
Velocyto.R	La Manno et al., 2018 [262]
Viridis	Garnier et al, 2021 [263]

1.6 Patient samples

Written informed permission from all patients in this study was obtained from clinical partners. Cells of *MLL-X* patients were kindly provided by Prof. Dr. Michael Lübbert (Department of Internal Medicine I, University of Freiburg). These cells were collected from peripheral blood or bone marrow and enriched for mononuclear cells (MNCs) via Ficoll-Hypaque. *EDC4-MLL* fusion cells were additionally depleted from $CD3^+$ cells using autoMACS (Miltenyi Biotec), as described previously [264]. Cells of *FLT3-ITD* or *IDH1mut* patients were kindly provided from Prof. Dr. Konstanze Döhner and Prof. Dr. Hartmut Döhner (Department

of Internal Medicine III, University Hospital Ulm). These samples were obtained from bone marrow aspirates from of the respective patients.

Table 7 Overview of *MLL*-X patients

Patient	1	2	3	4
Aberrations	*MLL-EDC4*	*MLLT3-MLL*	*MLLT3-MLL*	*MLL-ELL*
Sex/age	F/56	F/28	F/58	F/57
AML type	Secondary AML evolving from MDS (no prior CTx)	T-AML (after RCTx for Hodgkins lymphoma), no MDS phase	*De-novo* AML, no MDS phase	T-AML (after RCTx for ovarian cancer), no MDS phase
Prior AML treatment	Dezitabine (DAC)	-	-	-
Timepoint	Before 4th cycle of DAC, 7 weeks no treatment	Diagnosis	Diagnosis	Diagnosis
Karyotype	46, XX, t(11;16)(q23;q12)	46, XX, t(9;11)(p22;q23)	46, XX, t(9;11)(p22;q23)	46,XX, t(11;19)(q23;p13)
% of fusion cells	85%	90%	95%	82%
Cell source	PBMC CD3 depleted	BM MNC	BM MNC	PBMC
% blasts (% nuclei with MLL fusion detected by FISH)	90 (85)	78 (90)	95 (95)	57 (82)
scRNA-seq #cells	4 974	1 928	3 906	5 502
scRNASeq median genes/cell	951	624	856	1 225

Table 8 Overview of *FLT3*-ITD *and IDH1^{mut}* patients

Patient	*FLT3*-ITD AML-1		*FLT3*-ITD AML-2			*IDH1^{mut}* AML		
Timepoint	Diagnosis	Relapse	Diagnosis	Remission	Relapse	Diagnosis	Remission	Relapse
Treatment	Untreated	Midostaurin	Untreated	Midostaurin	Midostaurin	Untreated	AG-120	AG-120
scRNA # cells # filtered cells # median genes/cell	7 601 3855 1 960	1 244 413 2 649	5 445 4 918 3 589	5 172 2 700 1 681	8 315 7 011 2 969	1 194 485 387	3 610 2 673 991	2 678 1712 1350
scATAC # cells # filtered cells # median fragments/cell	1 709 1 493 6 987	3 970 3 647 4 610	2 911 2308 6 828	212 55 6 688	2 329 1 552 7 593	1 122 505 6 234	1 746 782 3 652	--
scDNA #cells #toal reads #reads/amplicon/ cell	1 169 28M 36	3370 35M 30	686 42M 124	3 384 110M 18	2 767 117M 13	765 32M 110	4 327 62M 37	1 275 85M 19

2 Oligonucleotide sequences

Table 9 Sequencing primers

Oligo name	Sequence (5´→ 3´)	Application	Reference
1	GACTTCTT	scDNA-seq, myeloid V1	Mission Bio, USA
2	TTATTCTT	scDNA-seq, myeloid V1	Mission Bio, USA
3	CGCGGCTT	scDNA-seq, myeloid V1	Mission Bio, USA
4	AGGAGCTT	scDNA-seq, myeloid V1	Mission Bio, USA
5	CCTTCATT	scDNA-seq, myeloid V1	Mission Bio, USA
6	CAGCTCGT	scDNA-seq, myeloid V1	Mission Bio, USA
7	TAGGACGG	scDNA-seq, myeloid V1	Mission Bio, USA
8	TTCCTAGG	scDNA-seq, myeloid V1	Mission Bio, USA
Nextera index PCR primers Index 1 read (custom)	CAAGCAGAAGACGGCATACGAGAT[i7]GTCTCGTGGGCTCGG	ATAC-seq	Adapted from Illumina, USA
Nextera index PCR primers Index 2 read (custom)	AATGATACGGCGACCACCGAGATCTACAC[i5]TCGTCGGCAGCGTC	ATAC-seq	Adapted from Illumina, USA
Nextera transposase adapter read 1	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG	ATAC-seq	Illumina, USA
Nextera transposase adapter read 2	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG	ATAC-seq	Illumina, USA
IDH2 forward primer	AATTTTAGGACCCCCGTCTG	Sanger sequencing	Merck, Germany
IDH2 reverse primer	TGTGGCCTTGTACTGCAGAG	Sanger sequencing	Merck, Germany

3 Experimental procedures

3.1 Cell culture, treatment and sample preparation

3.1.1 Cell lines and cell culture material

Table 10 AML cell lines

Cell line	RRID	Growth medium	Reference
TF-1	CVCL 0559	Gibco RPMI 1640 ATCC	Kitamura et al., 1989 [265]
TF-1 IDH2 R140mut	CVCL 0559	Gibco RPMI 1640 ATCC	Kitamura et al., 1989 [265]

Table 11 Cell culture material

Medium / supplement	Company
Gibco RPMI 1640 (ATCC Modification), 500mL	Thermo Fisher Scientific, USA
1x Penicillin-streptomycin, 10,000 U/mL penicillin, 10 mg/mL streptomycin	PAN-Biotech, Germany
Fetal bovine serum	PAN-Biotech, Germany
Human GM-CSF cecombinant protein	Thermo Fisher Scientific, USA
Recovery™ cell culture freezing medium	Thermo Fisher Scientific, USA

3.1.2 Cell culture of TF-1 cells

Cells were cultured in ATCC modified Gibco RPMI 1640 medium (Table 11). For cultivation, TF-1 cells (Table 10) were seeded at a density of 1-4 x10^5 cells/mL in T-25 flasks suitable for suspension cells. Cells were split alternating every second or third day or at densities over 1x10^6 cells/mL. Treatment with Enasidenib was conducted for 6 days at 37°C and 5% CO_2. Enasidenib was freshly diluted with dimethyl sulfoxide (DMSO) in a serial dilution from a 10 mM stock. For control conditions, DMSO was added to the media in equal volumes. Incubation with Enasidenib was refreshed by medium exchange after 3 days, ensuring constant drug concentration. For this, cells were spun down, old media was discarded and fresh media containing the drug or DMSO alone was added. Cells were counted in combination with trypan blue using a Luna cell counter and directly processed or frozen in 1 mL freezing medium (Table 11).

3.1.3 Freezing of viable cells

Viable cells were frozen by resuspension of the cell pellets containing a minimum of $1x10^6$ cells. The commercially available recovery cell culture freezing medium (Table 11) was used. The cell suspension was carefully homogenized by pipetting up and down 10 times and gently chilled to -80 °C overnight before being transferred to -120°C.

3.1.4 Mycoplasma test

Venor GeM advance kit was used to test cell lines for mycoplasma contamination on a regular basis, according to the manufacturer's instructions. For this, 500 µl of confluent cell culture medium was transferred and heated to 95°C for 10 minutes. The sample was centrifuged at 13,000 rpm for 10 minutes at room temperature, and the supernatant was moved to a new tube. 2 µl of sample and 23 µl of rehydration buffer were added to the PCR tube which was included in the kit. For positive controls, 25 µl of rehydration buffer were added. PCR was carried out according to manufacturer's instructions, then 5 µl of each sample were put onto a E-Gel (2%) and run twice for 8 minutes using the E-Gel safe imager. ChemiDoc XRS+ system was used to examine the gel.

3.1.5 Luminescent cell viability assay

1 200 cells were counted via Luna cell counter and seeded into sterile 96 well-corning plates in 100 µl in medium in total. One row of wells was reserved for medium only accounting for background signal. Serial dilution of Enasidenib was performed, starting with a concentration of 50 µM into row 11, yielding concentrations from 0.1 µM – 25 µM. Next, cells were incubated for six days at 5% CO_2 at 37°C. Media was exchanged after three days, ensuring constant treatment concentrations. DMSO was present in identical levels in all treatment dosages, and the maximal DMSO content in the medium was less than 0.1%. Cell viability was assessed by measuring luminescence generated due to ATP release using cell-titer glo. The manufacturer's instructions were followed accordingly. Before measurement, cells were transferred into a white opaque flat bottom 96 well plate and equilibrated at room temperature for 20 minutes. The lysis reagent was made from a 1:1 solution of complete

medium and equilibrated at room temperature for 30 minutes. The plate's supernatant was discarded, and 200 µL of lysis medium mixture was administered to each well. A plate shaker was used to agitate the plate for 6 minutes at 600 rpm. Wells were resuspended with a multichannel pipette with extreme caution to avoid forming bubbles. Luminescence was measured using a Tecan plate reader via the luciferase program with an integration time of 1ms. After averaging each condition, the background signal was removed, and each cell line's signal was standardized to DMSO controls.

3.1.6 Sanger sequencing of TF-1 cells

To ensure presence of *IDH2wt* and *IDH2mut* in TF-1wt and TF-1 *IDH2mut*, respectively, sanger sequencing of the IDH2 amplicon was performed. Therefore, $1x10^5$ were collected and spun down for 5 minutes at 10 000 rpm. Media was flicked off and 200 µl direct PCR buffer was added to the samples and resuspended. 1/100 of 0.2 mg/mL proteinase K was then added to break proteins and incubated overnight on a heat block at 55°C. The next day proteinase K was inactivated by heating the mixture at 85°C for 45 minutes. Using a Nanodrop concentration was measured. Subsequently, a PCR was performed with 2 µl of the mixture. Annealing temperature of 65°C was calculated based on the NEB Q5 Tm calculator.

Table 12 Reagents for direct PCR

Reagent	Volume
10 ng gDNA	x µl
25 mM dNTP's	0.4 µl
5x Q5 reaction buffer	10 µl
10 µM forward primer	2.5 µl
10 µM reverse primer	2.5 µl
Q5 Polymerase	0.5 µl
H_20	x µl
Total	50 µl

Table 13 PCR for sanger sequencing

Direct PCR	Temperature	Time	Cycles
Initial denaturation	98°C	30 sec	1x
Denaturation	98°C	10 sec	30x
Annealing	65°C	30 sec	
Extension	72°C	20 sec	
Final extension	72°C	2 min	1x
Hold	4°C	∞	1x

PCR products were then cleaned up using Wizard SV gel and PCR clean-up system. Subsequently the samples and the fast DNA ladder were transferred onto a 2% E-Gel and run twice for 8 minutes using the E-Gel safe imager. ChemiDoc XRS + system was used to examine the gel and the expected band at a size of 357 bp. The PCR products were then diluted with nuclease free water to a concentration of 7.5 ng/µl into two fresh 1.5 mL tubes. The tubes were then labeled with pre-bought barcodes for Sanger sequencing at Microsynth. Together with a tube of 4 µM IDH2 primer the amplicons were shipped to Microsynth and sequenced overnight. Sequencing data was then examined in IGV browser track.

3.2 ATAC-seq

ATAC-seq of AML patients treated *ex vivo* with BAY-1436032 was performed using 50 000 cryopreseved cells per technical replicate. After thawing a frozen vial, technical duplicates were created by dividing cells and performing ATAC-seq sequencing library preparation in parallel. Cells were spun down at 500 g for 5 minutes at 4°C and supernatant was discarded. Cells were lysed with a tagmentation mix consisting of 9.75 µl H_2O, 12.50 µl 2x transposase buffer, 0.50 µl protease inhibitor cocktail, 2 µl Tag DNA enzyme and 0.25 µl of 1% digitonin. After carefully resuspending the cell pellet with the tagmentation mix, samples were incubated for 30 minutes at 37°C and the reaction was then halted by transferring the samples on ice. The tagmented DNA was purified using Qiagen MinElute kit and were eluted in 12 µl of EB-buffer. Next, a qPCR was performed to obtain optimal cycle numbers for each sample. For the enrichment PCR 2.9 µl nuclease-free water, 2.5 µl each of custom Nextera index PCR primers 1 and 2, 25 µl NEBnext HF 2x ready master mix and 10 µl of tagmented sample was added. Following, libraries were purified using AMPure beads with a ratio of

1:1.6 and samples were eluted in 50 µl EB-buffer. Then a right sided AMPure beads selection with 1:0.5/1.4 was performed to get rid of high molecular peaks. Samples were then eluted in 15 µl of EB-buffer. DNA was quantified via Qubit 2.0 fluorometer and the Qubit dsDNA HS assay kit. The mean peak sizes were assessed via TapeStation system. Samples were pooled at equimolar concentrations and sequenced at the DKFZ Genomics Core Facility.

Table 14 Reagents for ATAC tagmentation mix

Reagent	Volume
H_2O	9.75 µl
2x Transposase buffer	12.50 µl
Protease inhibitor cocktail	0.50 µl
Tag DNA enzyme	2.00 µl
1% Digitonin	0.25 µl
Total	25 µl

Table 15 Reagents for ATAC qPCR

Reagent	Volume
H_2O	2.9 µl
Index primer 1	0.5 µl
Index primer 2	0.5 µl
100x SYBR green	0.1 µl
NEBnext HF 2x ready MM	5 µl
Tagmented sample	1 µl
Total	10 µl

Table 16 Settings for ATAC qPCR

qPCR	Temperature	Time	Cycles
Initial annealing	72°C	5 min	1x
Initial denaturation	98°C	30 sec	1x
Denaturation	98°C	10 sec	25x
Annealing	63°C	30 sec	
Extension	72°C	1 min	
Hold	4°C	∞	1x

Table 17 Reagents for ATAC enrichment PCR

Reagent	Volume
H_2O	10 µl
Index primer 1	2.5 µl
Index primer 2	2.5 µl
NEBnext HF 2x ready MM	25 µl
Tagmented sample	10 µl
Total	50 µl

Table 18 Settings for ATAC Enrichment PCR

Enrichment PCR	Temperature	Time	Cycles
Initial annealing	72°C	5 min	1x
Initial denaturation	98°C	30 sec	1x
Denaturation	98°C	10 sec	according to qPCR results
Annealing	63°C	30 sec	
Extension	72°C	1 min	
Final extension	72°C	1 min	1x
Hold	4°C	∞	1x

3.3 scRNA-seq of primary AML samples

The experiment was conducted according to the manufacturer's standard procedure for Chromium single cell 3' reagent kits v2 or v3. Frozen viable cells were thawed for 30 seconds in a water bath with a temperature of 37°C. Cells were then washed twice with 1xPBS and counted using the Luna cell counter. Cell numbers were chosen based on the manual to recover around 8 000 cells per condition and/or patient. A master mix consisting of 50 µl of RT reagent mix, 3.8 µl of reverse transcriptase (RT) primer, 2.4 µl of additive A and 10.0 µl of RT was generated and added to the samples resulting in a total of 90 µl. 90 µl of the cell/master mix suspensions were then added to the according wells of the 10x chips. A 40 µl solution of single cell 3' gel beads was vortexed for 30 seconds before being transferred to the respecting bead row. In the remaining row partitioning oil in a volume of 270 µl was added and then the rubber gasket was fixated on the chip. With the default settings, the chromium controller was run for approximately seven minutes. The chip holder was

positioned at a steep angle, and 100 µl of GEMs were transferred into a PCR tube, followed by GEM RT incubation. Subsequently, 125 µl of recovery agent was gradually added. The red agent-oil mixture was taken from the bottom of the reaction tube after one minute. After vortexing the DynaBeads for 30 seconds, 200 µl of water was added to the remaining samples in the PCR tubes. The samples were carefully pipetted up and down several times before a 10 minutes incubation at room temperature. The first elution solution was prepared by mixing 98 µl of EB buffer, 1 µl additive A and 1 µl of 10% tween-20. Supernatant was discarded after placing the samples on the 10x magnetic rack. Then the beads were thoroughly washed with 80% ethanol. The beads were air dried shortly, then tubes were removed from the magnet and DNA was eluted from the magnetic beads using 35.5 µl elution solution. After an incubation time of 2 minutes, the purified sample could be transferred into fresh tubes.

Table 19 Settings for GEM RT incubation

GEM RT Incubation	Temperature	Time
Step 1	53°C	45 min
Step 2	85°C	5 min
Hold	4°C	∞

Amplification of cDNA was carried out by pipetting 8 µl of nuclease free water, 50 µl of amplification master mix, 5 µl of cDNA additive, and 2 µl of cDNA primer mixed together with 35 µl purified GEM-RT product. 13 cycles were used to amplify the PCR products.

Table 20 cDNA amplification

PCR	Temperature	Time	Cycles
Initial denaturation	98°C	3 min	1x
Denaturation	98°C	15 sec	13x
Annealing	65°C	20 sec	
Extension	72°C	1 min	
Final extension	72°C	1 min	1x
Hold	4°C	∞	1x

The PCR products were cleaned-up using AMPure beads with a 0.6x ratio end eluted in 40 µl. Qubit dsDNA HS assay kit was used to quantify the concentration of purified GEM-RT products, and TapeStation D5000 ScreenTape was used to assess mean peak sizes. For library preparation, a fragmentation mix was produced by mixing 10 µl of fragment enzyme together with 5 µl fragmentation buffer. This mix was added to the cleaned-up GEM-RT product and end repair and A-tailing was conducted. Afterwards two AMPure bead size selections with a ratio of 0.6x and 0.8x, respectively, were performed. The samples were eluted in 50 µl EB buffer. 20 µl of ligation buffer, 10 µl of DNA ligase, and 2.5 µl of adaptor mix were combined and then added to 50 µl of sample. The samples were then incubated at 20°C for 15 minutes before being purified using 0.8x AMPure beads. Then samples were again eluted in 30 µl of EB buffer.

Table 21 Settings for end repair and A-tailing

End repair and A-tailing	Temperature	Time
Pre-cool block	4°C	∞
End repair	32°C	5 min
A-tailing	65°C	30 min
Hold	4°C	∞

For indexing, 2µl of SI-PCR primer, 8 µl nuclease free water and 50 µl amplification master mix were spiked to the 30 µl of cleaned-up samples. In addition, 10 µl of a distinct chromium i7 sample index was added ensuring successful multiplexing. The index PCR was run with 10 cycles and followed by 0.9x AMPure beads purification. After eluting in 35 µl of EB buffer, the concentration of the scRNA libraries were quantified via Qubit dsDNA HS assay kit and mean peak sizes were assessed by TapeStation D5000 ScreenTape. Samples with fragment sizes over 650 bp were purified once more using AMPure beads with a 0.9x ratio. The libraries were then pooled equimolar and sequenced by the DKFZ Genomic Core Facility.

Table 22 Sample index PCR

PCR	Temperature	Time	Cycles
Initial denaturation	98°C	45 sec	1x
Denaturation	98°C	20 sec	10x
Annealing	54°C	30 sec	
Extension	72°C	20 sec	
Final extension	72°C	1 min	1x
Hold	4°C	∞	1x

3.4 scATAC-seq of primary samples and cell lines

The experiment was conducted based on manufacturer's instructions. For primary samples, frozen viable cells were thawed for 30 seconds in a water bath with a temperature of 37°C. Cells were then washed twice with 1x PBS and counted using the Luna cell counter. For TF-1 cell lines, cells were harvested after 6 days of treatment with 5 μM of Enasidenib or concentration matched DMSO. After cell count, roughly a million cells were pipetted into a new 2 mL tube, topped up to 1 mL of volume with 1x PBS and then centrifuged at 300 g for 5 minutes at 4°C. Supernatant was removed and cells were resuspended in 100 μl chilled lysis buffer. Samples were then incubated on ice for 5 minutes prior adding 1 mL of cold 1x wash buffer and centrifuging at 4°C for 5 minutes at 500 g. Supernatant was then discarded, and nuclei were counted using the Luna cell counter. The needed volume of nuclei suspension for an output of approximately 8000 cells was calculated for each sample based on the nuclei stock concentration table provided by 10x Genomics. 5 μl of the stock concentrations were then used to proceed with the scATAC protocol.

Table 23 scATAC-seq buffers for nuclei isolation

Buffer	Component	Stock	Final	Volume
Diluted nuclei buffer	Nuclei buffer Nuclease-free H_20	20x	1x	50 µl 90 µl
Wash buffer	Tris-HCl (pH7.4) $NaCl_2$ $MgCl_2$ BSA Tween-20 Nuclease-free H_2O	1 M 5 M 1 M 10% 10%	10 mM 10 mM 3 mM 1% 0.1%	100 µl 20 µl 30 µl 1 mL 100 µl 0.75 mL
Lysis buffer	Tris-HCl (pH7.4) $NaCl_2$ $MgCl_2$ Tween-20 NP-40 Digitonin BSA Nuclease-free H_2O	1 M 5 M 1 M 10% 10% 5% 10%	10 mM 10 mM 3 mM 0.1% 0.1% 0.01% 1%	50 µl 10 µl 15 µl 50 µl 50 µl 10 µl 500 µl 4.315 mL

By combining 7µl of ATAC buffer with 3µl of ATAC enzyme on ice, a transposition mixture was made. Based on the sample stock concentration, 1 µl of diluted nuclei buffer was added to 4 µl of nuclei suspension and 10 µl of transposition mix. The mixes were combined and incubated for 60 minutes in a thermocycler at 37°C with a lid temperature of 50°C. The master mix was made up of 61.5 µl of barcoding reagent, 1.5 µl of reducing agent B, and 2 µl of barcoding enzyme. 65 µl of the master mix was added put to 15 µl transposed nuclei and this mixture was then transferred into row 1 of the chromium chip E. 40 µl of the gel beads were pipetted to row 2 after being vortexed for 30 seconds and spun down. Then 240 µl partitioning oil was transferred to the remaining row and followingly the rubber gasket was fixed onto the chip. After insertion of the chip into the 10x controller, the scATAC program was initiated and completed in 7 minutes. The chip was then positioned at a 45° angle and 100 µl of GEMs were aspirated using a multichannel pipette and transferred into fresh PCR strips. Then amplification of the GEMs was conducted.

Table 24 Settings for scATAC enrichment PCR

PCR	Temperature	Time	Cycles
Initial annealing	72°C	5 min	1x
Initial denaturation	98°C	30 sec	1x
Denaturation	98°C	10 sec	13x
Annealing	59°C	30 sec	
Extension	72°C	1 min	
Hold	4°C	∞	1x

Each sample was diluted with 125 µl recovery agent and the tube was inverted fifteen times. Samples showed a separation of a hydrophobic and hydrophilic layer. The bottom 125 µl which represented the recovery agent were carefully removed. Into each sample, 200 µl of cleanup mix consisting of 182 µl cleanup buffer, 8 µl dynabeads MyOne SILANE and 5 µl reducing agent B was pipetted. At room temperature, the samples were incubated for 10 minutes. The samples were then put on a magnet, supernatant was carefully discarded, and beads were washed twice with 80% ethanol. Subsequently 40.5 µl elution buffer comprising of 98 µl of EB buffer, 1 µl 10% tween-20 and 1 µl reducing agent B was pipetted to the beads and the mixture was set aside for 2 minutes at RT. Then the samples were returned to the magnet and 40 µl of supernatant were moved to new tubes. To the purified samples 48 µl of SPRIselect reagent was added and incubated for 5 minutes at room temperature. Samples were placed on the magnet and supernatant was removed. The beads were thoroughly washed with fresh 80% ethanol. DNA was then eluted from the magnetic beads in 40.5 µl EB-buffer and incubated for 2 minutes. Next 40 µl of purified sample was transferred to a fresh tube. Samples were kept overnight at 4°C and library preparation was proceeded the next day.

Table 25 Settings for scATAC library PCR

Library PCR	Temperature	Time	Cycles
Initial denaturation	98°C	45 sec	1x
Denaturation	98°C	20 sec	12x
Annealing	67°C	30 sec	
Extension	72°C	20 sec	
Final extension	72°C	1 min	1x
Hold	4°C	∞	1x

For library preparation, 50 µl AMP mix and 7.5 µl SI-PCR primer B were mixed. 57.5 µl of this mixture was transferred to 40 µl of purified sample. Then 2.5 µl sample specific chromium i7 sample index N was spiked in individually. The library was generated running a library specific thermocycler program. After that 40 µl of SPRIselect reagents were added and a 5 minutes incubation at room temperature was conducted. The supernatant was moved into new PCR tubes and 74 µl of SPRIselect reagent were added to the supernatant while the beads from the first tube were discarded. The samples in the fresh tubes were then put on a magnetic rack and incubated for 5 minutes. Then supernatant was discarded and beads were washed twice with 80% ethanol. After air-drying, the beads were eluted with 20.5 µl EB-buffer. Samples were put back on the magnet and after 2 minutes 20 µl of eluted DNA was transferred to a new tube. The library quality and size distribution were assessed by TapeStation D5000 ScreenTape and concentration were measured via Qubit dsDNA HS Assay kit. Then, all libraries were pooled in equimolar concentrations and sequenced by the DKFZ Genomics Core Facility.

3.5 Targeted single cell DNA-seq

Both the Tapestri Single-Cell DNA kit V1 and V2 and the Mission Bio myeloid panel were used to perform targeted single-cell sequencing on the Mission Bio Tapestri Platform. The myeloid panel covered 45 genes with 330 amplicons for V1 and 312 amplicons for V2 (amplicon length range: 375-550 bp). The experiment was conducted based on manufacturer's instructions. Frozen viable cells were thawed in a water bath with 37°C. Then the cell suspension was washed twice with 1x PBS and centrifuged at 400 g for 5 minutes at 4°C. The cell pellet was then resuspended in 40-100 µl cell buffer that was provided by the kit. Cells were counted using the Luna cell counter and cell suspension was further diluted to a concentration of 2000-4000 cells/µl. Then 35 µl of the diluted cell mix were loaded on the cartridge of the Tapestri platform. For library preparation cells were encapsulated, lysed and barcoded. After cleavage of the barcoding primers using a UV light for 8 minutes, the target PCR amplification was started. For V1 a UV lamp from CellenONE (Cellenion, France) was used, for V2 the UV light was implemented in a newer version of the Mission Bio Tapestri platform. The PCR products were purified and then sequencing adaptors were added individually in a second library PCR. The size distribution of the libraries were determined via High Sensitivity D5000 ScreenTapes using a TapeStation. Samples that did not show the desired library profiles were processed several times with different approaches (see Results). DNA was measured via Qubit 2.0 fluorometer and the Qubit dsDNA HS assay kit. The samples were then pooled equimolar and submitted to sequence. Libraries generated from the V1 kit were sequenced at the Genomics Core Facility at EMBL whereas libraries generated from V2 were sequenced by the DKFZ Genomic Core Facility.

3.6 Sequencing

Table 26 Sequencing of AML samples

Samples	Experiment	Sequencing type	Instrument model
AML samples treated *ex vivo* with BAY1436032	ATAC-seq	75 bp PE	NextSeq 550 (Illumina)
Primary AML samples (*IDH1*, *FLT3*-ITD)	scDNA-seq	150 bp PE	NovaSeq 6000 (Illumina)
Primary AML samples (*IDH1*, *FLT3*-ITD)	scDNA-seq	150 bp PE	
Primary AML samples (*IDH1*, *FLT3*-ITD)	scATAC-seq	49+8+16+49 bp PE	NovaSeq 6000 (Illumina)
Primary AML samples (*IDH1*, *FLT3*-ITD)	scRNA-seq	28+94 bp PE	NovaSeq 6000 (Illumina)
Primary AML samples (*MLL*)	scRNA-seq	28+94 bp PE	NovaSeq 6000 (Illumina)
TF1-cell line	scDNA-seq	150 bp PE	NovaSeq 6000 (Illumina)
TF-1 cell line	scATAC-seq	50+8+16+50 bp PE	NovaSeq 6000 (Illumina)

3.7 Data analysis

3.7.1 Analysis of bulk ATAC samples

Dr. Lara Klett oversaw development and maintenance the ATAC pipeline. Trimmomatic was used to remove adaptor sequences from demultiplexed reads, and Bowtie2 was used to map the remaining reads using indices for the human genome hg38. Samtools sorted, indexed, and excluded PCR duplicates from the mapped reads. BEDTools was used to eliminate reads mapped into blacklisted areas, and Samtools was used to apply a quality threshold. At the positive strand, each read was shifted by four basepairs, and at the negative strand, it was shifted by five. This was necessary since the transposase's integration site was a few nucleotides distant from the eventually discovered fragment start point. The bam files were then converted to a bed format. Reads mapping to the mitochondrial genome were deleted with the bash code *awk*. The filtered reads were utilized for peak calling using MACS2 and visualization via IGVtools. FastQC, MultiQC, NSC, RSC, and FRiP score computations were used to assess quality at various stages.

3.7.2 Annotation of scRNA-seq and scATAC-seq with CellRanger

Demultiplexing and alignment to the human genome hg38 of scRNA-seq and scATAC-seq raw data was performed using Cellranger, a tool provided by 10x Genomics. It was used with default settings and the parameter *--jobmode=torque* was chosen for quicker analysis.

3.7.3 Quality control of scRNA-seq data

Raw expression data was transferred into R version 4.0.2 and analyzed via Seurat [259] following the recommended parameters. Single-cell profiles with fewer than 500 identified genes (showing dying cells or no cell in a droplet), over 3 000 detected genes (indicating cell doublets), or over 15% of UMIs originating from mitochondrial reads were not considered for further analysis. The following settings were used to eliminate cells having a doublet score >0.4 obtained using the Python program Scrublet [256] : *sim doublet ratio = 2; n neighbors = 30; expected doublet rate = 0.1*. Raw-counts from scRNA-seq bone marrow data from eight healthy persons were acquired through the HCA portal [266] for comparison of leukemic cells with healthy hematopoietic progenitors. The HCA dataset's gene symbols were translated from GENCODE v27 to v28 for compatibility and the 3' reagent v2 chemistry was used.

3.7.4 Cell type annotation of microenvironment based on marker expression profiles

After determining the number of principal components used for downstream clustering using *ElbowPlot*, UMI counts were normalized via *single cell transform.* The percentage of mitochondrial reads per cell and the number of UMIs per cell were regressed out as part of the standard Seurat workflow. Then unsupervised dimensionality reduction using UMAP technique was performed to integrate the transcriptomes . Genes, that were differentially expressed, were identified using Wilcoxon Rank Sum test ($p_{adj} < 0.05$, log2FoldChange > 0.1) via the command *FindMarkers* implemented by Seurat. In the scRNA-seq data, cell type specific marker genes such as MS4A1, HBB, CD14, NKG7 and CD3D were evidently identifiable, allowing for a robust marker-based assignment of non-malignant cell types.

3.7.5 Annotation of leukemic cells

Leukemic cells of patient samples harboring a *FLT3*-ITD could be detected by their overexpression of FLT3, making FLT3 expression a robust tumor marker in these entities. For other AML subtypes cell clusters with unusual transcriptome signatures that could not be attributed to the microenvironment with confidence were defined as leukemic cells. Additionally, healthy microenvironment within different patients and timepoints clustered together, whereas malignant cells did not. Aneuploidy of AML cells was validated for each sample individually using the R package copyKat [243] using “Monocytes” and “T-cells/NK-cells” as diploid reference. The following settings were used to infer ploidy: *n.gene.chr=5, win.size=25, KS.cut=0.2, distance=”euclidean”*.

3.7.6 Cell type prediction to infer differentiation state of leukemic cells

Utilizing a down-sampled HCA data set (1 000 cells per cell type) and associated cell identity labels as training data set, cell type prediction of leukemic cells was determined using automated cell type annotation through the SingleR [260] tool. Specifically for each test cell; (i) Spearman correlation between the cell’s expression profile and that of each reference sample was calculated. This was done across the union of marker genes identified between all annotated labels. (ii) A per-label score was defined as fixed quantile of the distribution of correlations. (iii) Steps i and ii were repeated for all labels until the cell was annotated with the highest scoring label. Default settings were used. As a verification, module scores for six cell types along the HSC to myeloid trajectory, namely HSC-, GMP, MPP-, Monocytes, preDC- and cDC2 genes, were calculated. This was conducted using *AddModuleScore* from Seurat and a signature gene list which was derived from the HCA data set (see Appendix).

3.7.7 Gene set enrichment analysis

For gene set enrichment analysis Hallmark, Biocarta, KEGG, Reactome, GO:BP, GO:CC and GO:MF sets were downloaded with the R package msigdbr and the command *msigdb gsets.* The gene signatures from the distinct AML clusters and the listed gene sets were then used for downstream analysis, either by using the R packages hypeR [251] or escape [247]. For hypeR a hyp object was generated using the *hypeR* command and top enriched gene sets were

plotted using *hyp dots*. For escape a data frame was produced using the *enrichIt* command with the parameter *group=1000*. Visualization was conducted with complex heatmap [242].

3.7.8 Inference of transcription factor activity

Transcription factor activity was inferred from scRNA-seq data by using Bioconductor's DoRothEA package [246]. DoRothEA is a comprehensive resource that includes a well curated list of TFs and their transcriptional targets. A regulon is a collection of genes controlled by a particular transcription factor. DoRothEA regulons were coupled with the statistical method VIPER [267] and the TF activities were calculated based on the target's mRNA expression. As a result, TF activity was used as a proxy for a certain transcriptional states. This program classifies TF-target interactions in five confidence levels ranging from A to E, with A having the highest confidence. Human regulons were downloaded using *get(data("dorothea hs", package = "dorothea"))*. For my analysis only confidence levels A and B were chosen using the command *filter(confidence %in% c("A","B"))*. A list of the top 30 TFs depicted on the heatmaps that were shown to be regulated by each other were then imported to Cytoscape [268] for a network based visualization.

3.7.9 Trajectory inference

To determine the pattern of dynamic process within AML cells and subsequent arrangement of leukemic cells based on their differentiation state different packages were tested.(i) Velocyto [262] generates a trajectory by estimating spliced and unspliced counts by enumerating reads that incorporate within intronic sequence. To keep the metadata of predicted cells using singleR, a Seurat object was created in R from the python derived loom file. This object was then subsetted (e.g. unspliced cells were removed) so that singleR metadata could be added to the Seurat object. The object was then saved in .h5ad format using the commands *saveH5Seurat* and *Convert(Seurat object, overwrite=TRUE, dest="h5ad")*. This converted file and the original loom file were then read into python as annotated data matrices and merged. Filtering, normalization and velocity computing was then performed according to standard workflow. (ii) Partition-based graph abstraction (PAGA)[253] estimates connectivity of manifold partitions while preserving the global topology of data. For analysis, the saved h5ad embedding generated prior with velocyto was used.

Denoising, clustering and recomputing the embedding using PAGA initialization were conducted using the default settings.

(iii) Slingshot [261] accepts cluster labels as input and creates a minimal spanning tree to organize these clusters into lineages. Paths across the tree were smoothed by simultaneous principal curve fitting. The pseudotime value of a cell was derived by its projection onto one or more of these curves. The option to determine initial or terminal clusters was omitted for the analysis, otherwise default settings were used.

3.7.10 Integrating primary AML samples with healthy bone marrow controls

For the 8 healthy bone marrow samples the *SCTransform*-based workflow implemented in Seurat was chosen and run with the default parameters. For integrating individual AML samples with a healthy bone morrow control from the HCA dataset, different approaches were tested. (i) Integration was performed using harmony according to standard settings with *lambda=0.05*. (ii) For integration with the R package FNN embeddings both from the healthy control as well as from the AML sample were extracted. Then nearest neighbors were calculated using the command *get.knnx(embedding1, embedding2, k=1, algorithm="kd tree)*.

The resulting column of the nearest neighbor index *nn.index* was extracted as a data frame with the column name *position*. The knn data frame and the data frame extracted from the healthy control embedding were then merged using *left join* and *by="position"*. The cell ids from this data frame were vectorized and filtered using *make.unique* . Finally, the embedding of the healthy control was plotted using *DimPlot* and the cells of the distinct AML samples were highlighted with the vectorized unique cell ids parameter *cells.highlight* .

(iii) By using the *SCTransform*-based workflow features for downstream integration were selected using a gene list published by van Galen *et al.* [149] with the command *SelectIntegrationFeatures*. Then *PrepSCTIntegration* was executed to ensure that all necessary Pearson residuals had been calculated. Integration anchors were computed using *FindIntegrationAnchors* by adding the "Bernstein signature" again as *anchor.features.* Visualization of these UMAPs was performed via the utilization of the R packages ggpointdensity and viridisLite.

Results

1 Workflow for experimental approaches and data analysis methods for multi-omics analysis of primary AML cells

1.1 Development of a workflow for handling and quality assessment of primary human cells

The proper handling of patient samples is critical to preserve the integrity of cells for subsequent single cell sequencing analysis. Single-cell approaches not only require a certain number of cell input, but cells must also be intact, and their membrane should not be damaged before processing. Parameters such as the time between sample collection and processing/freezing, the chosen freezing method, density of cells, or the thawing process need to be adjusted due to their impact on the cell's state. Additionally, extensive treatment with chemotherapeutic drugs can lead to an increase in apoptosis that influences cell viability and consequently, sample quality.

I received viable frozen cells from the clinic that were either collected from peripheral blood or bone marrow. When assessing the quality of my samples, I first investigated the color of the cell pellet. Depending on the used freezing medium, a red, non-see-through color instead of a clear yellow to pink color, indicated that MNCs were contaminated with other blood components. Although it is possible to wash these samples with 1x PBS rigorously, many precious cells are lost during this step. Additionally, evaluation of the size distribution of cells using a Luna cell counter still showed a presence of more than 50% being erythrocytes. High numbers of erythrocytes distort cell counts, contributing to the total number of loaded cell input for the various droplet-based single-cell experiments. Although it would be possible to remove red blood cells during data analysis, this would unnecessarily increase the needed amount of sequencing reads per sample. Thus, I decided to deplete red blood cells using a self-made erythrocyte lysing buffer (Table 2) with an optimized protocol (Figure 10). This lysis step successfully removed the majority of erythrocytes in the samples, without affecting the quality of MNCs.

Cell viability was then assessed using the Luna cell counter combined with trypan blue, which stains dead cells. Both 10x Genomics and Mission Bio recommend processing samples

with viability higher than 70%. However, the quality of some samples did not match this criterion. Due to limitations in sample availability and time, it was impossible to perform a live/dead exclusion using flow cytometry. For further experiments, I would strongly advise not to skip this step to optimize sample quality.

Visualization with the Luna counter showed the ratio of live/dead cells and whether cells were building a clump or other debris was present. Debris was removed using a cell strainer with an appropriate pore size, and when cell clumps were observed, by gently mixing the cells by pipetting up and down 15 times. To decrease the formation of lumps, I started to add 0.1% BSA to every sample when washing the cells after thawing and when resuspended in a working solution. The remaining steps for sample preparation were performed as described in Materials and Methods.

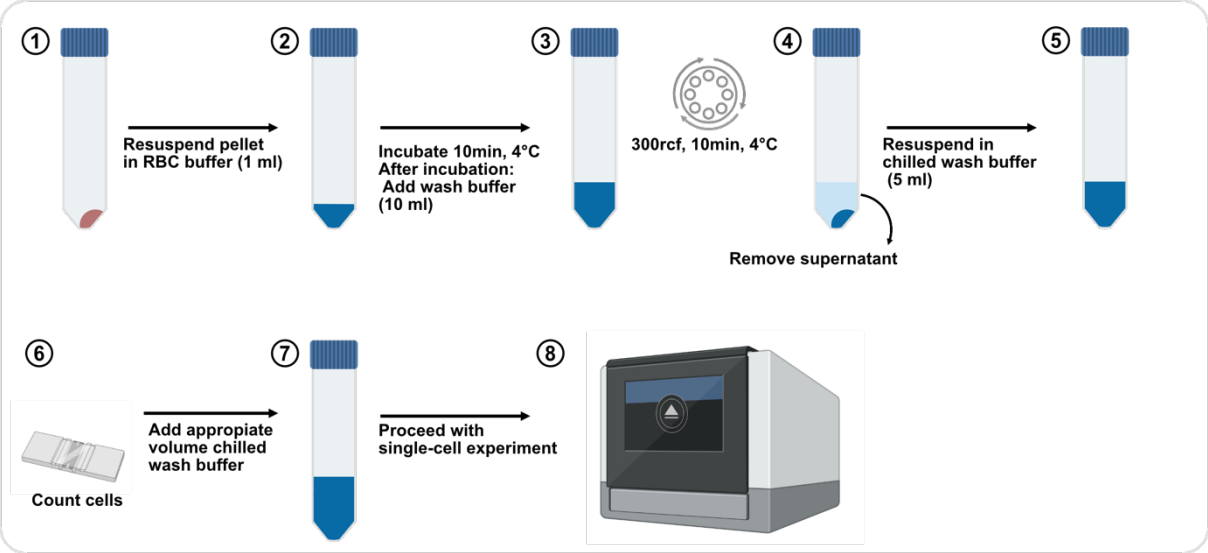


Figure 10

Schematic overview to lyse red blood cells. First, the cell pellet containing erythrocytes is resuspended in RBC lysis buffer and then incubated for 10 minutes at 4°C. Then cells are spun down at 300 rcf for 10 minutes at 4°C. The supernatant is then discarded, and the pellet is resuspended with chilled wash buffer. Cells are then counted with a hemocytometer or an automated cell counter, and the appropriate amount of wash buffer is added to obtain optimal cell concentrations needed for further experiments. Proceed with a droplet-based single-cell experiment. RBC; red blood cell.

In the present study, some whole blood samples were sent or incubated at room temperature before MNCs were isolated, processed or frozen down. Since RNA is very unstable and the transcriptome is severely affected by this, reliable scRNA-seq results obtained from these samples were not possible (see below). Usually, scRNA-seq sequencing data with mitochondrial reads over 20% indicate that samples were not adequately handled

from the clinics and thus should not be considered reliable for further bioinformatic downstream analysis.

1.2 scDNA-seq library protocol with increased ratio of on-/off target fragments

Since the sample quality varied, I adapted the protocol provided by Mission Bio to improve the quality scDNA-seq libraries that had too many off-target fragments and too few on-target amplicons. Expected library sizes differed depending on the used panel to target regions of interest. For this experiment, the commercially available myeloid panel from Mission Bio (see Materials and Methods) was used. Usually, these libraries should depict one prominent peak between 350-550 bp in size. Yet, large off-target fragments can be present in libraries generated with panels containing more than 300 amplicons. These fragments must be considered when quantifying library concentrations.

Additionally, fragments smaller than the target size, e.g., excess primer dimers, severely influence cluster efficiency on Illumina flow cells and have to be removed prior to sequencing [269]. Different scenarios arose during library generation for scDNA-seq and were addressed as following: Library preparation for the cell line TF-1 with wild-type *IDH2* generated high-quality on-target amplicons but high excess primer dimers around with a peak at 219 bp were present (Figure 11A). Thus, I performed another 0.69x clean-up using magnetic beads as described above. This procedure was sufficient to remove low off-target fragments without interfering without affecting the quality of the desired peak, around 450 bp in this case. For the relapse samples *FLT3*-ITD AML2 and *IDH1mut,* the size distribution of the libraries generated with the standard protocol already indicated that large off-target fragments are present in addition to primer dimers. An additional clean-up led to a massive loss of amplicons at the target site (Figure 11B, Figure 11C). This indicates that the normally distributed fragments on the left panel do not display actual large fragments but might instead be daisy chains [270], a phenomenon when DNA forms tangles when being amplified during PCR. This prompted me to further compare size distributions from the target PCR of the samples, which are usually not checked before proceeding with library PCR according to the manufacturer's protocol.

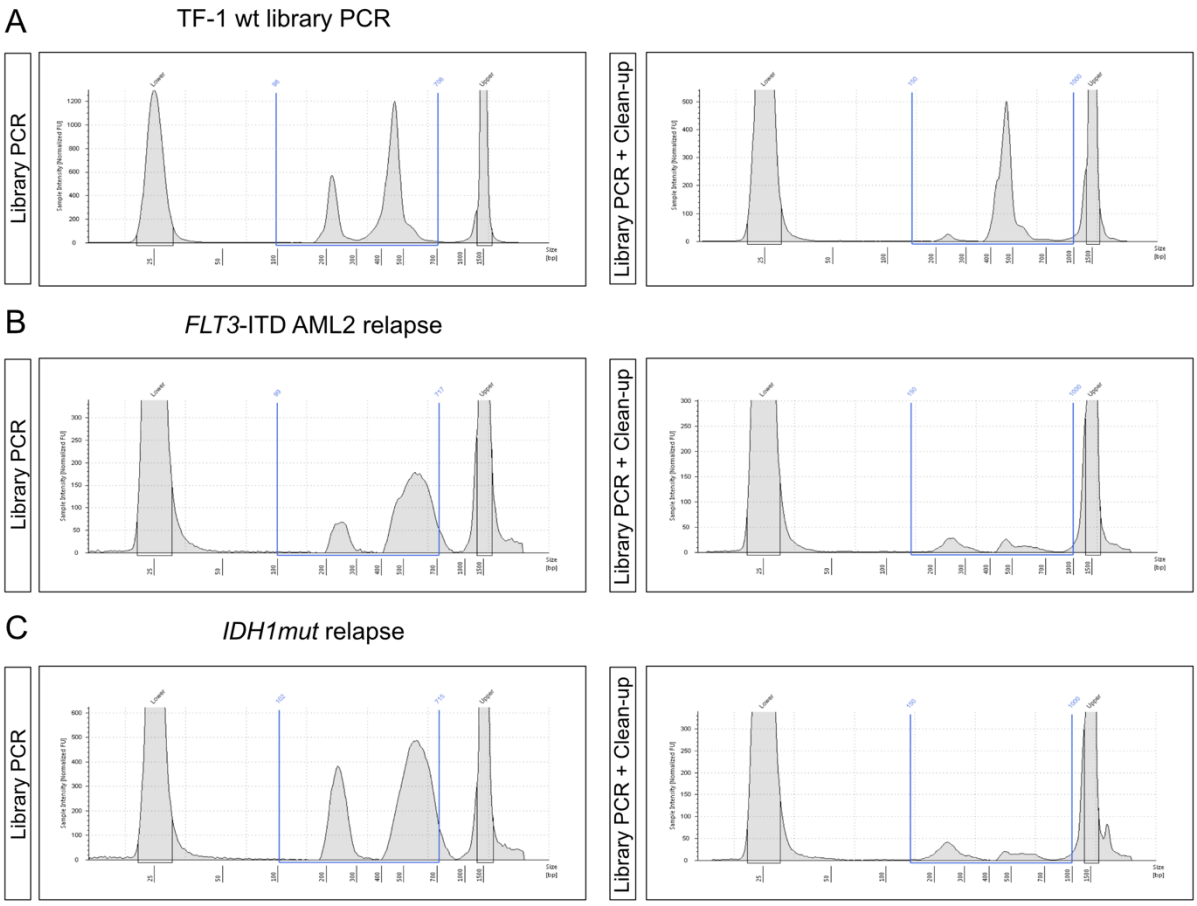


Figure 11
Size distribution of scDNA-seq libraries. (**A**) Size distribution of TF-1 wt library after the standard procedure (left) and after an additional 0.69x bead clean-up (right). (**B**) Size distribution of *FLT3*-ITD relapse AML2 library after standard procedure (left) and after an extra 0.69x bead clean-up (right). (**C**) Size distribution of *IDH1mut* relapse library after the standard procedure (left) and after an additional 0.69x bead clean-up (right).

It became evident that some samples like *FLT3*-ITD AML2 relapse and *IDH1mut* relapse contained an excess amount of primer dimers (Figure 12). Thus, another 0.66x clean-up was performed for these two samples to eliminate small size off-target fragments. Size selection for the *FLT3*-ITD AML2 relapse sample resulted in a peak at the desired length between 350-500 bp with hardly any primer dimers evident (Figure 12). However, for the *IDH1mut* relapse sample, hardly any on-target amplicons were visible, and the largest proportion of amplicons was around 150-200 bp (Figure 12). Library PCR was then conducted according to standard protocol. As expected, library construction scored better results for *FLT3*-ITD AML2 relapse sample compared to *IDH1mut* relapse, displaying a sharp peak at 470 bp. *IDH1mut* relapse sample again showed a normal distribution in amplicons of rather large fragments instead of a sharp peak at the expected size (Figure 12).

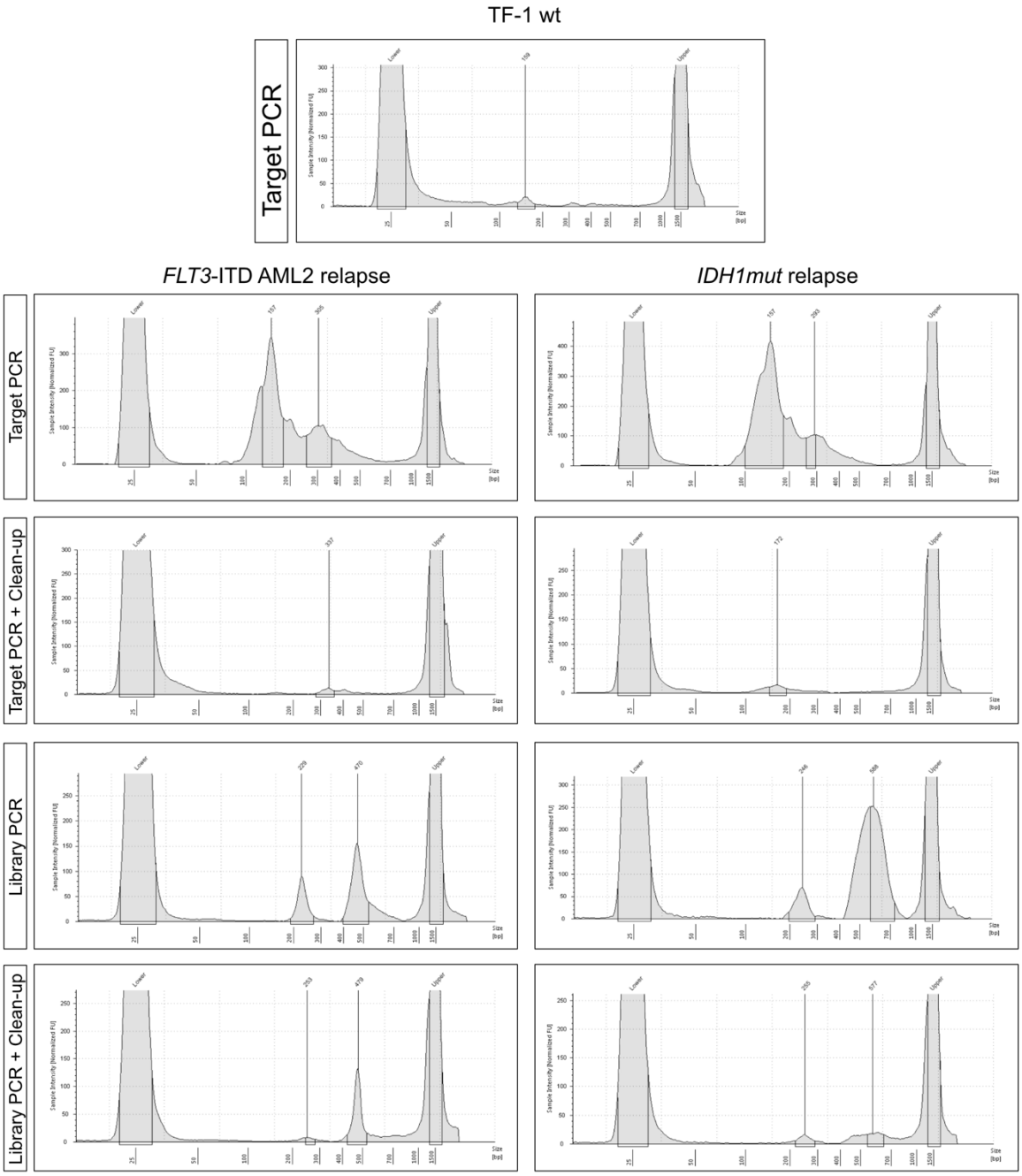


Figure 12

Size distribution of scDNA-seq target PCR products and libraries after adjustments. Size distribution of TF-1 wt target PCR product as a reference compared to *FLT3*-ITD AML2 relapse and *IDH1mut* relapse. The following steps were then only performed for *FLT3*-ITD AML2 relapse and *IDH1mut* relapse.

To remove excess primer dimers, an additional 0.66x clean-up with magnetic beads was performed. The quality of the size distribution for *FLT3*-ITD AML2 relapse library was good, containing mainly on-target amplicons and hardly any primer dimers. Again, clean-up of *IDH1mut* relapse library showed loss of the fragments between 400-500 bp. In conclusion, an additional clean-up of the target PCR products and after library PCR could recover one sample. To recover the last sample, *IDH1mut* relapse, I then decided to take all 15µl of the remaining cleaned-up library from Figure 12 and rerun another PCR with the same settings as indicated in the protocol using the same index primer for 5 more cycles. After this additional step, an on-target peak at around 460 bp and a large off-target primer peak were visible (Figure 13). Another 0.66x clean-up reduced primer dimers extensively, resulting in a high-quality library, ready for sequencing.

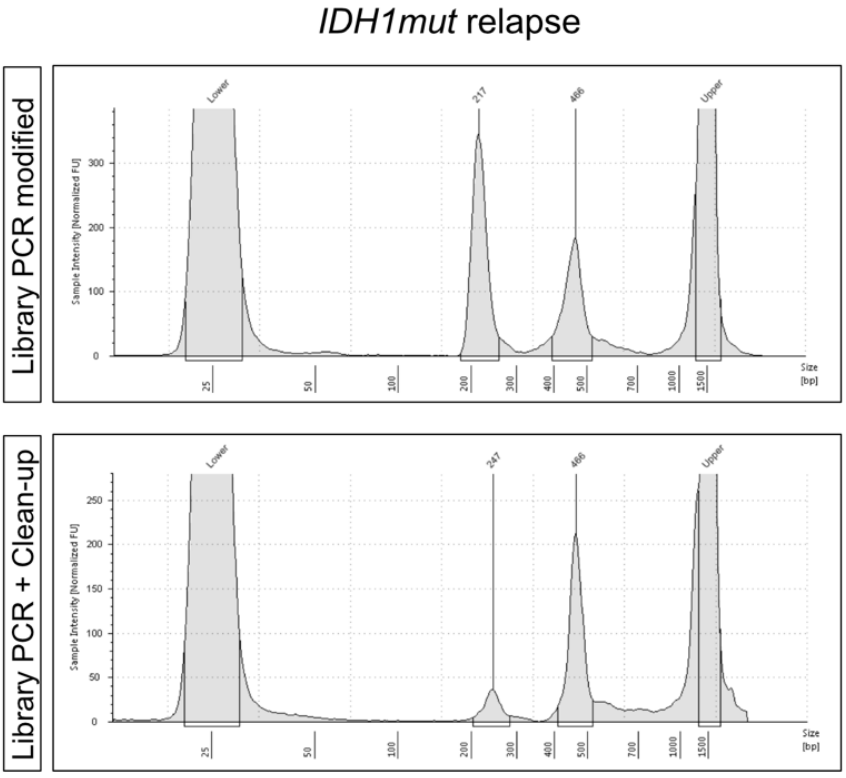


Figure 13
Size distribution of *IDH1mut* relapse libraries. Size distribution for *IDH1mut* relapse sample after rerunning the library with 5 additional PCR cycles and one additional 0.66x magnetic beads clean-up.

1.3 Development of a bioinformatic workflow for multi-omics approaches

In the presented thesis, primary AML samples were analyzed, focusing on the following readouts: scATAC-seq, scRNA-seq, and targeted scDNA-seq. Various methods to obtain gene expression information, especially in the context of blast heterogeneity from scRNA-seq, were evaluated and optimized individually. For AML samples with *FLT3*-ITD or *IDH1* mutations, all three readouts could be performed, and therefore, an additional integration of scRNA and scATAC-seq data was possible. Analysis was focused on scRNA-seq for samples harboring MLL fusions. For bulk ATAC-seq sequencing, primary samples, comparing *IDH1* wt and mutated, were treated ex vivo with BAY143603. The three key aspects of the computational analytic workflow were preprocessing and quality control, individual analysis, and integrated analysis (Figure 14). First, sequencing data was either preprocessed using publicly available or an in-house pipeline. Quality control was conducted separately for each AML sub-type. Second, scATAC-, scRNA- and scDNA-seq data were examined, evaluated, and optimized individually. Lastly, scATAC and scRNA data were integrated whenever possible.

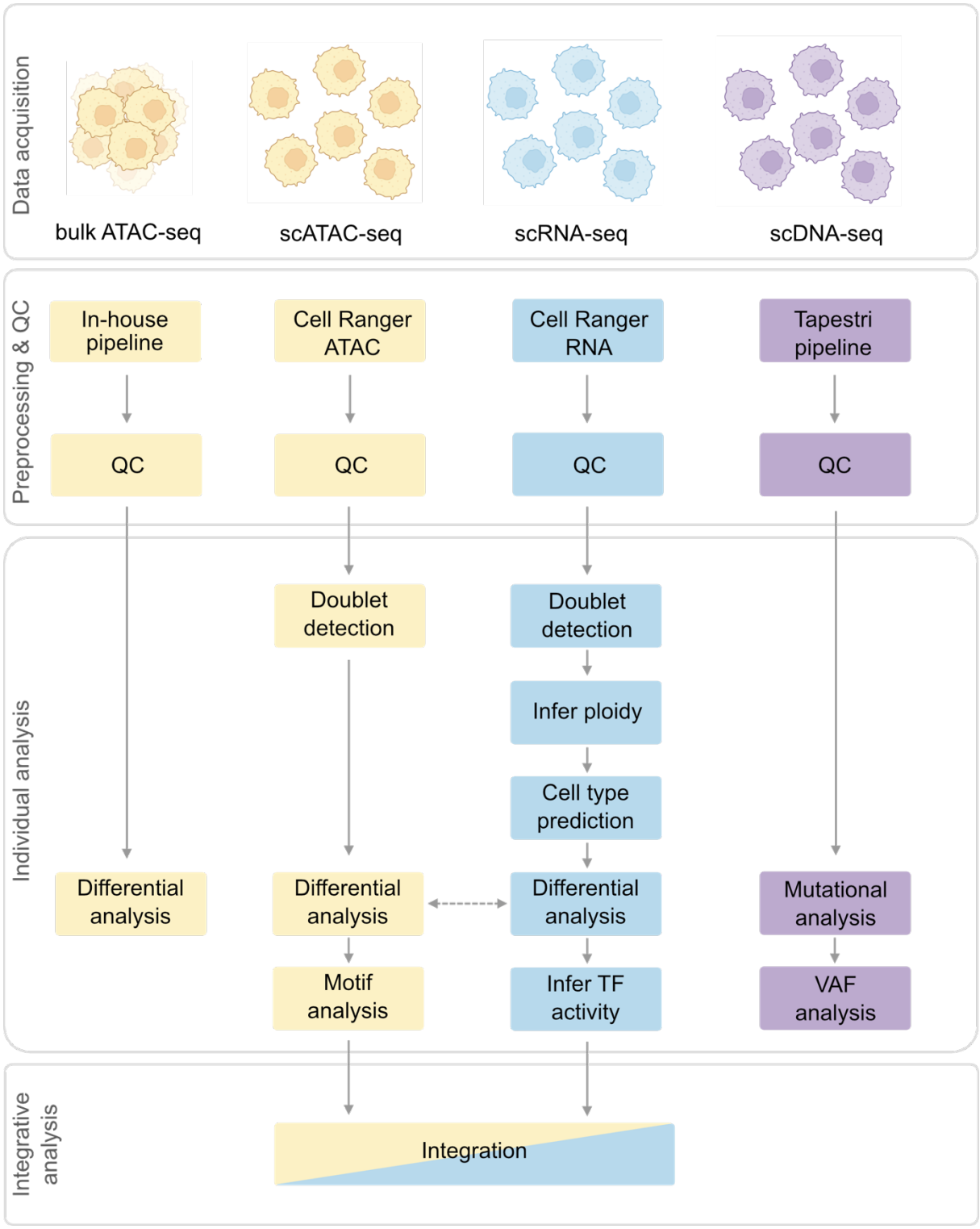


Figure 14

Schematic overview of bioinformatics workflow. ATAC-seq is depicted in yellow, RNA-seq in blue, and DNA-seq in purple. The computational workflow is categorized into preprocessing & QC, individual, and integrative analysis. QC, quality control; TF, transcription factor; VAF, variant allele frequency

1.4 Integration of AML cells with healthy bone marrow donors

Cell surface markers have been used extensively to investigate intra-tumoral heterogeneity [271]. It has recently been shown that scRNA-seq reveals malignant cell variety to a greater extent than it is possible from a limited number of cell markers [149]. Single-cell transcriptomes of leukemic cells can be used to project AML cells onto a myeloid trajectory of healthy reference cells based on similarity [149]. This approach not only characterizes distinct cell type compositions of each AML patient, but also might give insights into the cell of origin, presence of subclones, or to what extent treatment induces differentiation. This prompted me to find the best method to classify AML blasts with *MLL*-r. The following benchmarked the performance for integrating individual AML samples with healthy bone marrow donors using different tools.

1.4.1 Integration of various healthy bone marrow donors

The phenotype of AML blasts can resemble all cells along the HSC to myeloid differentiation axis. Thus, a suitable reference data set containing mature and myeloid progenitor cells was necessary. I chose a publicly available data set from the HCA[266] collected from eight healthy bone marrow donors while the AML cells used were collected from peripheral blood. Since AML blasts represent early progenitors of blood-forming cells that usually would be found in the bone marrow only, comparison to healthy peripheral blood samples was not suitable. Before integrating AML samples with healthy control, I tested the integration of donors alone to establish a work frame and to validate if trajectories would be adequately depicted. Therefore, I used two different integration methods, harmony [250] and reciprocal principal component analysis (rPCA) integration [259]. In total, ~70,000 cells, that had been annotated by Dr. Stephan Tirier, were used. Both harmony and the rPCA method successfully integrated myeloid cells of the bone marrow donors (Figure 15A-H). The various cell types were present in all donors (Figure 15A-B, Figure 15E-F), and cell cycle states were assessed (Figure 15C, Figure 15G). The stem cell marker AVP was highlighted (Figure 15D, Figure 15H) for better visualization of the HSC compartment. Generally, both methods revealed a differentiation trajectory ranging from HSCs to mature $CD16^+$ monocytes with distinct intermediate states. However, integration using harmony was significantly faster. This

analysis verified that both methods are suitable for hematopoietic cell data sets, and subsequently, integration with AML samples was performed.

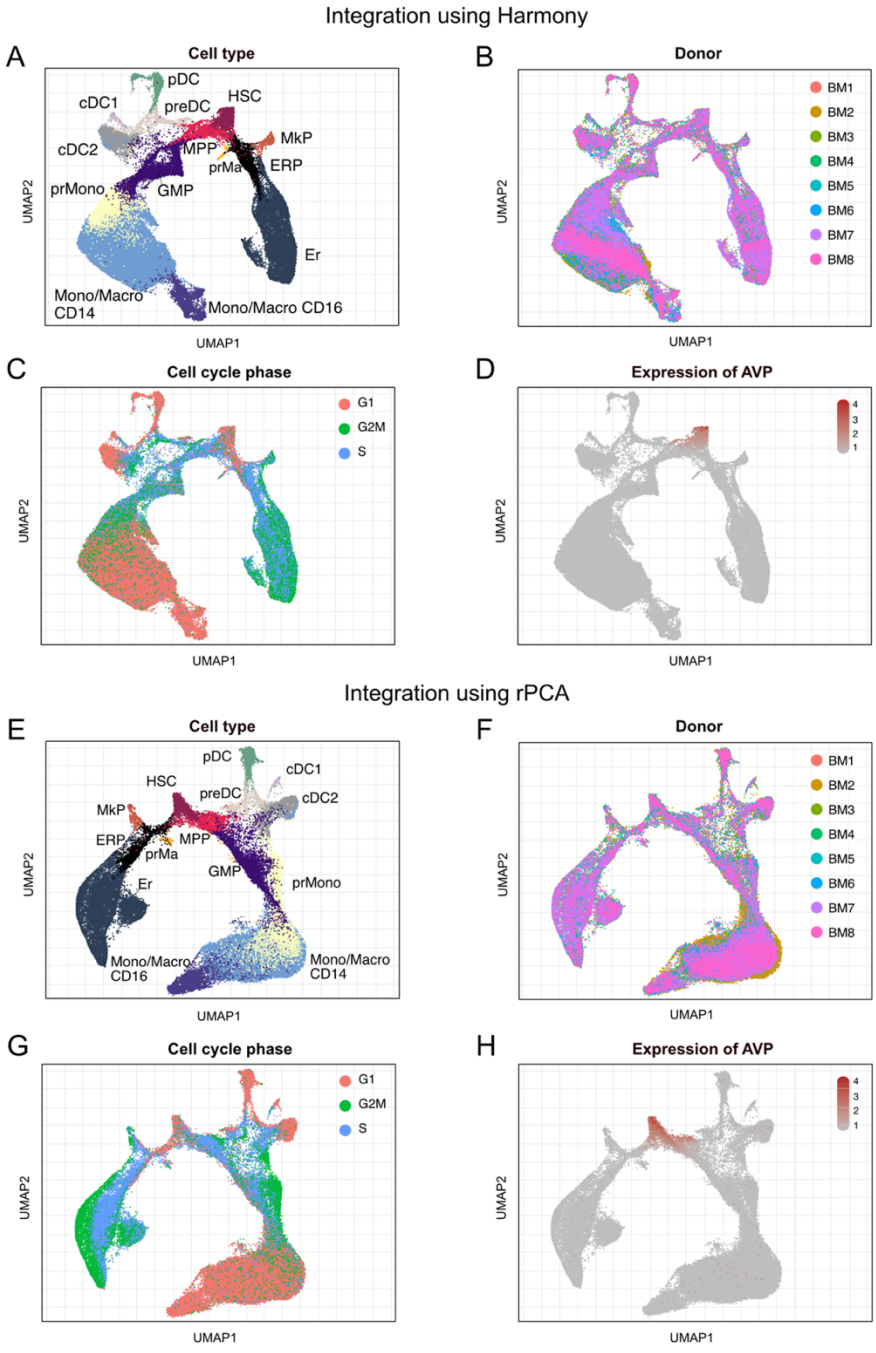


Figure 15

Integration of distinct bone marrow donors. (**A**) UMAP embedding of integrated data using harmony colored per cell type. (**B**) UMAP embedding of integrated data using harmony colored per sample. (**C**) UMAP embedding of integrated data using harmony colored per cell cycle. (**D**) Expression of HSC marker AVP. (**E**) UMAP embedding of integrated data using rPCA colored per cell type. (**F**) UMAP embedding of integrated data using rPCA colored per sample. (**G**) UMAP embedding of integrated data using rPCA colored per cell cycle. (**H**) Expression of HSC marker AVP. HSC, hematopoietic stem cell; MPP, multipotent progenitor; GMP, granulocyte monocyte progenitor; prMono, Monocyte; Mono/Macro CD14, Monocyte/Macrophage $CD14^{+}$; Mono/Macro CD16, Monocyte/Macrophage $CD16^{+}$; MkP, Megakaryocyte progenitor; prMa, mast cell precursor; ERP, erythroid progenitor; Er, Erythrocyte; preDC, pre dendritic cell; cDC1, conventional dendritic cell type 1; cDC2, conventional dendritic cell type 2; pDC, plasmacytoid dendritic cell.

1.4.2 Integration of AML samples with healthy bone marrow control

Next, several integration methods were tested to project AML blasts on a healthy myeloid trajectory. In particular, harmony, rPCA, and the fast nearest neighbor (FNN) approach were employed (Figure 16). To test which method is most suitable, AML cells from a patient with an *MLL-EDC4* fusion were subset from a healthy microenvironment and used to integrate with the healthy bone marrow control. Harmony was not able to incorporate the AML sample with bone marrow control. While there was still a trajectory of non-malignant myeloid cells evident (Figure 16A) and healthy donors integrated well (Figure 16B), the AML cells accumulated rather between the clusters of MPPs and plasmacytoid dendritic cells instead of being projected on the bone marrow. Similar results were shown for the integration using rPCA with standard parameters (Figure 16C, Figure 16D). However, rPCA seemed to work slightly better since the AML cluster was divided into two distinct subgroups indicating differences in the composition of AML cell types. While one AML subset clustered in proximity to HSCs and MPPs, the second one clustered next to $CD14^+$ monocytes/macrophages. This prompted me to refine integration parameters for rPCA further. Next, using the FNN tool, calculated the nearest neighbors of the AML samples based on the extracted embedding of the healthy donors (Figure 15F, Figure 16E) and then highlighted AML cells on the bone marrow embedding in red (Figure 16F). However, this method was only used with the parameter of *k=1.* Thus, AML cells were only projected onto the map based on one similar cell.

Additionally, further validation using the R tool SingleR and computation of module scores (see Classification of cell type abundance in *MLL-r* AML blasts) did not show matching results. Thus, I decided to proceed with rPCA integration using a specified gene signature as a list (see Appendix) for integration anchors. These genes were published by Van Galen *et al.* [149] and generated by a machine learning classifier for AML cellular hierarchies. AML cells were successfully integrated with healthy bone marrow controls using this approach (Figure 16G, Figure 16H). Thus, this method was used for further analysis.

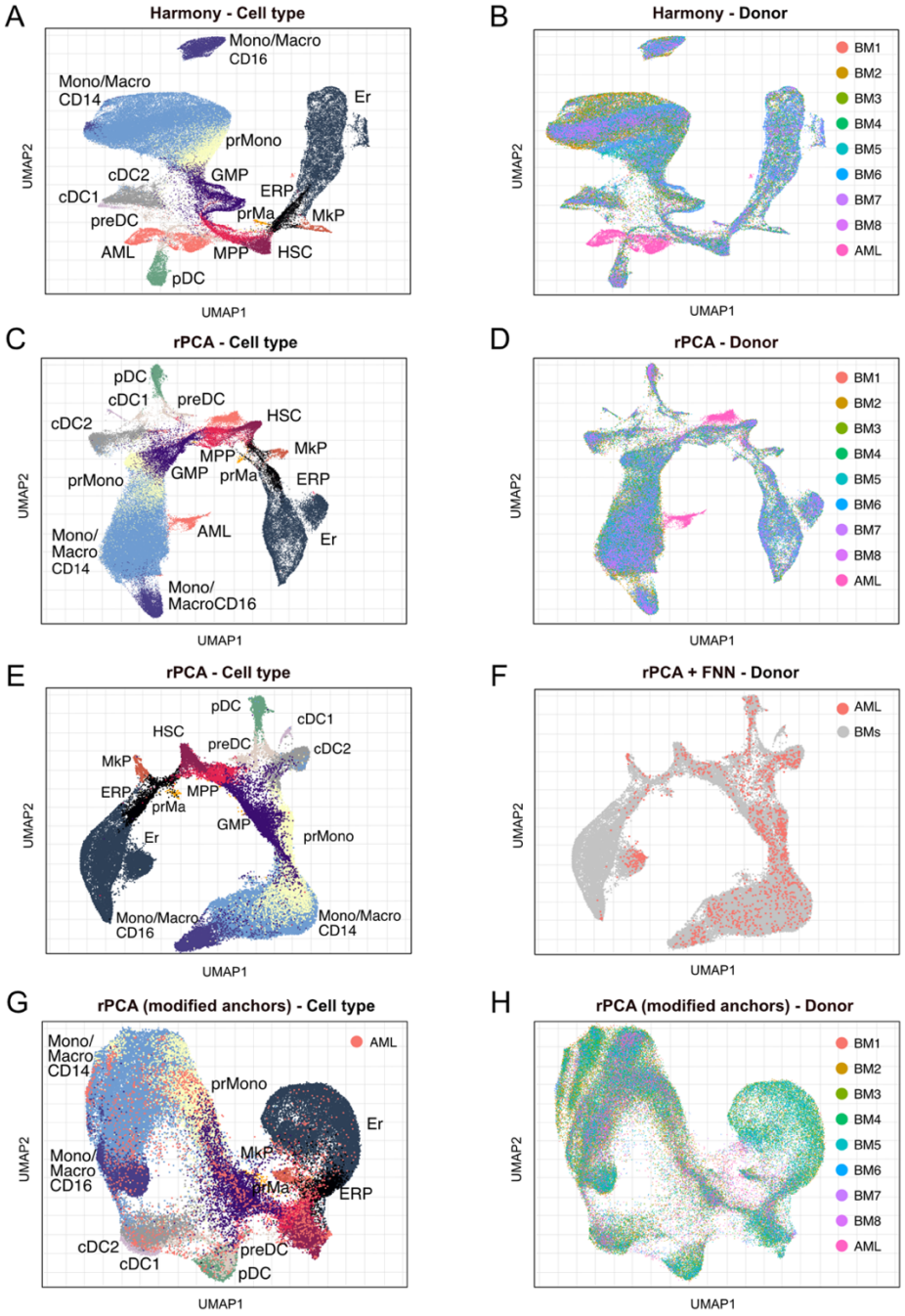


Figure 16

Integration of AML cells with healthy bone marrow control. (**A**) UMAP embedding of integrated data using harmony colored per cell type. (**B**) UMAP embedding of integrated data using harmony colored per sample. (**C**) UMAP embedding of integrated data using rPCA with standard parameters colored per cell cycle. (**D**) UMAP embedding of integrated data using rPCA with standard parameters colored per sample. (**E**) UMAP embedding of integrated data using rPCA in combination with FNN. AML cells are highlighted in red, whereas healthy bone marrow samples are colored grey. (**F**) UMAP embedding of integrated data using rPCA with standard parameters colored per sample. (**G**) UMAP embedding of integrated data using rPCA with a specified anchors list for integration colored per cell cycle. (**H**) UMAP embedding of integrated data using rPCA with a specified anchors list for integration colored per sample. HSC, hematopoietic stem cell; MPP, multipotent progenitor; GMP, granulocyte monocyte progenitor; prMono, Monocyte; Mono/Macro CD14, Monocyte/Macrophage $CD14^{+}$; Mono/Macro CD16, Monocyte/Macrophage $CD16^{+}$; MkP, Megakaryocyte progenitor; prMa, mast cell precursor; ERP, erythroid progenitor; Er, Erythrocyte; preDC, pre dendritic cell; cDC1, conventional dendritic cell type 1; cDC2, conventional dendritic cell type 2; pDC, plasmacytoid dendritic cell

1.5 Pseudotime analysis of leukemic myeloid cells

Single-cell transcriptomics can track dynamic changes in cell fates and understand temporal transitions of cell states by inferring trajectories. This approach has been used to analyze numerous biological processes such as cell cycle, cancer formation, and cell differentiation by displaying transcriptional changes along the calculated trajectory or pseudo-time [272]. However, it remains challenging to reliably model high-resolution cell trajectories and automatically recognize the relevant cell fates and lineage [273]. Prior knowledge of gene expression signatures can facilitate determining the directionality of differentiation trajectories in physiological conditions. In healthy individuals, stem cells would be expected to represent a starting point spanning a trajectory over progenitor cells that ends with highly differentiated cell states [274]. In cancer, trajectories can contain many bifurcations, dead ends, or reversible paths reflecting trans- and dedifferentiation. Using scRNA-seq data from two patients with distinct *MLL* fusions, I tested three different approaches to infer myeloid leukemic cell trajectories.

First, RNA velocity analysis was performed using sc-velocyto (Figure 17), which uses the relative ratio between intronic and exonic reads to infer dynamics in transcript abundance to anticipate future transcriptional cell states [262]. For *MLLT3-MLL* AML-2, velocity streams were estimated to wander towards two different directions (Figure 17A), with a breaking point between the transition of promonocytes and $CD14^+$ monocytes. Pseudotime analysis of this patient determined the root of cells at $CD16^+$ monocytes and the stop point at GMPs and promonocytes (Figure 17C), which is counter-intuitive under physiological conditions but could still represent dedifferentiation processes which is a common hallmark in cancer [275]. Visualization of PAGA graph with velocity-directed edges depicted connectivities between all cell types (dashed lines) but no transitions (solid lines) (Figure 17E). Similar results were obtained for leukemic cells from *MLL-ELL* AML regarding velocity streams and connectivities (Figure 17B, Figure 17F). Nonetheless, pseudotime analysis marked the cluster of undifferentiated promonocytes as starting cells and $CD14^+$ monocytes as the endpoint (Figure 17D). Subsequently, data from the same samples as analyzed with PAGA and slingshot (Figure 18). In contrast to sc-velocyto, PAGA [253] uses a cluster-graph representation to capture the underlying topology. Slingshot [261] employs a minimum spanning tree to connect clusters following principal curves are fitted for detected branches.

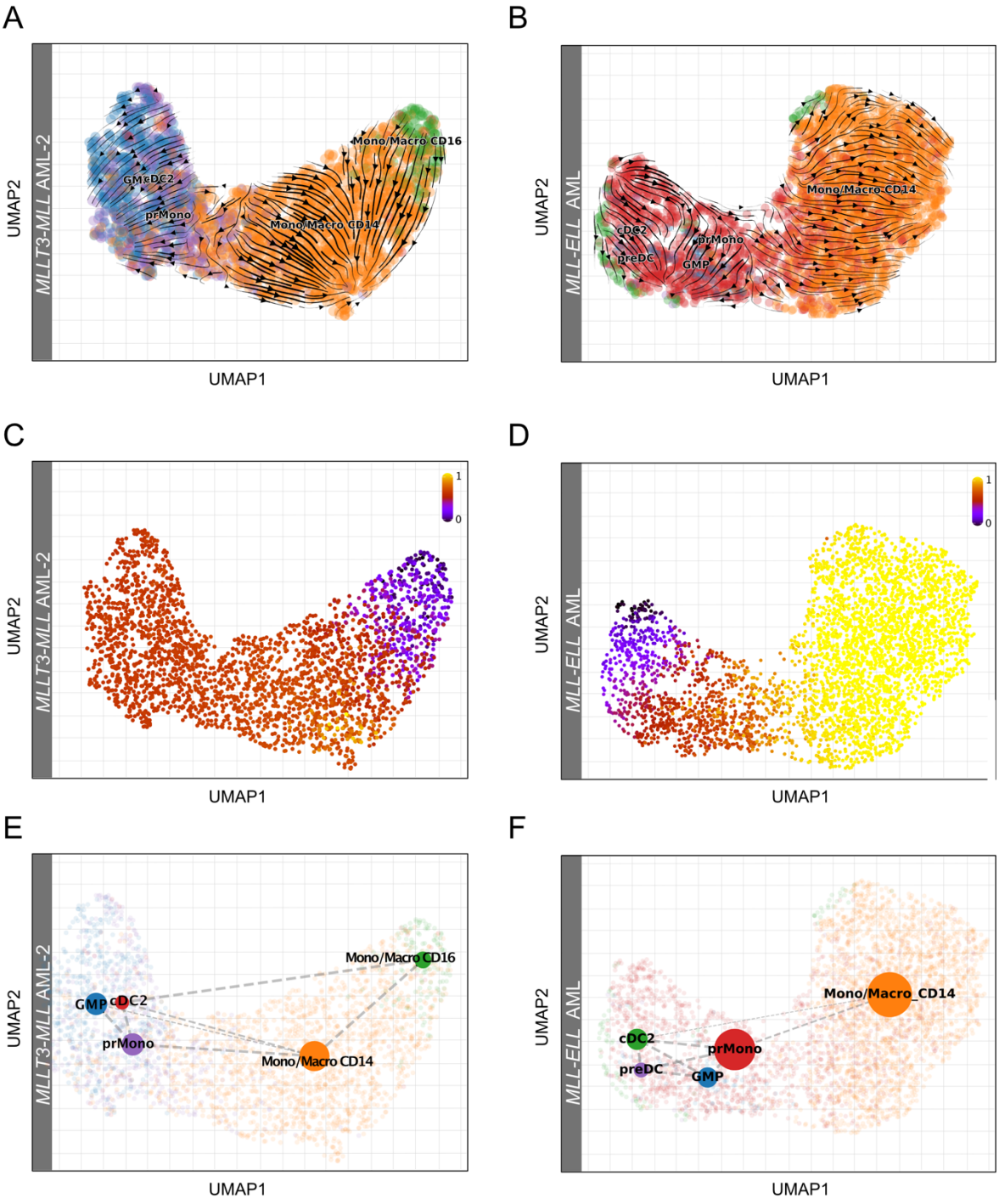


Figure 17

Trajectory inference of myeloid leukemic blasts using sc-velocyto. (**A**) Velocities of leukemic cells from patient *MLLT3-MLL* AML-2 projected onto UMAP embedding. (**B**) Velocities of leukemic cells from patient *MLL-ELL* AML projected onto UMAP embedding. (**C**) Computed velocity pseudotime of *MLLT3-MLL* AML-2 cells. Root cells and end points are colored based on pseudotime color scale. (**D**) Computed velocity pseudotime of *MLL-ELL* AML cells. Root cells and end points are colored based on pseudotime color scale. (**E**) PAGA graph on UMAP embedding with velocity-directed edges for *MLLT3-MLL* AML-2 cells. (**F**) PAGA graph on UMAP embedding with velocity-directed edges for *MLL-ELL* AML cells.

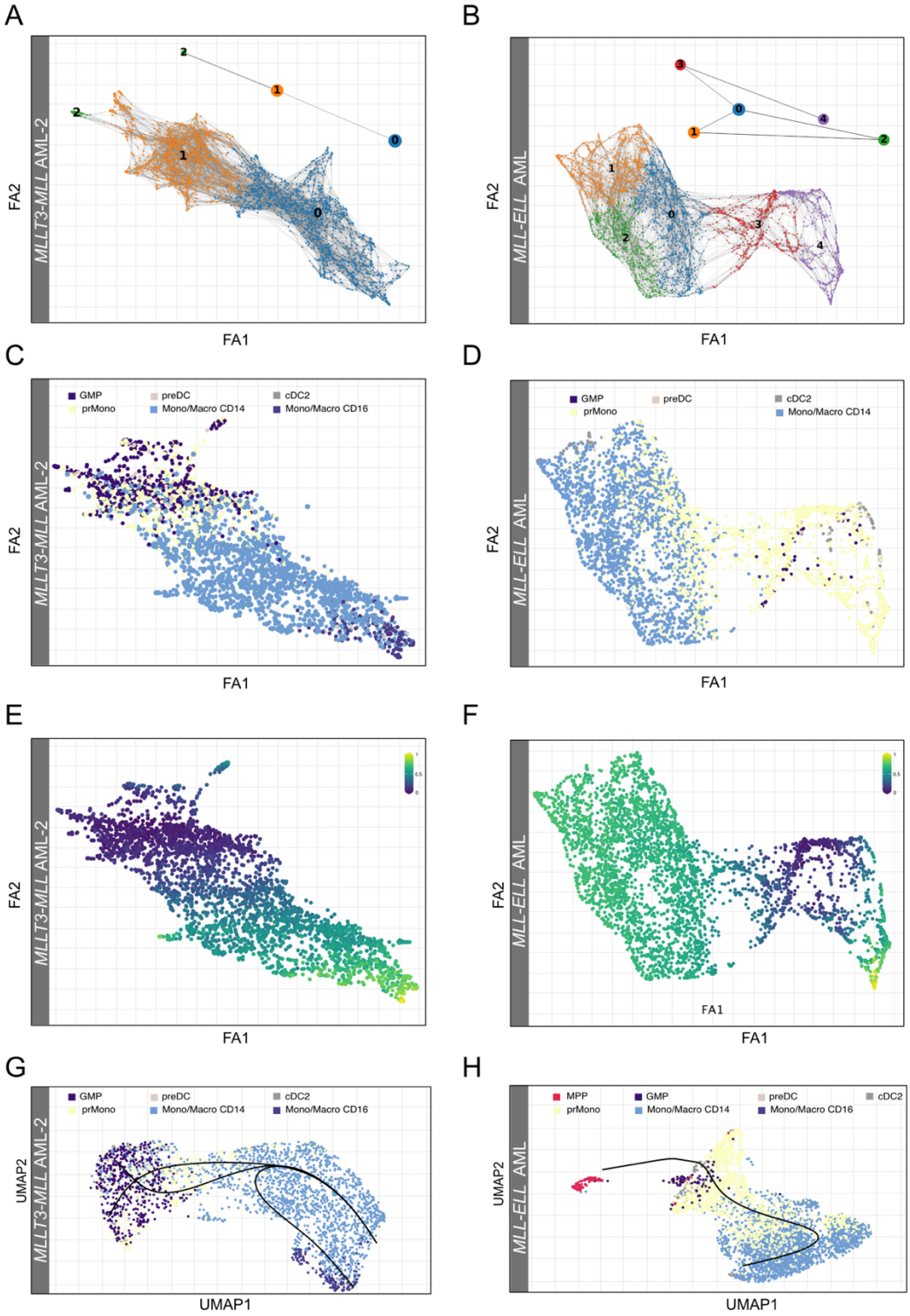


Figure 18

Trajectory inference of myeloid leukemic blasts using PAGA and slingshot. (**A**) Recomputed FA embedding using PAGA-initialization of MLLT3-MLL AML-2 cells colored by detected clusters. (**B**) Recomputed FA embedding using PAGA-initialization of *MLL-ELL* AML cells colored by detected clusters. (**C**) FA embedding of *MLLT3-MLL* AML-2 cells colored by predicted cell types. (**D**) FA embedding of *MLL-ELL* AML cells colored by predicted cell types. (**E**) FA embedding of *MLLT3-MLL* AML-2 cells colored by pseudotime. (**F**) FA embedding of *MLL-ELL* AML cells colored by pseudotime. (**G**) UMAP embedding of *MLLT3-MLL* AML-2 cells for trajectory inference using slingshot. Principal curves depict smoothed representations of different cell types. (**H**) UMAP embedding of *MLL-ELL* AML cells for trajectory inference using slingshot. Principal curves portray smoothed representations of varying cell types. FA, factor analysis.

For patient *MLLT3-MLL* AML-2, PAGA detected three clusters and cell types were aligned in a linear differentiation trajectory with terminal states being predicted at the monocytic compartment (Figure 18A, Figure 18C, Figure 18E). In contrast, slingshot added additional branches visualized by smoothed principal curves (Figure 18G). PAGA detected 5 clusters for leukemic cells from *MLL-ELL* AML, with cluster 0 having the most connectivities to other clusters (Figure 18B). Cells from this cluster might represent intermediate cells transitioning between the stated from promonocytes to more differentiated monocytes, according to their marker profile (Figure 18D). For pseudotime analysis inferred from PAGA, starting and end points were again contrary to results obtained from sc-velocyto (Figure 18F). Trajectory inference using slingshot resulted in one branch along the classical myeloid differentiation axis (Figure 18H). In conclusion, all tested tools led to distinct outcomes of trajectory and pseudotime inference, impeding estimation of accurate trajectories. Although there is a plethora of tools to infer the temporal dynamics of cells, there is still an opportunity to refine algorithms for precise trajectory detection.

2 Cell type analysis of *MLL-r*

MLL rearrangements generally confer a poor prognosis in AML [30]. Recently, the enhancer of mRNA decapping 4 (*EDC4*) gene was described as a new *MLL* fusion partner (*MLL-EDC4*)[276]. *EDC4* is pivotal for the decapping process and required for the assembly of RNA binding proteins into higher-order complexes termed processing (P)-bodies [277]. P-bodies contain untranslated mRNAs, exonucleases, the mRNA decapping machinery, and translational repressors [278]. Stabilization or decay of P-body associated mRNA is regulated in a context-dependent manner or can be stress-induced [279].

EDC4 plays a critical role in the posttranscriptional modulation of the pro-inflammatory cytokine IL-6 in leukemic macrophage cell lines [280] and IκB kinase-*EDC4* interaction has been connected to P-body formation and control of mRNAs encoding inflammatory cytokines[281]. Since *MLL*-r AML usually occurs *de novo,* the clinical presentation of the patient with a preceding myelodysplastic syndrome before developing secondary AML was quite unusual. Additionally, it was hypothesized that these features combined with an indolent clinical course, could relate to a more stem-cell-like disease. Thus, I wanted to predict developmental stages of AML blasts inferred from the respective transcriptome changes and compare them to three other patients (2 patients with *MLL-MLLT3*, 1 patient with *MLL-ELL*) with initial presentations typical for *MLL* rearranged AMLs, e.g., without a preceding MDS phase. Only a few single-cell RNA-seq analyses on AML cases with rearranged *MLL* have been performed so far. Thus, this project serves as an additional single-cell reference to better characterize different *MLL*-r AMLs.

2.1 scRNA-seq resolves the intratumor heterogeneity in samples of different *MLL* fusions

Sequencing data were preprocessed, and quality was assessed as indicated in Material and Methods. The UMAP approach for unsupervised dimensionality reduction and clustering was used to integrate ~16 000 single cell transcriptomes into a two-dimensional map. UMAP visualization of the four merged patient samples revealed intra-tumoral heterogeneity for patient 2 with *MLLT3-MLL AML* and the patient carrying *MLL-ELL* fusion cells, whereas *EDC4-MLL* AML and patient 1 with *MLLT3-MLL AML* showed a more homogenous phenotype

(Figure 19A). Non-malignant microenvironment clustered per cell type independently, whereas AML blasts clustered individually (Figure 19B). Thus, a further batch correction was not necessary. Manual marker-based assignment of cell identities was distinguishable for the microenvironment (Figure 19C). A limited set of predefined markers may not accurately distinguish tumors from normal cells, due to the diversity of malignant myeloid differentiation states in AML and the possibility of transcriptional expression in healthy cells [282]. To further validate the separation between healthy and cancer cells, I employed CopyKAT, an R package that identifies genome-wide aneuploidy using an integrative Bayesian approach [243]. CopyKAT successfully inferred leukemic cells that matched my annotation based on marker gene expression profiles (Figure 19D).

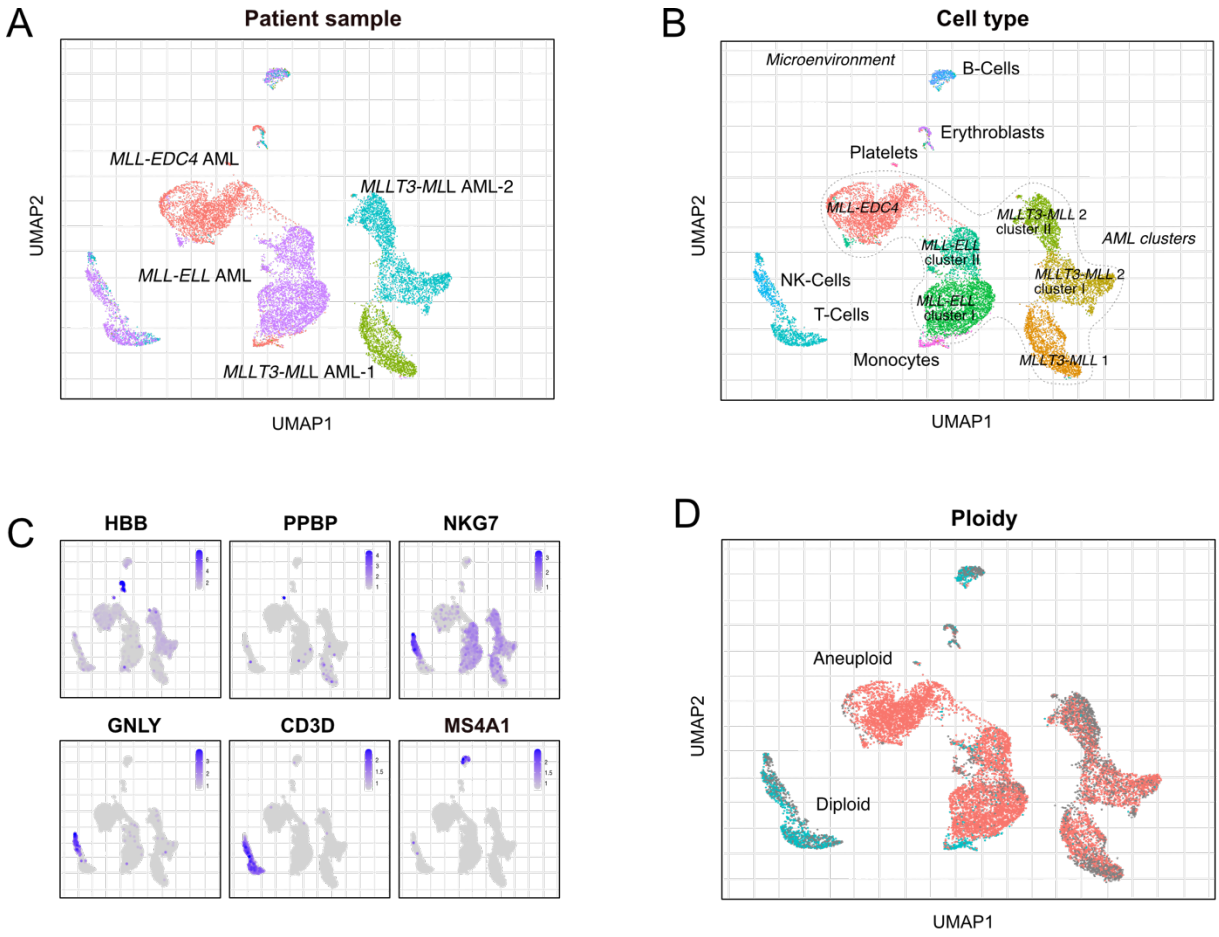


Figure 19

Annotation of patients with MLL-r AML. (**A**) UMAP embedding of merged AML samples colored by patient. (**B**) UMAP embedding of merged AML samples colored by cell types. Non-malignant cell types from all samples cluster together, whereas AML cells form distinct clusters for each patient (dashed line). (**C**) UMAP embedding of cell type specific markers. NK-Cells, Natural Killer -Cells.

2.2 Characterization of transcriptome profiles

Next, I performed differential gene expression analysis to assess inter and intra-patient tumor heterogeneity on a transcriptomic level. The transcriptional patterns of the leukemic cells of *MLL-MLLT3* and *MLL-ELL* fusions depicted many commonalities in the upregulation of classical monocytic markers, while there were great transcriptional differences evident in comparison to the *MLL-EDC4* fusion. Differential gene expression analysis identified both a significant upregulation of transcription factors and genes that are known to have an impact on cell-fate decision and cellular differentiation (*NPM1, CDK6, SOX4, GATA2, MYC, DACH1*) and upregulation of a set of genes that were described to play a critical role in hematopoiesis or leukemic stem cell activation (*FLT3, HOPX*, *HOXA9, RUNX1*) [283-288] (Figure 20A). Additionally, the expression of several downstream targets of the hematopoietic key regulator *RUNX1* (*UBB, PSNE1, ARID1B, KIAA0125*) which are reported to be involved in the differentiation of myeloid cells were upregulated [289]. Interestingly, the most differentially expressed gene in patient *MLL-EDC4* was lactate dehydrogenase B (*LDHB),* an enzyme mediating the switch on the anaerobic glycolysis and lactate production after hypoxia challenge [290,291]. Analysis of the microenvironment showed high expressions of CD36, cathepsins, and CLEC receptors in the monocytic compartment (Figure 20B).

Gene set enrichment analysis (GSEA) also showed a downregulation of the innate immune system, myeloid leukocyte mediated immunity, and activation in the leukemic population of *MLL-EDC4* (Figure 20D). Conversely, *MYC* targets, interferon-alpha response, and genes involved in eukaryotic translation initiation and elongation were upregulated (Figure 20C, Figure 20D). Additionally, GSEA highlights an upregulation of *EIF2AK4* response to amino acid deficiency. Furthermore, pathways linked to oxidative phosphorylation and reactive oxygen species (ROS) were upregulated in *MLL-EDC4* aberrated cells.

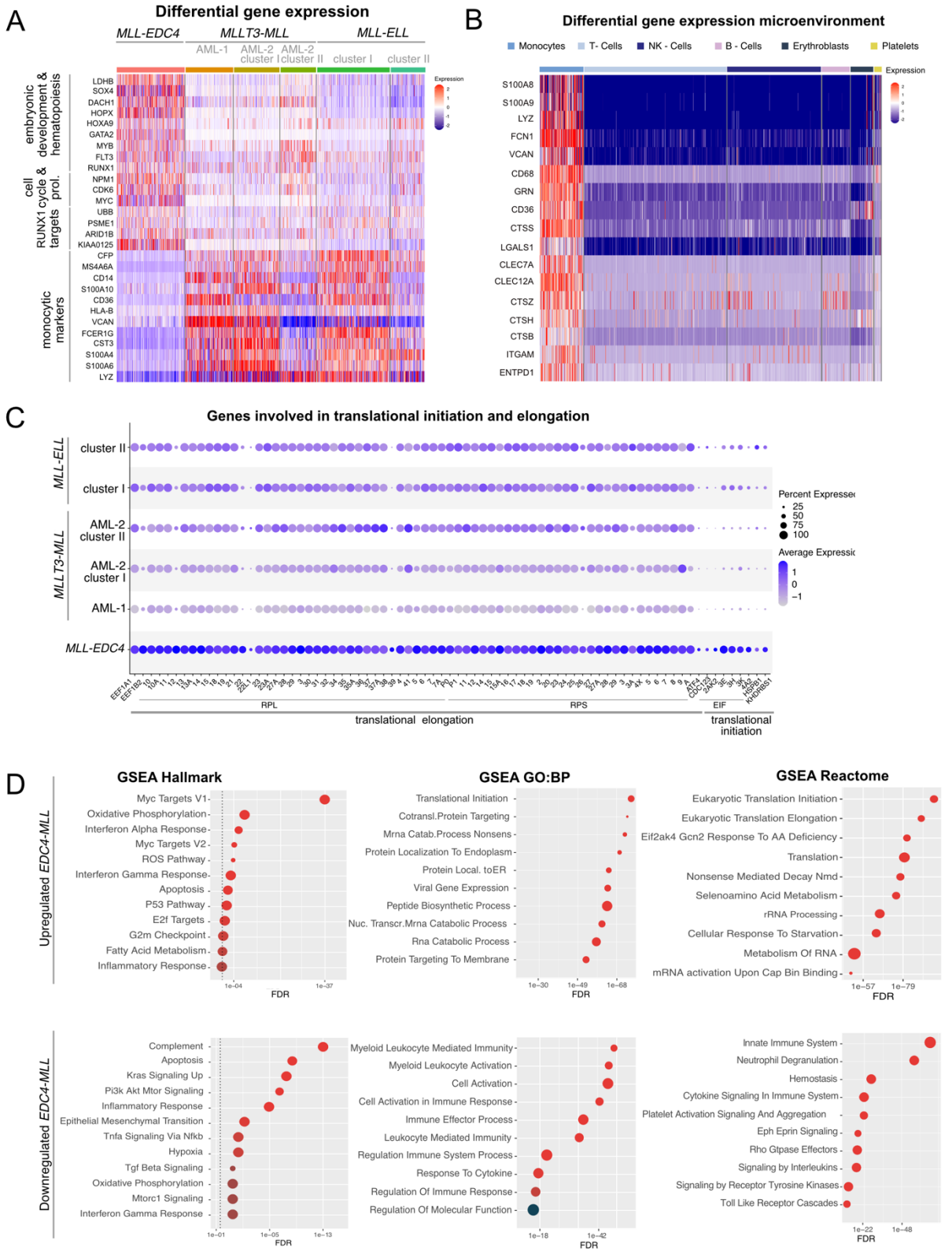


Figure 20

Differential gene expression in *MLL-EDC4* compared to *MLL* fusions with *MLLT3* and *ELL*. (**A**) Clustered single-cell transcriptomic heatmap for the most differentially expressed genes between all AML clusters. (**B**) Clustered single-cell transcriptomic heatmap for the most differentially expressed genes between non-malignant cell types. (**C**) Dot plot of genes involved in translational initiation and elongation expressed in distinct AML clusters. (**D**) Dot plot of up and down regulated gene sets of *MLL-EDC4* AML cells compared to all other AML clusters. Gene sets from Hallmark, Reactome, and GO:BP were used.

2.3 Transcription factor activity in *MLL-EDC4*

The specific upregulation of well-established master regulators of stem cell programs such as *SOX4*, *GATA2*, *MYC, and RUNX1* in *EDC4-MLL* led to a systematic evaluation of TF expression and their activity based on target gene expression. This analysis highlighted an upregulation of interferons such as *STAT1*, *STAT2,* and *IRF9*, the oncogenes *MYB ,MYC* as well as other TFs like *E2F1, E2F4, ETS1, GATA1, NFYA, POU2F1, SPI1, TAL1* that have been linked to stemness in hematopoietic cells [292-294] (Figure 21A).

A network of interacting TFs was created using this information (Figure 21B). Based on unsupervised clustering, MYC was visualized as a central node in the network related to numerous TFs as a first or second edge. *MYC* plays a crucial role in cell proliferation, growth, and tumorigenesis [295]. In addition, *POU2F1* TF activity was shown to be high in *MLL-EDC4* cells, which was reported to control cell growth, stem cell identity, cellular stress response, and immune regulation [296]. Furthermore, a signature of high transcription factor activity in *MLL-EDC4* was detected for TFs such as *E2F1, E2F4, ETS1, MYB, MYC, GATA1,* and *TAL1* that could potentially be linked to a more stemlike phenotype [292-294].

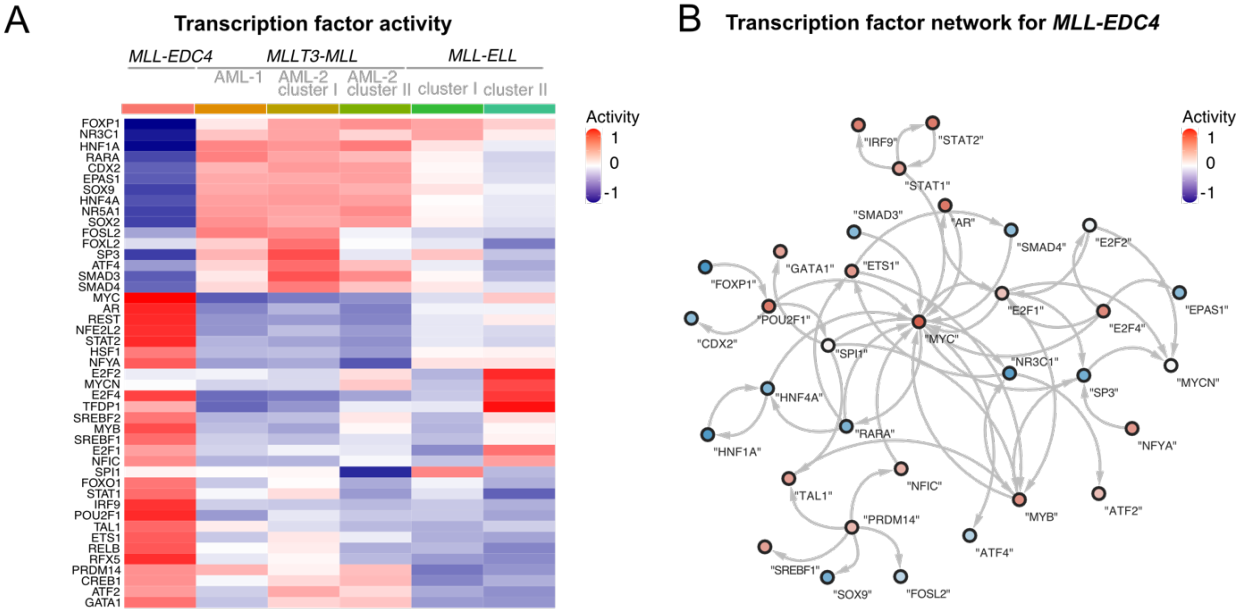


Figure 21

Transcription factor activity in *MLL-EDC4* compared to *MLL* fusions with *MLLT3* and *ELL*. (**A**) Heatmap depicting TF activities of distinct AML clusters inferred from scRNA-seq data. (**B**) TF network colored by TF activities obtained from A.

2.4 Classification of cell type abundance in *MLL-r* AML blasts

I employed automated cell type prediction [260] using the HCA [266] bone marrow dataset from eight healthy donors as a training dataset to further characterize the leukemic cells' differentiation states and to dissect intra-tumoral heterogeneity. This cell type prediction showed a distinct phenotype for MLL-EDC4 cells compared to other *MLL* fusions (Figure 22A, Figure 22B). Leukemic cells with *MLL-EDC4* translocation were almost exclusively classified as HSCs, MPPs, or ERPs. While differentiated monocytes predominated in *MLLT3-MLL* AML-2 and *MLL-ELL* AML subclusters I and *MLLT3-MLL* AML-1, cell type abundance in *MLLT3-MLL* AML-2 and *MLL-ELL* AML subclusters II was more diverse. *MLLT3-MLL* AML-2 and *MLL-ELL* AML subclusters II composed of cells like GMPs and promonocytes, respectively.
Metadata of the predicted cell types were then plotted onto the UMAP embedding previously shown (Figure 22C). This revealed a trajectory from myeloid progenitors to monocyte-like cells for *MLLT3-MLL* AML patient 2 subcluster II to subcluster I and *MLL-ELL* subcluster II to subcluster I, respectively (Figure 22D).

Then, I constructed gene signatures (see Appendix) for the distinct cell types based on the highest expressed cell type indicators from the HCA data set, which was conceptually similar to a prior approach [149]. These signatures were used to stratify leukemic cells of all four patients by calculating the average expression levels of each tumor cluster on a single cell level, subtracted by the aggregated expression of control feature sets. HSC- and MPP-gene module scores were raised in the *MLL-EDC4* patient, but there was no elevation in monocytic CD14 genes (Figure 22E). Leukemic cells from *MLLT3-MLL* AML patients and cluster I from *MLL-ELL* AML patients, on the other hand, displayed an entirely reversed pattern. Only cluster II from *MLL-ELL* AML revealed elevated progenitor module scores compared to *MLL-MLLT3* fusions. However, *MLL-EDC4* ranked highest scores in the stem cell and progenitor compartments.

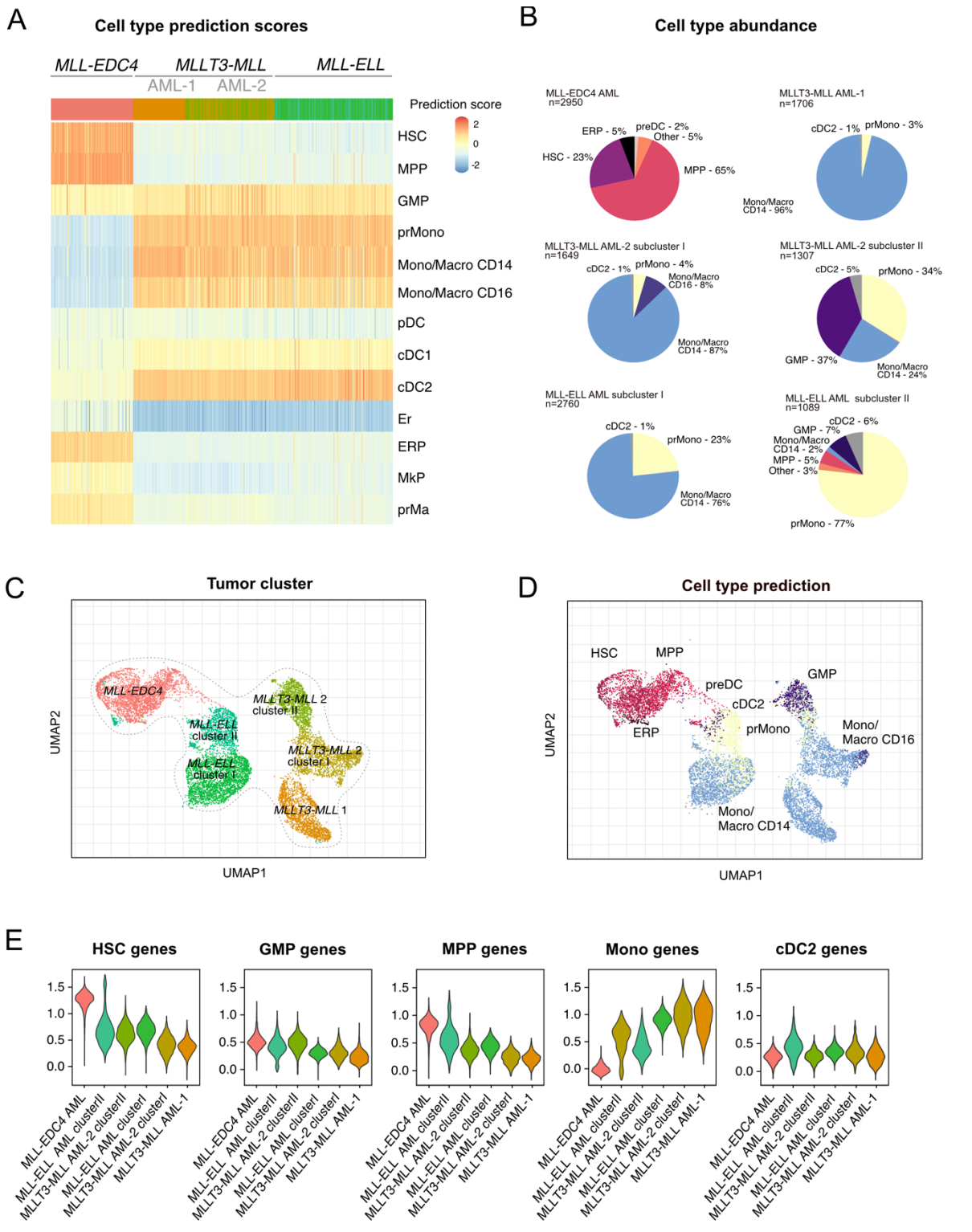


Figure 22

Cell type prediction of AML samples with *MLL* fusions. (**A**) Clustered single-cell heatmap showing prediction scores for all *MLL* samples. (**B**) Pie charts of relative abundance of predicted cell types. (**C**) UMAP embedding of malignant cells with *MLL*-fusions, colored by intratumoral clusters. (**D**) UMAP embedding of malignant cells with *MLL*-fusions, colored by predicted cell type. (**E**) Violin plots showing module scores for different gene signatures. HSC, hematopoietic stem cell; MPP, multipotent progenitor; GMP, granulocyte monocyte progenitor; prMono, Monocyte; Mono/Macro CD14, Monocyte/Macrophage CD14^{+}; Mono/Macro CD16, Monocyte/Macrophage CD16^{+}; ERP, erythroid progenitor; preDC, pre dendritic cell; cDC2, conventional dendritic cell type 2.

2.5 Integration of *MLL*-r cells with healthy bone marrow donors

AML cells were projected along a myeloid trajectory, based on the empirical finding that integration using rPCA with modified anchors worked best for this data set. All eight healthy bone marrow donors were integrated with the four AML samples with distinct *MLL* fusions (Figure 23). After integration, visualization using UMAP depicted a clear separation of each cell type along a differentiation trajectory, and leukemic cells were speckled along this HSC to monocyte axis rather than forming their own independent AML clusters (Figure 23A). AML blasts were highlighted individually for better visualization, and cells were colored based on density. *MLL-EDC4* cells seemed to accumulate especially at the HSC to MPP and the pro monocyte to $CD14^{+}$ monocyte compartments (Figure 23B), indicating great heterogeneity within blasts. Additionally, this analysis supports the hypothesis that *MLL-EDC4* AML cells confer a more progenitor-like phenotype. In contrast, cells from *MLLT3-MLL* AML-1 (Figure 23C) and *MLL-ELL* AML (Figure 23E) showed the highest density at the cluster of monocytes, implying that the phenotypes of these tumors are more differentiated. Cells from *MLLT3-MLL* AML-2 were shown to be spread around the myeloid trajectory with high densities in the GMP, promonocyte, and monocyte clusters (Figure 23D). These data integration results are in accordance with outcomes obtained from cell type prediction as described above. This analysis underlines intra-tumoral heterogeneity by the presence of various subgroups of blasts with distinct differentiation states within one AML entity.

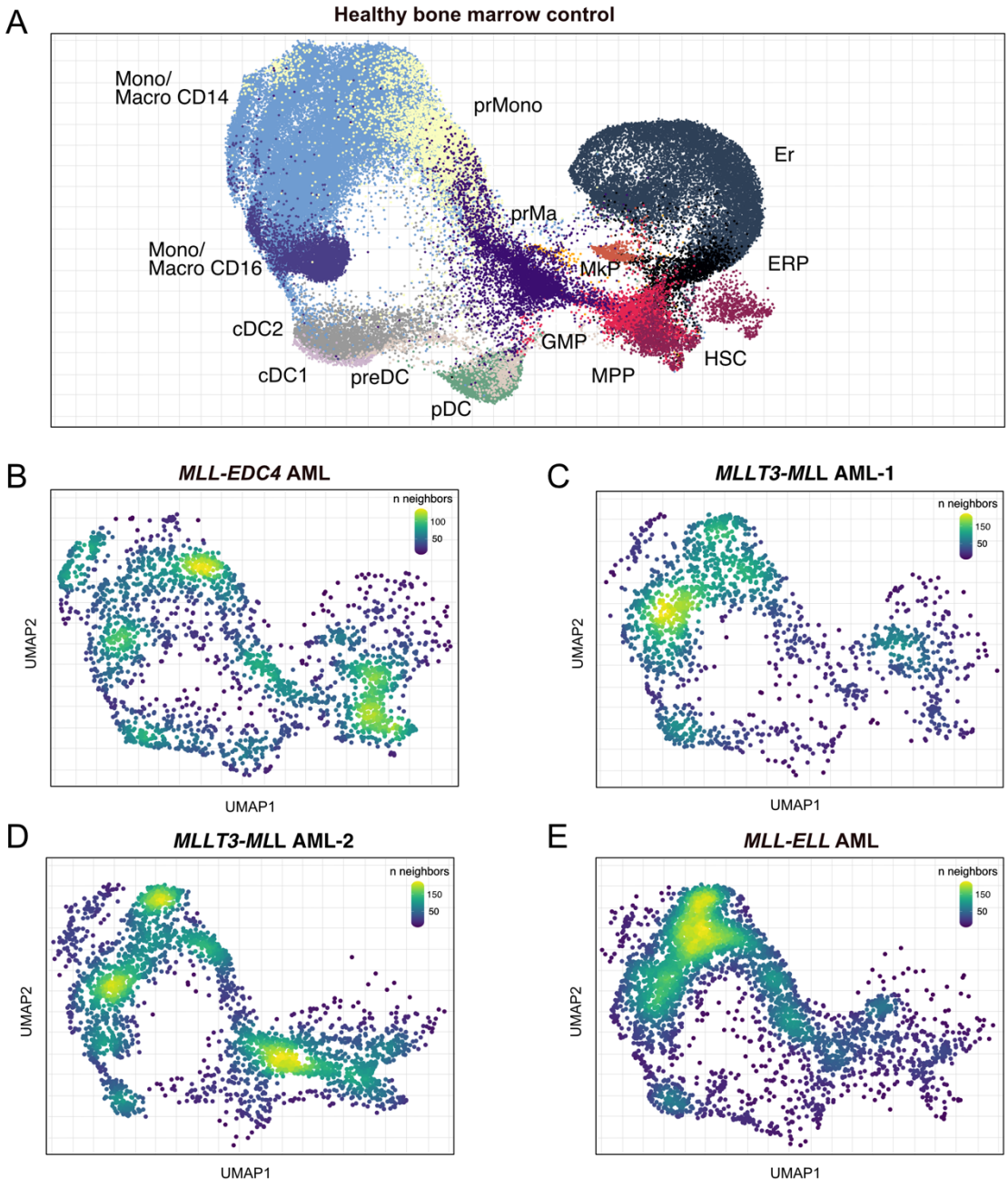


Figure 23

Integration of *MLL-r* AML cells with healthy bone marrow control. (**A**) UMAP embedding of integrated bone marrow samples with all four AML patients colored by cell type. (**B**) Density plot of *MLL-EDC4* AML cells on a UMAP embedding of integrated data. (**C**) Density plot of *MLLT3-MLL* AML-1 cells on a UMAP embedding of integrated data. (**D**) Density plot of *MLLT3-MLL* AML-2 cells on a UMAP embedding of integrated data. (**E**) Density plot of *MLL-ELL* AML cells on a UMAP embedding of integrated data. HSC, hematopoietic stem cell; MPP, multipotent progenitor; GMP, granulocyte monocyte progenitor; prMono, Monocyte; Mono/Macro CD14, Monocyte/Macrophage $CD14^{+}$; Mono/Macro CD16, Monocyte/Macrophage $CD16^{+}$; MkP, Megakaryocyte progenitor; prMa, mast cell precursor; ERP, erythroid progenitor; Er, Erythrocyte; preDC, pre dendritic cell; cDC1, conventional dendritic cell type 1; cDC2, conventional dendritic cell type 2; pDC, plasmacytoid dendritic cell.

3 Multi-omics analysis of AML patients harboring *FLT3*-ITDs

AML drug therapy has the ability to modify dysregulated epigenetic signaling as an important aspect of the disease. Treatment with midostaurin, which displays one of the new target therapy alternatives for *FLT3* aberrations, has a lot of promise for AML patients. However, midostaurin resistance may emerge. Intra-tumoral genetic heterogeneity is common in malignancies and significantly influences tumor growth and therapy response. Although midostaurin is classified as targeted therapy, it also interacts with other kinases. Thus, treatment can impact clonal genetic structures, epigenome, and transcriptome, among other molecular layers. As a result, the multi-omics effects of midostaurin in AML patients with *FLT3*-ITDs during treatment were investigated. Using longitudinal samples collected at diagnosis, remission, and relapse from two patients, I assessed the induced impacts on gene expression, chromatin accessibility, and clonal composition throughout treatment

3.1 Gene expression signature changes during midostaurin treatment

First, the effect of midostaurin treatment in two *FLT3*-ITD AML patients on transcriptomic alterations was studied. A scRNA-seq analysis was performed to dissect intra-tumoral heterogeneity further. In total, 23 210 cells from five timepoints passed stringent quality filtering criteria. Internal tandem duplications lead to constitutive expression of *FLT3* gene and thus *FLT3* expression serves as an excellent tumor marker in these identities (Figure 24C, Figure 24D). The microenvironment from all timepoints clustered together, and AML blasts from the relapse sample clustered with the diagnosis cluster for patient *FLT3*-ITD AML-1 (Figure 24A). Cells were then annotated using SingleR as previously described. Most tumor cells composed of GMPs and mast cell precursors (prMa) at timepoint of diagnosis and erythroblasts and prMas at relapse (Figure 24E). For patient *FLT3*-ITD AML-2, tumor cells clustered individually per timepoint, and the relapse sample was divided into two major subclusters based on their levels of *FLT3* expression (Figure 24B, Figure 24D). However, some overlapping cells were evident between the diagnosis and relapse I cluster. While the diagnosis cluster contained primarily undifferentiated cells, a differentiation trajectory was displayed along the two relapse clusters (Figure 24F). Further analysis of AML blast composition was performed in chapter 5.

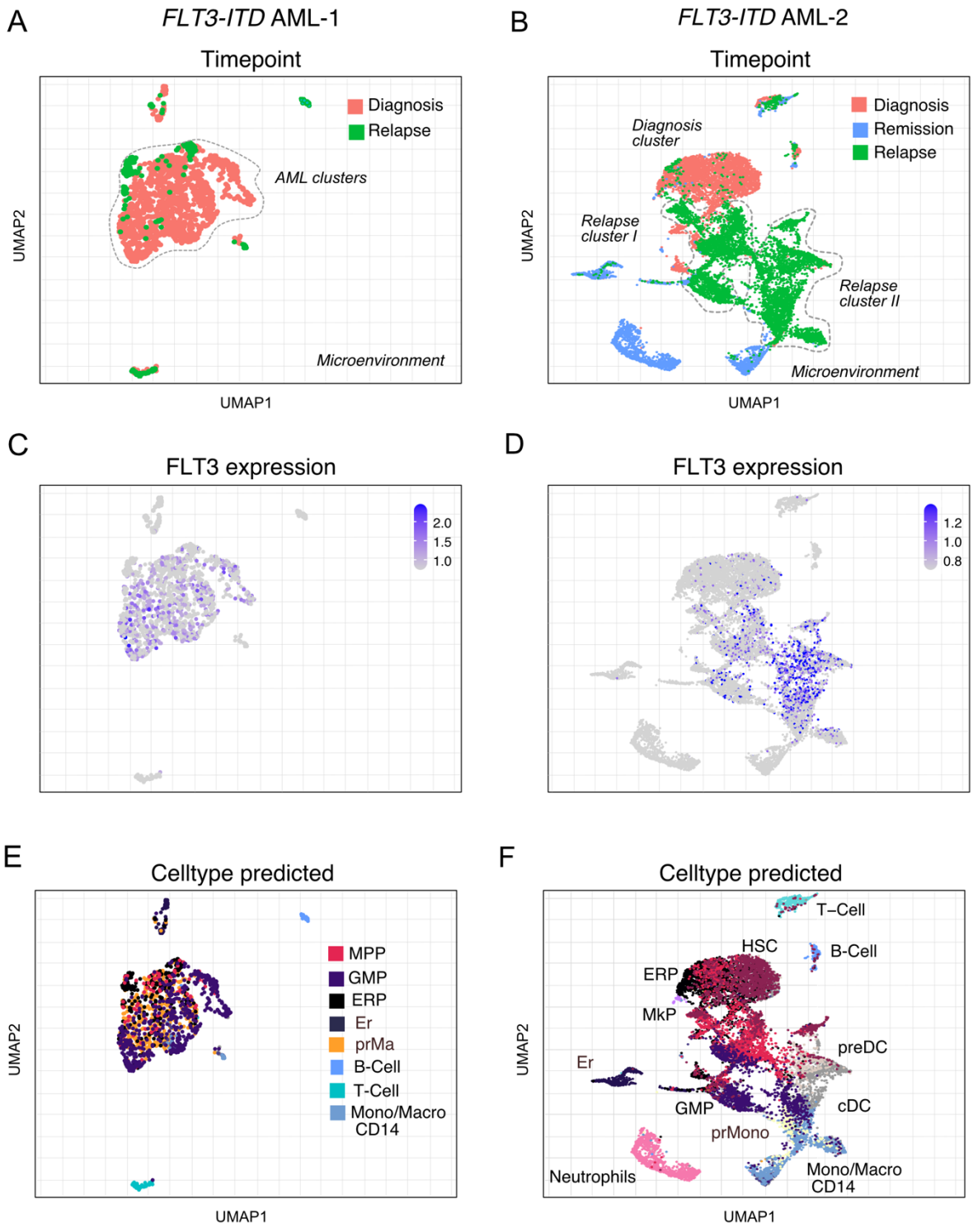


Figure 24

Annotation of *FLT3*-ITD AML samples. (**A**) UMAP embedding of patient *FLT3*-ITD AML-1 colored by timepoint (**B**) UMAP embedding of patient *FLT3*-ITD AML-2 colored by timepoint (**C**) Feature plot of patient *FLT3*-ITD AML-1 depicting FLT3 expression (**D**) Feature plot of patient *FLT3*-ITD AML-2 depicting FLT3 expression (**E**) UMAP embedding of patient *FLT3*-ITD AML-1 colored by predicted cell type (**F**) UMAP embedding of patient *FLT3*-ITD AML-2 colored by predicted cell type. HSC, hematopoietic stem cell; MPP, multipotent progenitor; GMP, granulocyte monocyte progenitor; prMono, Monocyte; Mono/Macro CD14, Monocyte/Macrophage CD14^{+}; ERP, erythroid progenitor; preDC, pre dendritic cell; cDC, conventional dendritic cell type, prMa; mast cell precursor.

Next, I determined the 10 most up and down regulated marker genes per timepoint and visualized their expression in two heatmaps (Figure 25). Gene expression programs showed a dynamic change between diagnoses and relapses for both patients. During diagnosis, especially *STAT* genes were upregulated in *FLT3*-ITD AML-1 (Figure 25A). STAT transcription factors, especially STAT3, STAT5A, and STAT5B, which are downstream of FLT3 were reported to have strongest connections to the most common AML drivers, highlighting their role in AML pathogenesis [297]. At relapse, however, STAT genes were downregulated while TFs such as POU2F2, RELB, FOXA2, NFE2L2, PAX8 were upregulated. As expected, GSEA analysis also highlighted the upregulation of genes involved in STAT signaling when comparing diagnosis to relapse (Figure 26A). Pathways linked to ROS, apoptosis, or the proto-oncogene p53 were upregulated at the time of relapse (Figure 26B).

For patient *FLT3*-ITD AML-2 gene expression signature of diagnosis and relapse cluster II showed large similarities, whereas relapse cluster I showed an opposite pattern (Figure 25B). In addition to upregulation of *STAT* TF, the hematopoietic master regulator *RUNX1* was highly expressed in relapse cluster II. Moreover, molecular regulators of HOXA9, such as *USF2, USF1,* and the proto-oncogene *CREB1* and other *FOX* genes implicated in differentiation, proliferation, and senescence, were significantly expressed [298-301].

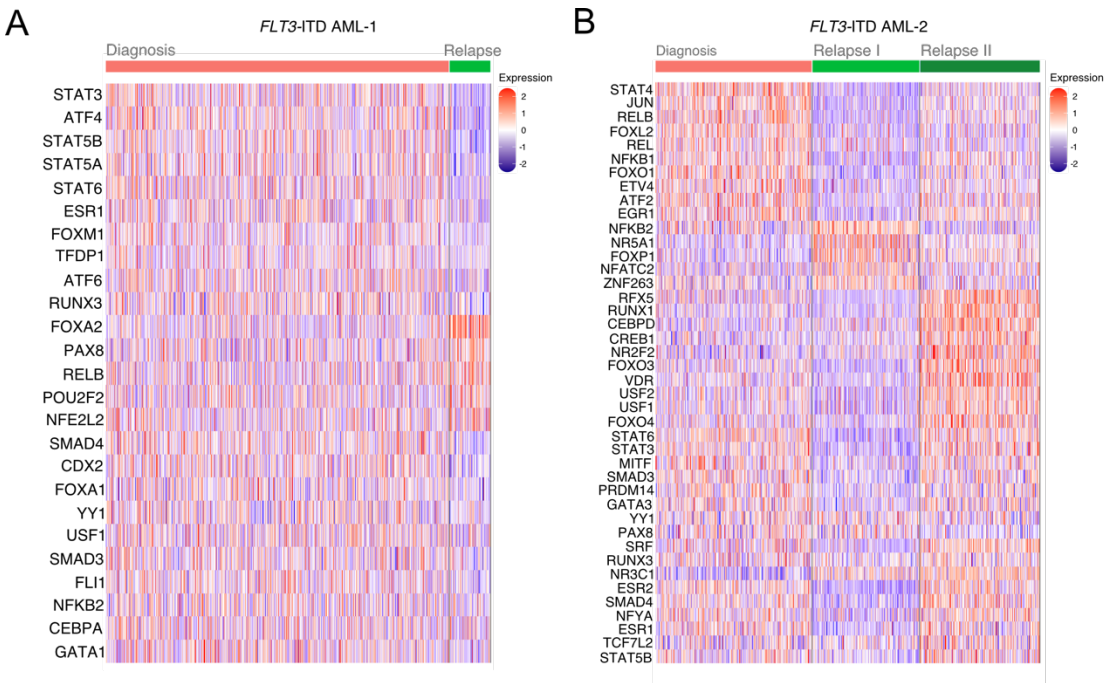


Figure 25

Differential gene expression of *FLT3*-ITD AML samples. (**A**) Clustered single-cell heatmap showing prediction scores for *FLT3*-ITD AML-1 at timepoints of diagnosis and relapse. (**B**) Clustered single-cell heatmap showing prediction scores for *FLT3*-ITD AML-2 at timepoints of diagnosis and relapse.

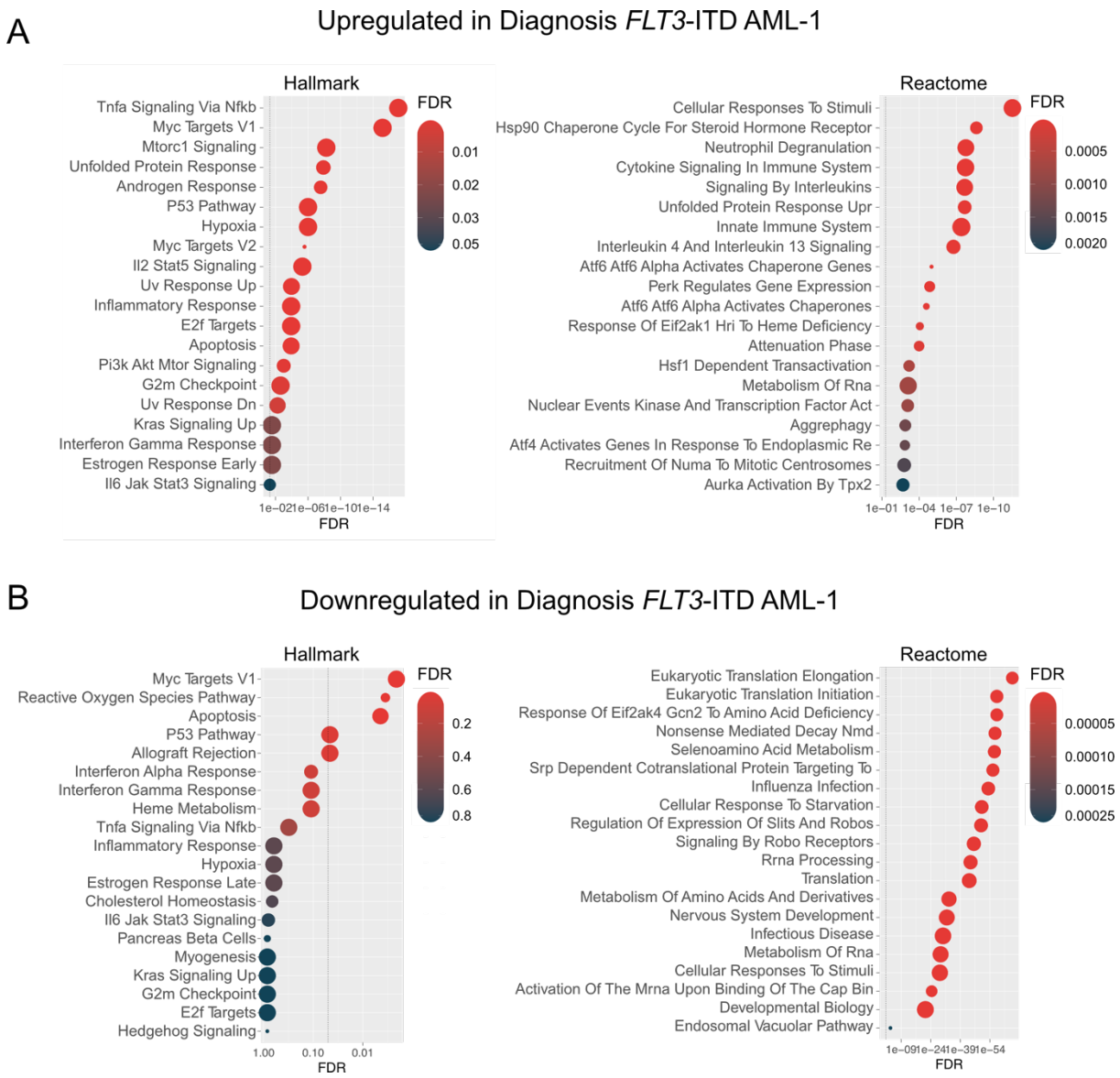


Figure 26

Gene set enrichment analysis of FLT3-ITD AML-1. (**A**) Dot plot of up regulated gene sets of AML cells collected at diagnosis compared to relapse. Gene sets from Hallmark and Reactome were used. (**B**) Dot plot of down regulated gene sets of AML cells collected at diagnosis compared to relapse. Gene sets from Hallmark and Reactome were used. The dashed line depicts significance threshold.

When comparing diagnosis to both relapse clusters, pathways associated, the proto-oncogene *p53* and responses to inflammation were up regulated (Figure 27A). Genes linked to the innate immune system or ROS response, were down regulated. Gene sets such as TNFα signaling via NFκB, mTOR signaling, or STAT signaling, were highly expressed in both timepoints (Figure 27A, Figure 27B). These pathways are linked to STAT signaling and promote cell proliferation and contribute to tumor initiation and progression [302-304].

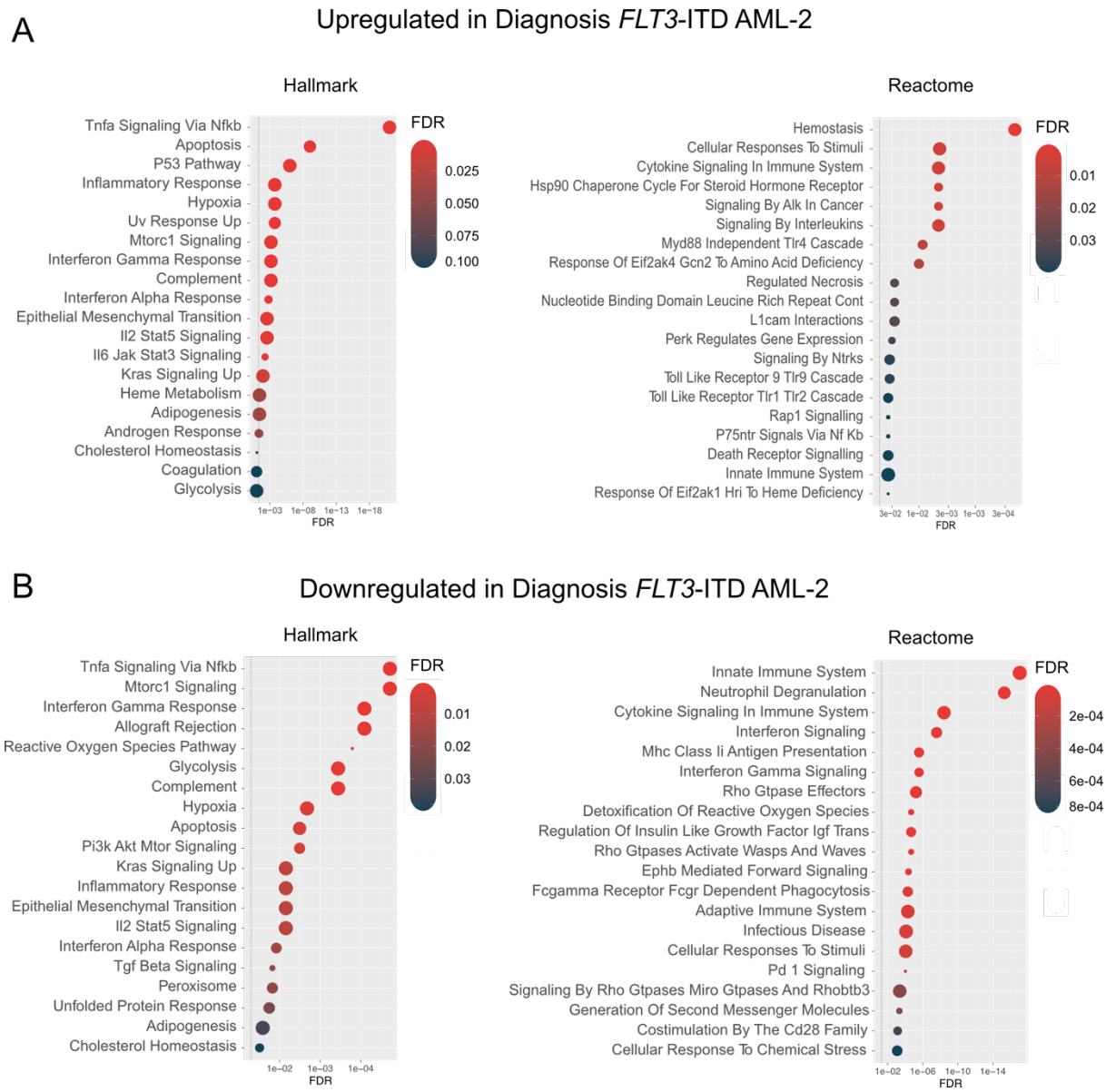


Figure 27

Gene set enrichment analysis of FLT3-ITD AML-2. (**A**) Dot plot of up regulated gene sets of AML cells collected at diagnosis compared to relapse. Gene sets from Hallmark and Reactome were used. (**B**) Dot plot of down regulated gene sets of AML cells collected at diagnosis compared to relapse. Gene sets from Hallmark and Reactome were used. The dashed line depicts significance threshold.

I then compared AML cells from relapse cluster II to relapse cluster I (Figure 28). Again, relapse cluster II displayed upregulation of genes linked to *STAT* such as TNFα signaling via NFκB, mTOR signaling, or STAT signaling (Figure 28A). Relapse cluster I however, showed upregulation of pathways linked to *p53*, ROS, *MYC* targets, and eucaryotic translation initiation and elongation (Figure 28B).

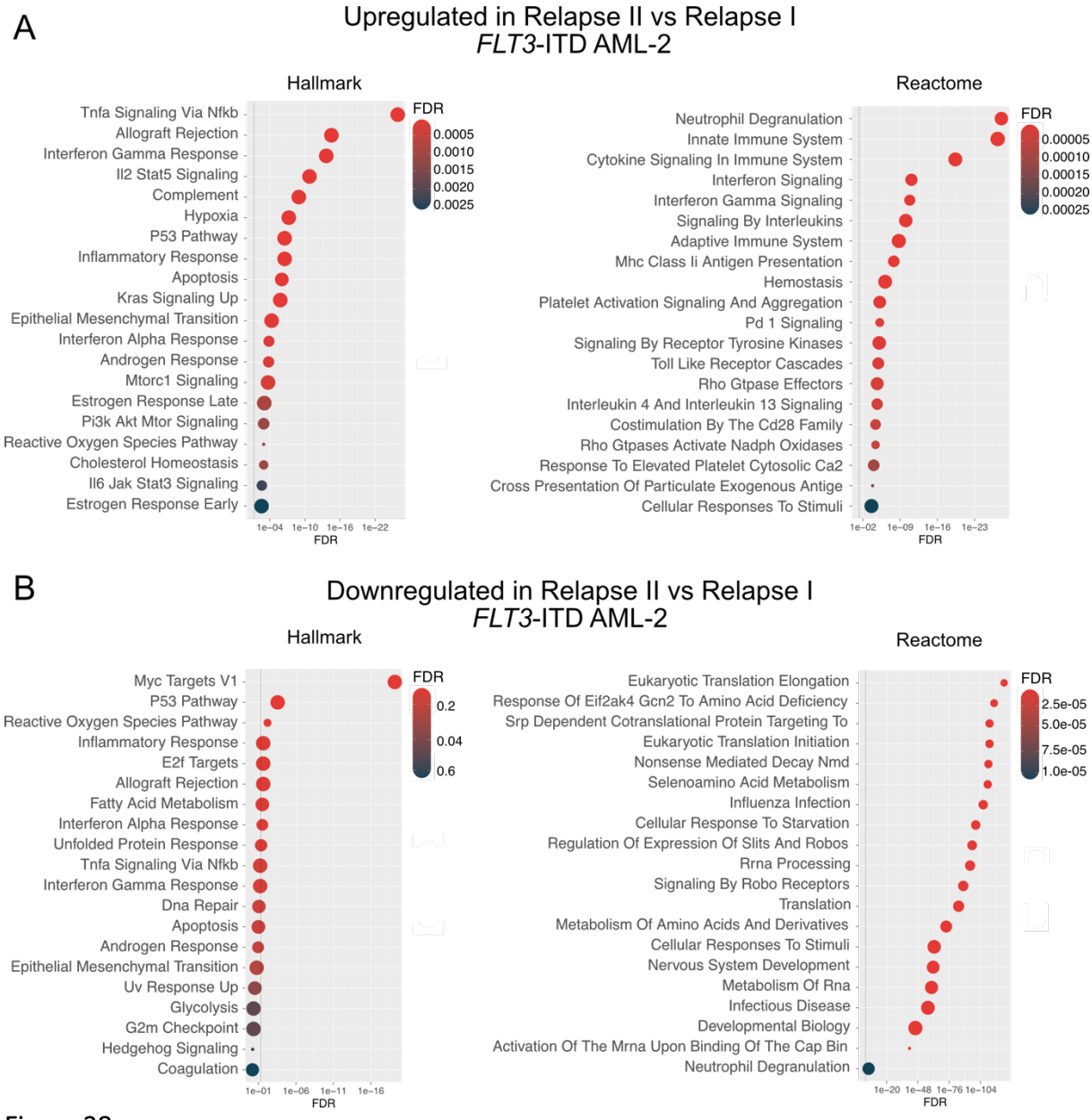


Figure 28

Gene set enrichment analysis of FLT3-ITD AML-1. (**A**) Dot plot of up regulated gene sets of AML cells from relapse cluster II compared to relapse cluster I. Gene sets from Hallmark and Reactome were used. (**B**) Dot plot of down regulated gene sets of AML cells from relapse cluster II compared to relapse cluster I. Gene sets from Hallmark and Reactome were used. The dashed line depicts significance threshold.

Analysis of transcription factor analysis highlighted many TFs involved in JAK/STAT signaling, NFκB signaling, hematopoiesis, cell cycle, proliferation, leukemogenesis and tumor progression in both *FLT3*-ITD patients (Figure 29). At the time of relapse in *FLT3*-ITD AML-1, most TFs showed decreased activity compared to diagnosis (Figure 29A). However, single TFs are linked to leukemogenesis or NFκB signaling had enhanced activity.

TF activity signature in patient *FLT3*-ITD AML-2 drastically changed during midostaurin treatment (Figure 29B). Cluster relapse I showed a completely opposite pattern compared to diagnosis, with most TF activity being down regulated. Relapse cluster II resembled more TF signature as observed in diagnosis, but genes involved in STAT signaling, cell fate and hematopoiesis showed significantly higher activity.

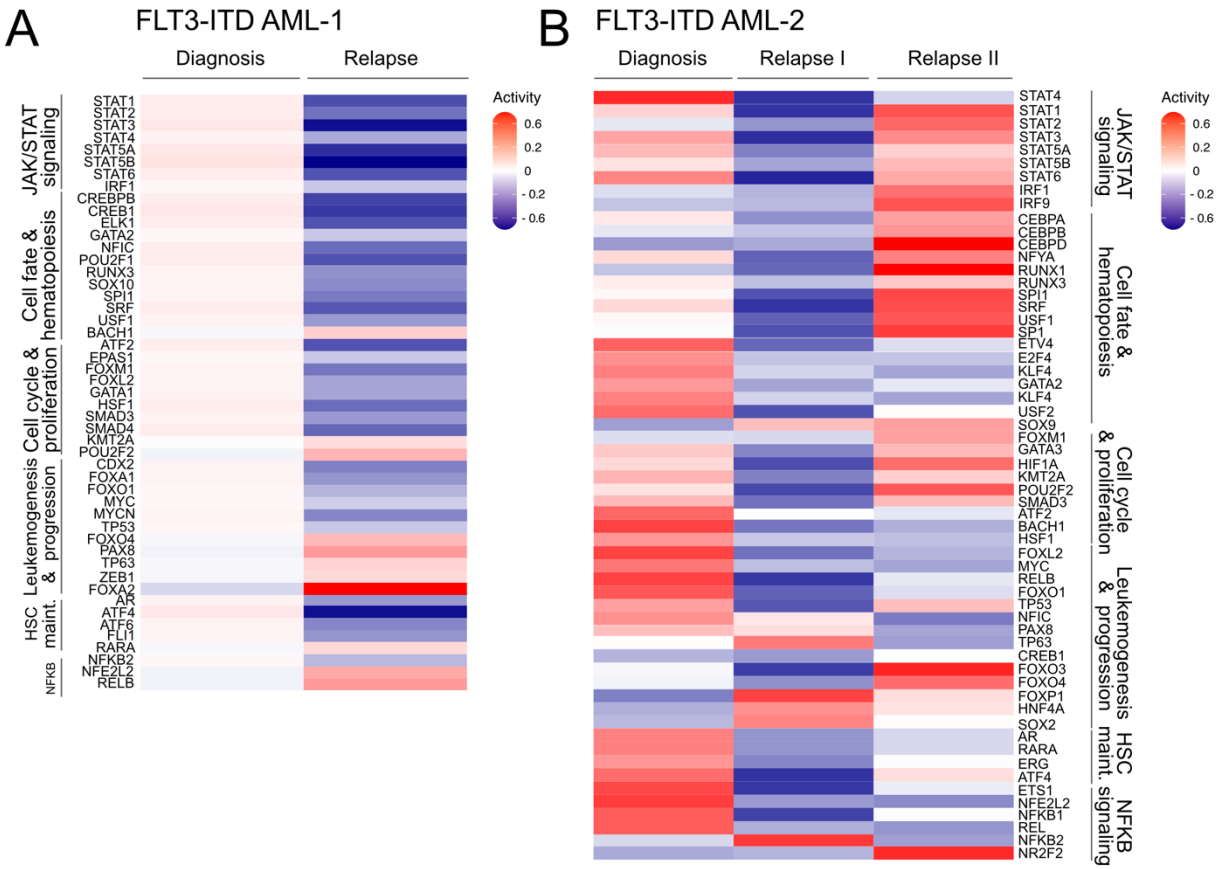


Figure 29

Transcription factor activity in patients with *FLT3*-ITD. (**A**) Heatmap depicting TF activities at diagnosis and relapse in patient *FLT3*-ITD AML-1 inferred from scRNA-seq data. (**B**) Heatmap depicting TF activities at diagnosis and relapse in patient *FLT3*-ITD AML-2 inferred from scRNA-seq data.

3.2 Chromatin accessibility changes during midostaurin treatment

Chromatin accessibility measured by scATAC-seq gives insights on the activity states of cis-regulatory elements and makes it possible to connect them to TF occupation [194]. Here, I examined the enrichment of TF binding motifs in significant differentially accessible peaks between of relapse compared to diagnosis. For both patients carrying *FLT3*-ITDs, scATAC-seq was performed for samples collected at diagnosis and relapse. For *FLT3*-ITD AML-1, diagnosis and relapse formed two distinct clusters (Figure 30A). Subsequently, DNA sequence motif analysis was performed by Signac, detecting overrepresented motifs of differentially accessible peaks between timepoints (Figure 30B). At diagnosis, STAT1/STAT2 and SP transcription factors were overrepresented. Overrepresentation of the SP TF family was even more prominent at the timepoint of relapse.

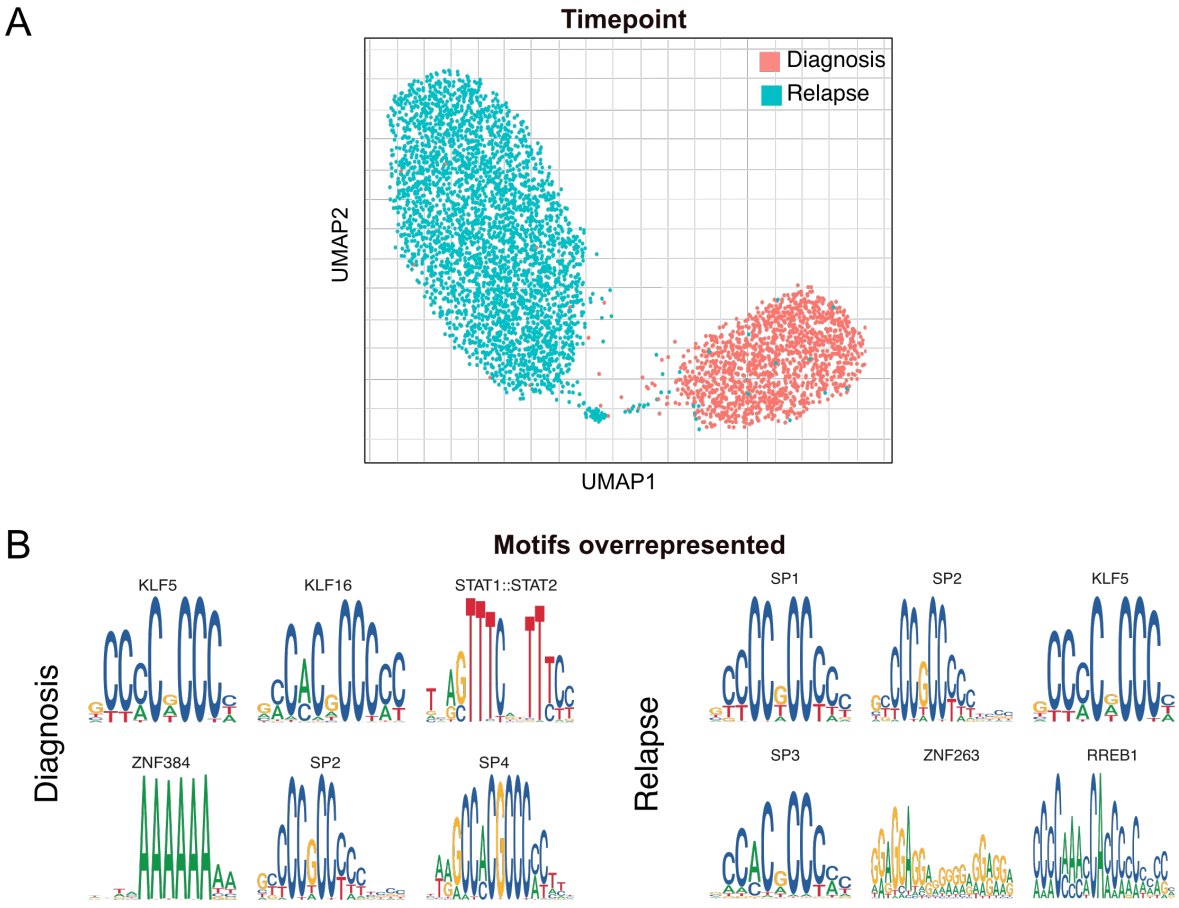


Figure 30

Chromatin accessibility in patient *FLT3*-ITD AML-1 determined by Signac. (**A**) UMAP embedding colored by timepoint (**B**) Overrepresented motifs of at diagnosis and relapse derived from the JASPAR database

Data analysis using ArchR also resulted in two divided clusters for diagnosis and relapse and one cluster comprising cells from both timepoints (Figure 31A, Figure 31B). Unique marker peaks within the whole dataset only generated 25 significant features (Figure 31C). Thus, pairwise testing between the relapse (C3) and diagnosis (C2) cluster without cluster 3 was performed. Differential testing identified ~0.5% upregulated and 0.01% downregulated significant marker features in relapse compared to diagnosis (Figure 31D). Then motifs enriched in these peaks were investigated. The TF motifs found in the Signac analysis, were also present in the analysis using Signac. However, Signac predicted other motifs such as SMAD5, WT1 or DNMT1 to be more upregulated in relapse compared to diagnosis with higher p values than the SP1 motifs (Figure 31E). While Signac mainly found KLF motifs upregulated in diagnosis, ArchR found motifs such as SOX10, BCL, GATA, HOXA3 and RUNX3 (Figure 31F). Differences in motif analysis obtained from Signac and ArchR might be explained by the usage of different motif databases, JASPAR and CIS-BP, respectively.

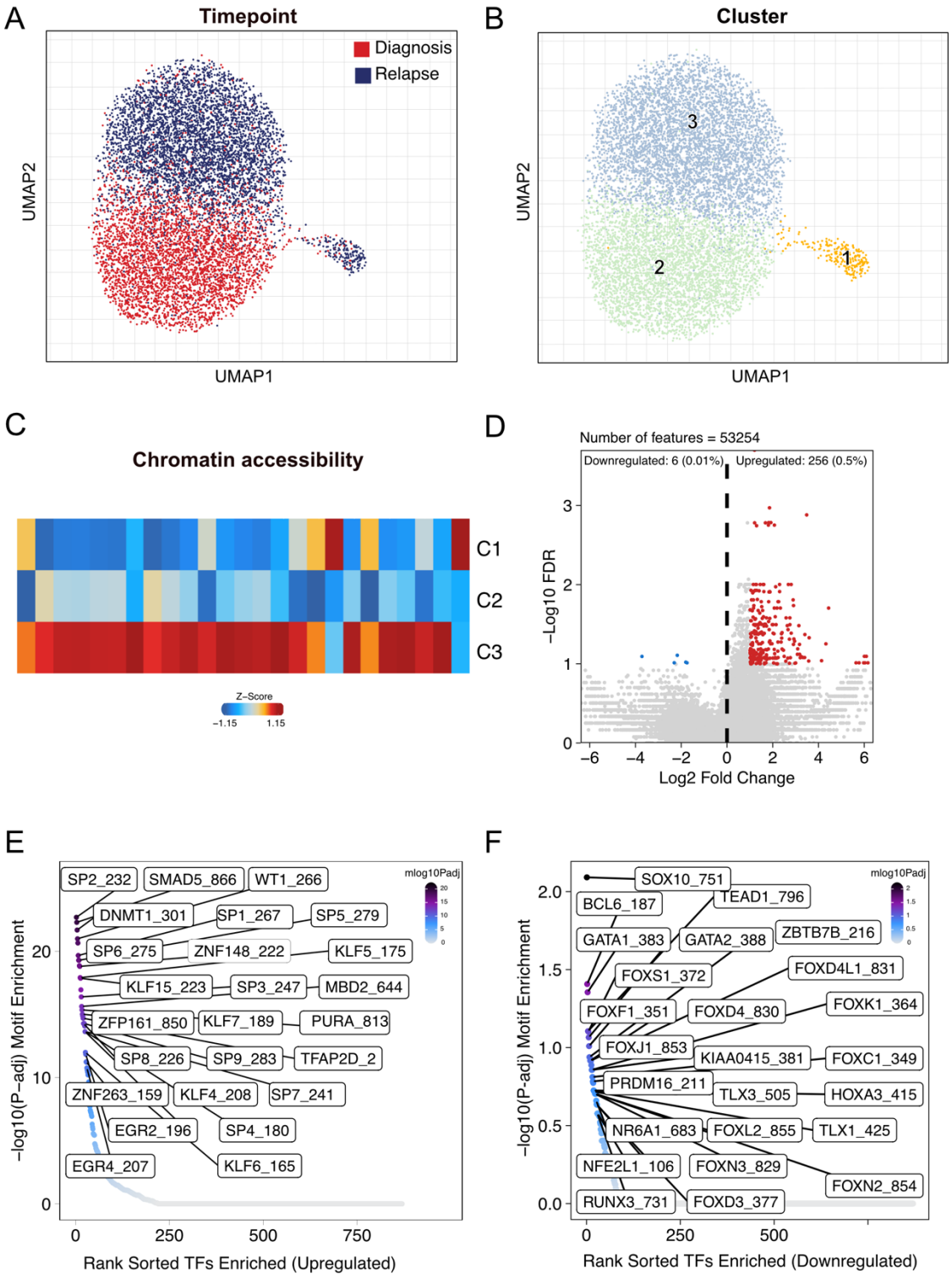


Figure 31

Chromatin accessibility in *FLT3*-ITD AML-1 determined by ArchR. (**A**) UMAP embedding colored by timepoint (**B**) UMAP embedding colored by clusters (**C**) Peak marker heatmap (**D**) Volcano plot of differential peaks (relapse vs diagnosis) (**E**) Upregulated motifs in relapse obtained from CIS-BP database (**F**) Downregulated motifs in relapse obtained from CIS-BP database

For *FLT3*-ITD AML-2, the relapse sample was divided into two distinct clusters, relapse cluster I was in proximity to diagnosis (Figure 32A). Additionally, a small cluster containing cells from both timepoints was present. Motif analysis revealed an overrepresentation of motifs downstream of FLT3 signaling, such as TFs from CEB/P family, IRF1, STAT1/2, and GATA for relapse cluster II (Figure 32B).

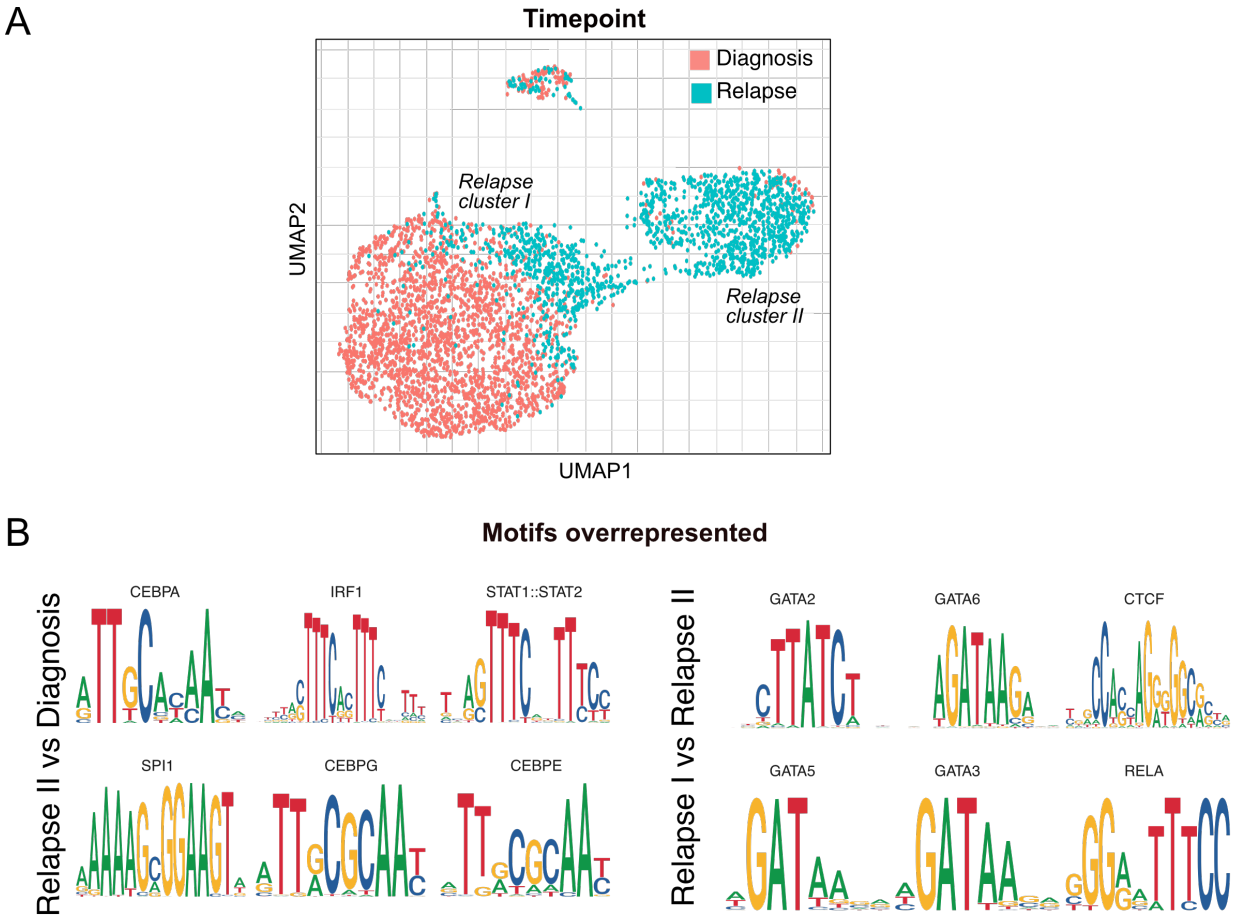


Figure 32
Chromatin accessibility in patient *FLT3*-ITD AML-2 determined by Signac. (**A**) UMAP embedding colored by timepoint (**B**) Overrepresented motifs of at diagnosis and relapse derived from the JASPAR database

To further evaluate and integrate with scRNA-seq data, ArchR was employed. Again, cells from both timepoints were visualized using UMAP, forming two primary and one small cluster (Figure 33A). The upper main cluster consisted of cells both from diagnosis and relapse (relapse cluster I) that seemed to follow a trajectory, while the second main cluster only comprised cells from the relapse sample (relapse cluster II). The minor cluster contained cells from both timepoints and did not seem to be in a specific order. Cell types and relapse

clusters were annotated by integration with scRNA-seq data (Figure 33C). This integration compares the scATAC-seq gene score matrix with the scRNA-seq gene expression matrix to directly link cells from scATAC-seq with cells from scRNA-seq. Based on this integration, the minor cluster represents non-malignant microenvironment while the other two main clusters represent leukemic blasts. Relapse cluster II comprised mainly differentiated monocytes, while cells from relapse cluster I were more undifferentiated MPPs. Marker peaks were identified using pairwise differential testing between relapse cluster II vs relapse cluster I and diagnosis (Figure 33D).

In total, 192 635 peaks were detected, of which ~1 % were significantly upregulated and 0.5 % downregulated in relapse cluster II. After differential peak set identification, motif enrichment analysis was performed to predict what TFs might mediate the binding events responsible for the accessible chromatin sites.

Peaks that gained accessibility in relapse cluster II compared to the other main cluster showed an enrichment of binding motifs for BACH1/2, C/EBP, BCL-X, IRFs, and RUNX1 (Figure 33E). RUNX1 and C/EBP motifs were previously reported to be the most common binding motifs in *FLT3*-ITDs [305], and BACH1 [306] were linked to mediating *FLT3*-dependent gene expression. While BCL11A has been shown to control FLT3 expression in early hematopoietic progenitors [307], IRFs are common mediators of receptor signaling. The highest enriched motifs in sites with lost accessibility were motifs from the Krüppel-like family of transcription factors (KLFs).

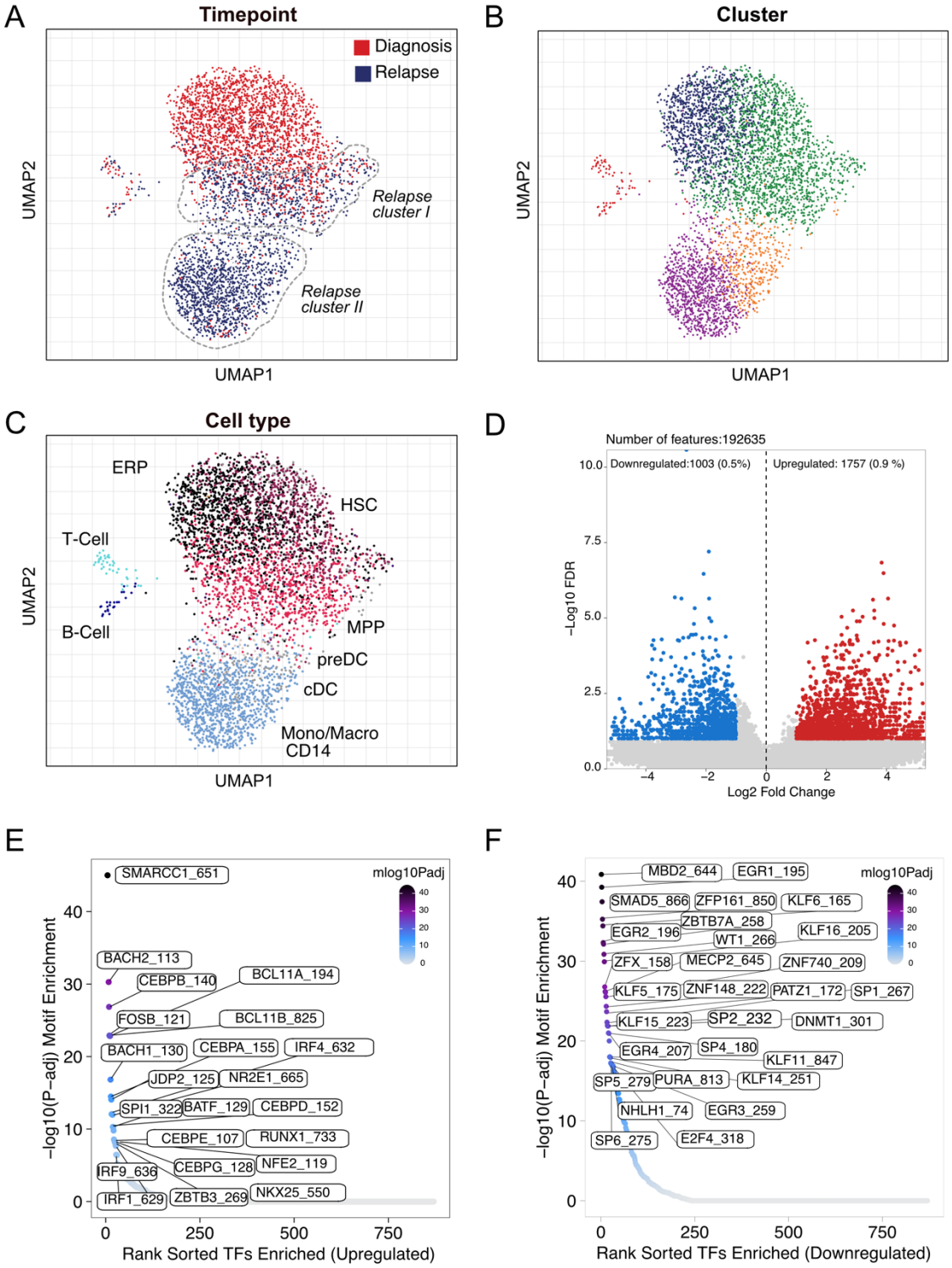


Figure 33

Chromatin accessibility in *FLT3*-ITD AML-2 determined by ArchR. (**A**) UMAP embedding colored by timepoint (**B**) UMAP embedding colored by clusters (**C**) UMAP embedding colored by predicted cell type (**D**) Volcano plot of differential peaks (**E**) Upregulated motifs obtained from CIS-BP database (**F**) Downregulated motifs obtained from CIS-BP database

3.3 Resolution of clonal evolution in midostaurin-treated AML patients by targeted scDNA-seq

The malignant cell population is frequently very heterogeneous, harboring several distinct subclones. Clonal dynamics might alter over time, notably during treatment and resistance development [308]. Using targeted scDNA-sequencing, I evaluated the clonal evolution of two AML patients carrying *FLT3*-ITDs, that were treated with midostaurin. *FLT3*-ITD AML-1 was shown to carry five different subclones with distinct *FLT3*-ITDs, four harboring an additional *TET2* mutation (Figure 34A). However, this patient did not show great changes in clonal diversity during treatment, and variant alleles frequencies were not altered significantly.

For patient *FLT3*-ITD AML-2 the highest proportion of subclones harbored mutations in *BCOR* (chrX:39933339:A/G), *DNMT3A* (chr2:25457242:C/T), *KDM6A* (chrX:44833841) and an ITD in *FLT3* (chr13:28608271) (Figure 34B). Another smaller clone only carried mutations in *BCOR* and *KDM6A* but not in *DNMT3A* and *FLT3*. This indicates that the larger clone developed from, the smaller one, and the additional *DNMT3A* and *FLT3* mutations presented a clonal advantage. At remission, all subclones carrying an additional *FLT3* mutation were eradicated while the small clone expanded until it represented more than 98% of cells analyzed. However, at relapse, subclones seemed to develop resistance against midostaurin and clonal dynamics shifted rigorously towards subclones harboring a *FLT3*-ITD replacing the large clonal population during remission. At relapse, the *FLT3*-ITD clone expanded again, indicating an acquired treatment resistance.

This prompted me to investigate these results further. Therefore, the results of scDNA-seq were compared to getITD analysis, a tool to identify ITDs in bulk performed by collaborators in Ulm. The comparison revealed that this patient harbored four different ITDs, whereas only of them was fully covered by the used myeloid panel (Table 27).

In conclusion, targeted scDNA-seq was used to detect rare subclones and resolve mutational co-occurrences within the same cell during midostaurin treatment. Although scDNA-seq could detect small sub clones, issues with the proper detection of *FLT3*-ITDs were evident. To further enhance detection efficiency, I generated a custom panel with a focus on targeting *FLT3*-ITDs.

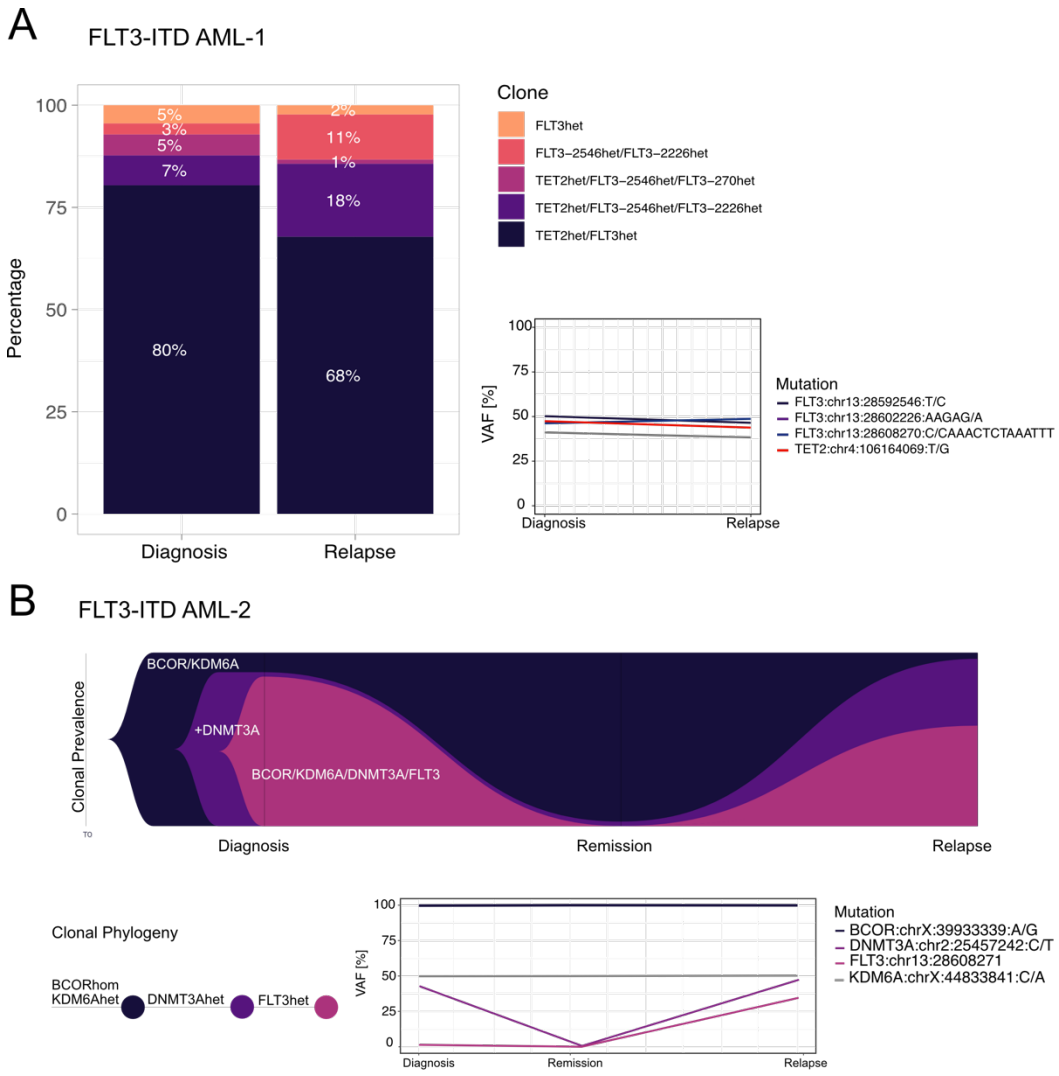


Figure 34

Clonal evolution during midostaurin treatment in *FLT3*-ITD patients. (**A**) Left: Dodge-plot depicting the clonal composition of AML cells before treatment and at relapse. Right: Aggregated variant allele frequency (VAF) calculated from single-cell data visualized as line plot. (**B**) Top: FISH-plot depicting the clonal composition of AML cells before treatment, during remission, and at relapse. Bottom left: Clonal phylogeny tree of clonal evolution. Bottom right: Aggregated VAF calculated from single-cell data visualized as line plot.

Table 27 Detected ITDs in *FLT3*-ITD AML-2 in bulk

ITD	ITD size	Insertion site	Timepoint	Covered by myeloid panel
ITD 1	141bp	28608129	Diagnosis, Remission, Relapse	No
ITD 2	197bp	28608110	Diagnosis	No
ITD 3	36bp	28608225	Diagnosis	Yes
ITD 4	39bp	28608277	Diagnosis, Relapse	Partially

3.4 Design of a custom-targeted scDNA-seq panel to optimize coverage of *FLT3*-ITDs in AML patients

To overcome the problem that some *FLT3*-ITDs present in the samples used, a custom DNA panel was designed in collaboration with Mission Bio. To ensure to cover as many ITDs as possible, a list of unique ITDs found in 110 patients from the University clinics in Freiburg was generated (see Appendix). In total, 247 distinct ITDs ranging from 9 to more than 200 bp were found. To successfully call an ITD, the whole ITD must be covered either by read 1 or read 2 of the designed amplicons [309] (Figure 35). ITDs with an overlap between the primer region or a gap between read 1 and 2 will only lead to partial amplicon coverage, impeding scDNA-seq analysis.

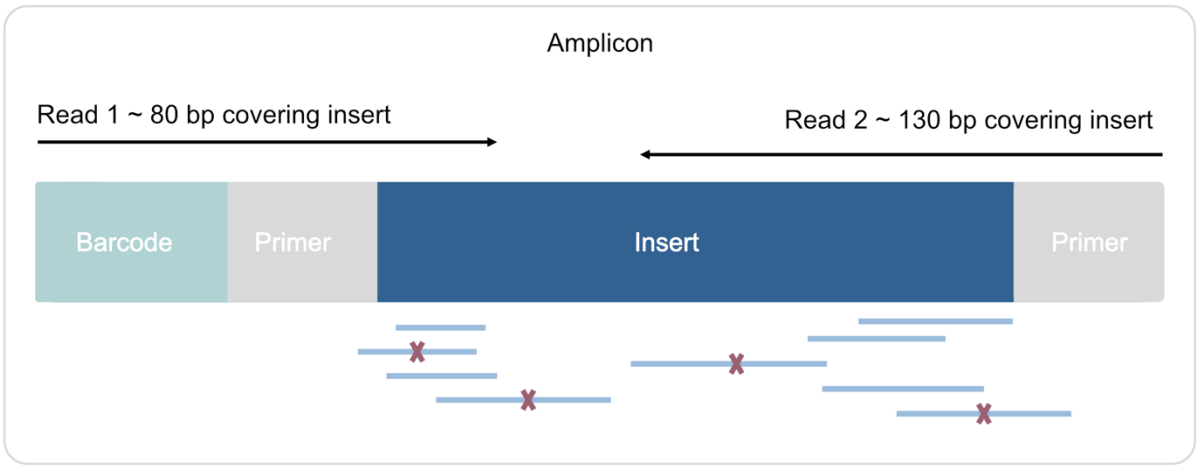


Figure 35
Schematic overview of amplicon. Amplicons for targeted scDNA-seq consist of a barcode and primer site on the 3' end and a primer site on the 5' end. Around 80 bp of the insert are covered by read 1 and 130 bp by read 2. ITDs only are covered when they do not overlap between primer and insert region or lie within the gap of primers 1 and 2.

Analysis can fail to identify amplicons based on two categories: amplicon or sequence-related targets (Table 28). In the first category, prioritization can frequently save the most-relevant targets. The second category is associated with the target's sequence. It poses a problem because any potential rescue requires lowering design constraints (e.g., GC percentage), which may influence amplicon performance. As already mentioned above, amplicons are not detected if primers cover region. To overcome this, new amplicons can be created by adding additional targets near the original target. Center gap regions usually occur for amplicons longer than 250bp. Since ITDs can vary tremendously in size, large ITDs are at risk of lacking coverage. However, these targets could be recovered using 250bp PE

sequencing instead of the recommended 150bp PE. In some cases, the addition of one more target could help to solve this issue as well.

Additionally, coverage will be lost when targeting complete genes or exons due to primer regions and gaps between neighboring amplicons. This happens due to Mission Bio's design of tiled amplicons that never overlap. To ensure that the most desired targets are not missed, targets should be broken into smaller parts. Usually, for all these amplicon-related issues, there are easy solutions to recover the target. However, recovering targets that are missed due to sequence-related characteristics, recovery is more challenging. Primers for scDNA-seq are limited to a GC content between 27-62%. Regions outside this range are harder to amplify using PCR, and their resulting amplicons are more likely to underperform regarding sequencing depth. Although the design of amplicons with a GC content of up to 74% is possible, the performance of these amplicons will be poor. Sequences at centromeres or telomeres, often display the base N instead of A, T, C, G making them so-called masked regions. If a single target has a masked zone, the entire panel may fail to produce, rather than showing as missing coverage. Masked regions can be restored by supplying missing sequence information with a custom reference genome. The same solution applies to homologous regions. Additionally, standard sequencing cannot generate amplicons in highly repetitive regions. To address this, 250bp PE sequencing must be used instead of 150bp. Then the new custom panel was designed to miss as few targets while keeping a number of amplicons and resulting sequencing costs as low as possible.

Table 28 Interpreting missed targets

Category	Reason	Solution
Amplicon related	Primer region	Adding new target
	Center gap region	250bp PE-sequencing, Adding new target
	Amplicon tiling	Break up longer targets
Sequence related	High GC content	Targets with GC content >74% can't be rescued
	Masked region	Custom reference genome
	Repetitive region	250bp PE-sequencing
	Homologous region	Custom reference genome

All 247 distinct ITDs found in the 100 patients were in a small range covering exome 14 – intron 14 – exome 15 (see Appendix). When comparing the commercially available myeloid to the custom panel, the intronic region between exome 14 and 15 lacks coverage of an amplicon (Figure 36). By adding one amplicon in this region, 70-95% of ITDs from the patient list can be tracked, depending on the selection of sequencing chemistry. In addition, 8 amplicons to target 33 more hot spot mutations in the *FLT3* genes were designed. In summary, the new custom panel comprises most amplicons from the myeloid panel, 9 amplicons to detect *FLT3*-ITDs, and selected hot-spot mutations for *ANKRD26, ASXL2, BCORL1, CEBPA, DDX41,* and *ZBTB7A* (Table 29). This panel will be used for a future study to determine treatment-relevant off-target effects and differences in molecular events induced by *FLT3* inhibition either by midostaurin or gilteritinib.

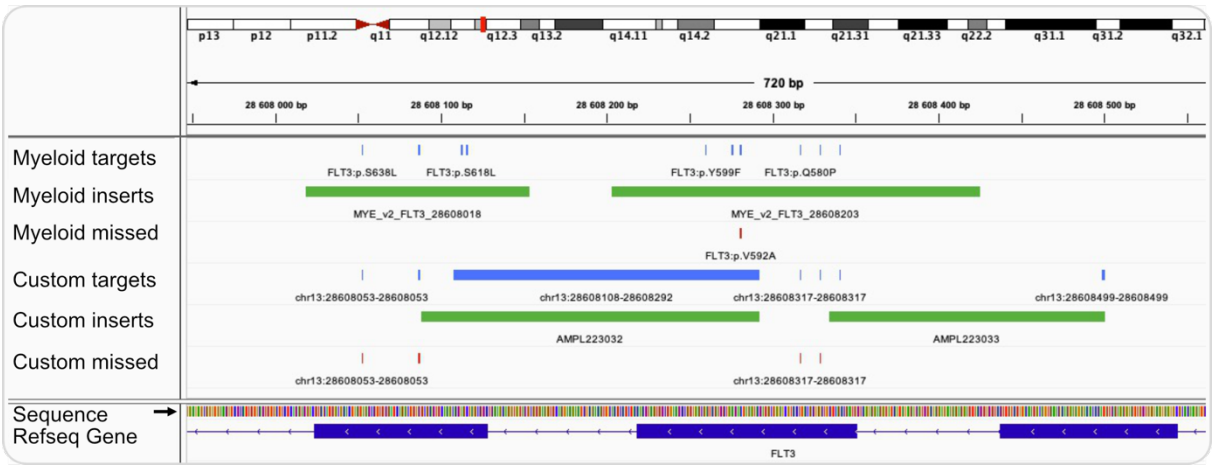


Figure 36

Amplicon design for a custom scDNA-seq panel. Targets (blue), inserts (green), and missed targets (red) are shown in IGV browser both for the myeloid and custom panel at location of exon 14–exon 16 in the *FLT3* gene.

Table 29 Genes covered in the custom panel

Genes			
ANKRD26	DNMT3A	KMT2A	SETBP1
ASXL1	ETV6	MPL	SF3B1
ASXL2	EZH2	MYC	SMC1A
BCOR	FLT3	NF1	SRSF2
BCORL1	GATA2	NPM1	STAG2
BRAF	IDH1	NRAS	TET2
CALR	IDH2	PHF6	TP53
CBL	JAK2	PPM1D	U2AF1
CEBPA	KDM6A	PTPN11	WT1
CSF3R	KIT	RAD21	ZBTB7A
DDX41	KRAS	RUNX1	ZRSR2

4 Multi-omics analysis of IDH^{mut} in AML

IDH genes play critical roles in the Krebs cycle and in maintaining cellular homeostasis. Mutations in *IDH1/2* lead to the conversion of α-KG to the oncometabolite R-2HG, which inhibits histone lysine demethylases and DNA demethylases, resulting in large-scale epigenetic alterations. *IDH^{mut}* inhibitors that counteract hypermethylation have recently emerged as promising clinical studies or have been licensed for the treatment of AML and glioblastoma [124,310-312]. Here, I investigated the effects of targeted *IDH* inhibitors, namely BAY1436032, AG-120, and AG-221, either on myeloid leukemia cell lines or in one primary AML patient harboring an *IDH1* mutation. For this patient, longitudinal samples were collected at diagnosis, remission, and relapse.

4.1 Partial reversion of epigenetic changes caused by *IDH1^{mut}* with BAY1436032

AML cells of one patient carrying an *IDH1* R132C mutation and one patient with IDH WT were isolated from peripheral blood. Both were treated *ex vivo* with the hypomethylating pan-mutant-*IDH1* inhibitor BAY1436032 over a period of 14 days [124]. As a negative control, samples were treated with DMSO for 14 days. Upon inhibitor treatment, cells with *IDH1^{mut}* showed a global decrease in H3K4me3, H3K6me3, H3K9me3, and H3K27me3 whereas this phenomenon was not present in *IDH^{WT}* cells treated with inhibitor. This prompted me to evaluate reversible and self-sustained epigenetic changes caused by mutant *IDH*1 on chromatin accessibility using ATAC-seq. Dr. Lara Klett provided frozen viable cells that had already been treated with the inhibitor. ATAC-seq libraries were processed in technical duplicates. Apart from technical replicate 1 of *IDH1^{mut}* +DMSO, samples had FRiP-scores between 12-23%, and more than 100 000 peaks were called (Figure 37). Due to the low quality of technical replicate 1 of *IDH1^{mut}* + DMSO, technical replicate 2 of *IDH1^{mut}* + DMSO was split into two pseudoreplicates and used for further analysis. In all four treatment conditions, high chromatin accessibility sites were discovered. The ChromHMM [313] annotation by ENCODE [314], which is based on the myelogenous leukemia cell line K562, displayed that these locations mostly appeared at transcription start sites or enhancers.

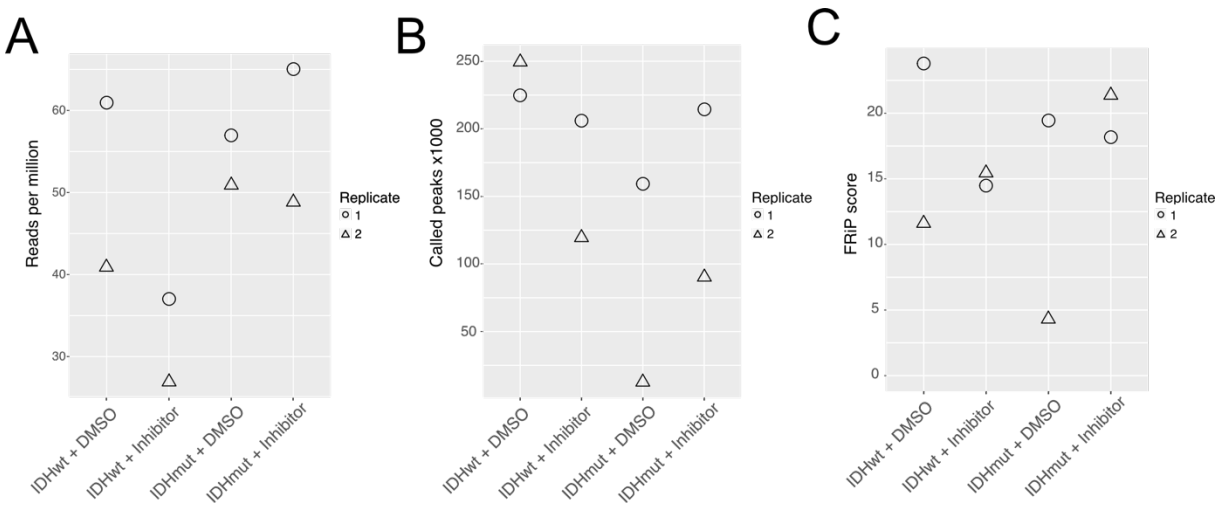


Figure 37

Quality control parameters for ATAC-seq pipeline. (**A**) Sequencing reads per million (**B**) The number of called peaks x1000 (**C**) FRiP score for each replicate.

When *IDH1*^{mut} cells were compared to wild-type cells, several regions showed enhanced accessibility, which were reversed when the mutant-*IDH1* inhibitor was used to restore the wild-type level (Figure 38A). However, other regions displayed irreversible patterns of accessibility after treatment with BAY1436032 (Figure 38B). As expected, differential binding analysis using DiffBind [234] showed that treatment with BAY1436032 did not alter binding affinity in *IDH1* wild-type samples (Figure 38C), whereas *IDH1*^{mut} samples showed partial reversibility when compared to their counterparts treated with DMSO. *IDH1* mutated samples treated with DMSO showed the highest correlation with *IDH1*^{mut} treated with inhibitor (Figure 38D). Nevertheless, *IDH1*^{mut} treated with inhibitor also yielded high correlations with *IDH1* wild-type samples. For a better comprehension and more reliable identification of significant changes in accessibility caused by *IDH1*^{mut} or the inhibitor's action, more biological and technical replicates would be required. Nevertheless, this analysis demonstrates the ability of ATAC-seq to map the chromatin state of primary human AML patients and the potential of BAY1436032 to reverse epigenetic changes caused by mutant *IDH*1 on chromatin accessibility.

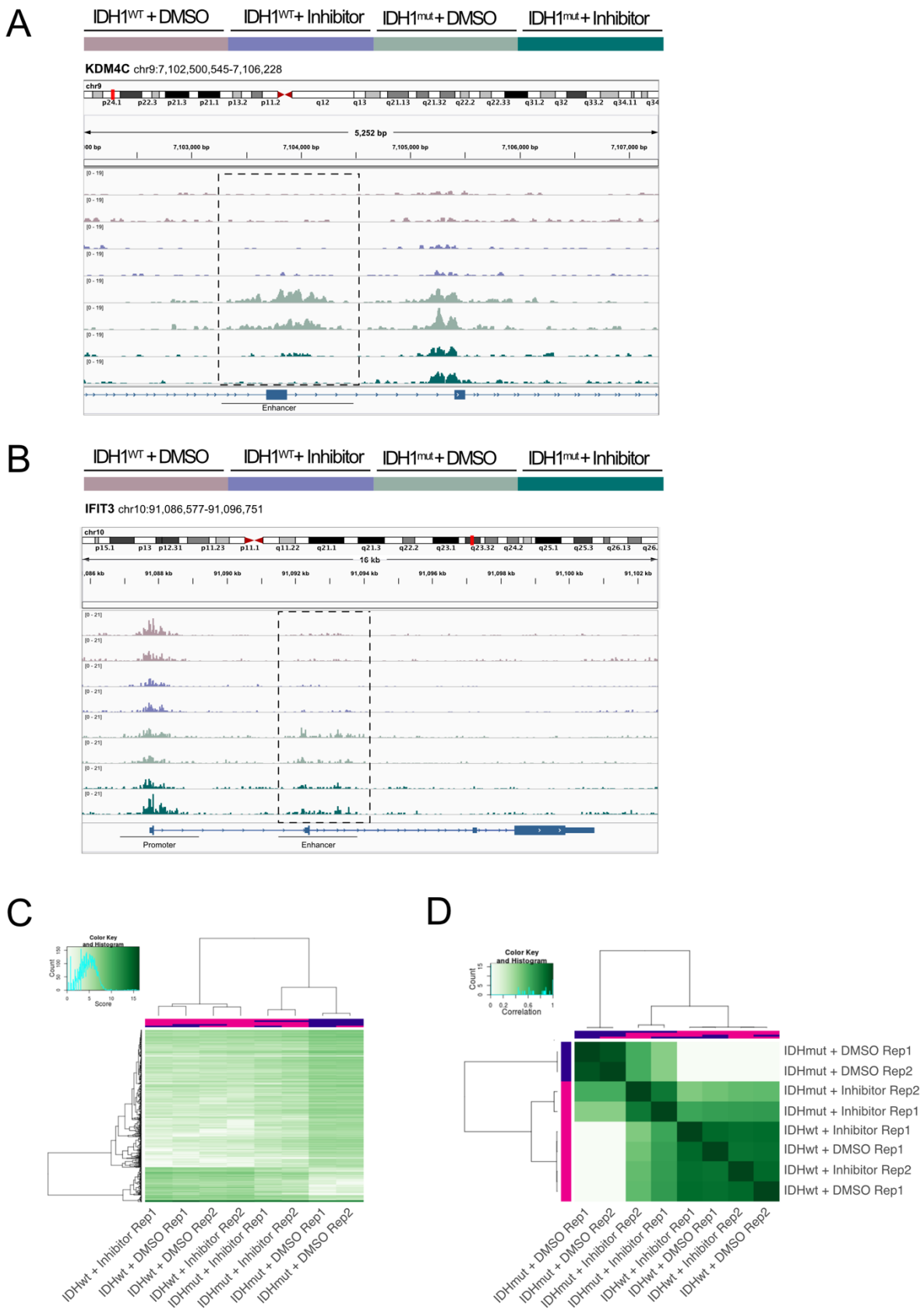


Figure 38

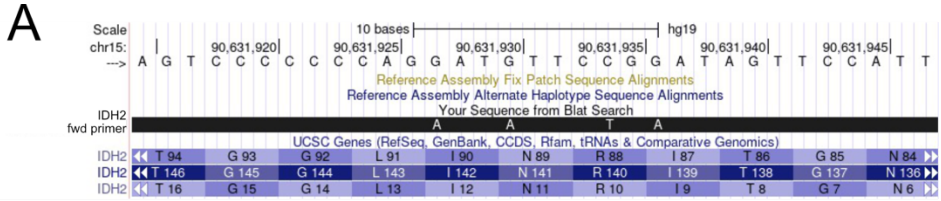
Differential chromatin accessibility assesses by ATAC-seq. (**A**)Example of called peaks in KDM4C (**B**) Example of called peaks in IFIT3 (**C**) Binding affinity heatmap showing affinities for differentially bound sites across all treatment conditions (**D**) Correlation heatmap plot for all treatment conditions.

4.2 Chromatin changes in *IDH2^{mut}* cell lines after AG-221 treatment

Cancer-induced changes in chromatin accessibility are potentially reversible. The *IDH1* mutant inhibitor BAY1436032 partially reverted chromatin signatures. Similarly, the *IDH2* mutant inhibitor AG-221 has the potential to revert deregulated epigenetic patterns. Thus, I investigated the impact of AG-221 on the chromatin state of TF-1 cell lines with or without *IDH2* mutations. Cell lines were collected from HI-STEM laboratories (DKFZ Heidelberg), and subsequently, presence of *IDH2^{mut}* was assessed via Sanger sequencing (Figure 39A). As expected, TF-1 wt did not show mutations in the *IDH2* genes, whereas TF-1 *IDH2^{mut}* harbored three additional point mutations to the anticipated R140 mutation.
Then, sc-DNA was performed to validate Sanger sequencing results and the sensitivity of the myeloid panel to target all *IDH2* mutations observed. Sc-DNA seq detected all four mutations in the *IDH2* gene in erythroblasts from TF-1 *IDH2^{mut}* (Figure 39B). In addition, mutations in *EZH2* and *WT1* were found in both cell lines.

Next, cells from both cell lines were treated with concentrations ranging from 0.06 to 32 μM for six days to assess toxicity. As a control, cells were treated with DMSO in matching volumes. Viability measured via trypan blue staining on a Luna counter showed viabilities around 90% for all cells treated with concentrations up to 10 μM (data not shown). Cell viability measured via luciferase assay obtained similar results, with AG-221 showing no impairment of viability up to 8μM while concentrations of 32μM decreased viability by 80% (Figure 39C). After ensuring that AG-221 does not kill cells at a concentration of 5μM, cells were treated with 5μM of AG-221 or DMSO for six days. Then, scATAC-seq was performed to evaluate reversible and self-sustained epigenetic changes caused by mutant *IDH* on chromatin accessibility.

A



B

Variant	TF-1 wt	TF-1 IDH2mut
EZH2:chr7:148543525:A/G	Homozygous	Homozygous
WT1:chr11:32417945:T/C	Homozygous	Homozygous
IDH2:chr15:90631934:C/T	WT	Homozygous
IDH2:chr15:90631936:G/A	WT	Homozygous
IDH2:chr15:90631927:G/A	WT	Homozygous
IDH2:chr15:90631930:G/A	WT	Homozygous

C

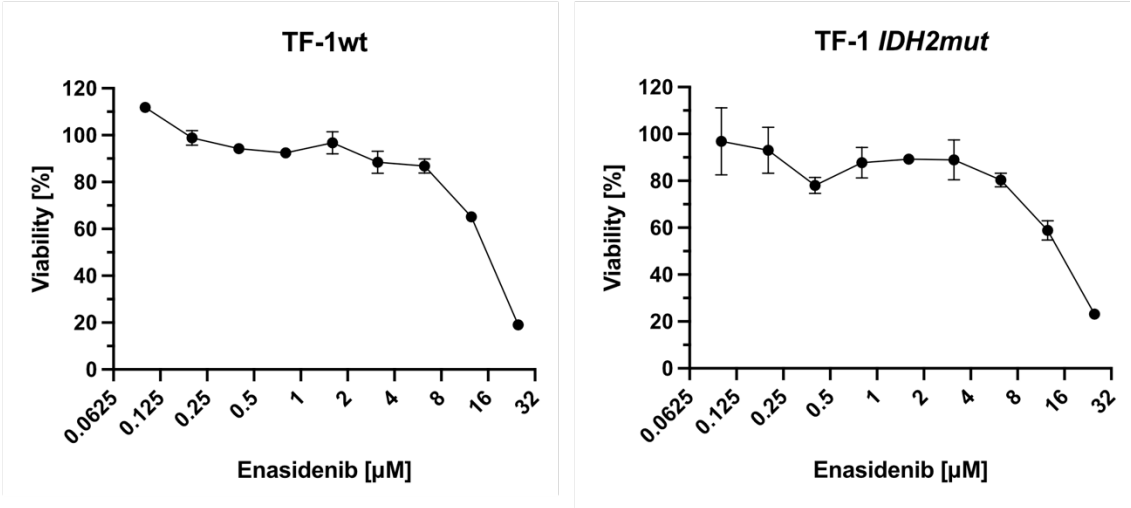


Figure 39

Mutations and viability in TF-1 cell lines. (**A**) Snapshot of point mutations in *IDH2* gene visualized in UCSC Genome Browser (**B**) scDNA-seq of both TF-1 wt and TF-1 $IDH2^{mut}$ (**C**) Viability of cells upon treatment with AG-221 for 6 days, normalized to DMSO control.

For the scATAC-seq run, wild-type and *IDH2*^{mut} cells were treated for 6 days with 5μM AG-221. As a negative control, both cell lines were treated with DMSO in same volumes. Wild-type and mutated cells formed two distinct clusters independent of treatment and once minor cluster with cells from all four conditions (Figure 40A). However, heterogeneity in chromatin accessibility was higher in the population of treated *IDH2*^{mut} cells (Figure 40B). As expected, treatment with AG-221 did not seem to have an impact on chromatin accessibility in wt cells. Cluster 2 comprised cells from wt and *IDH2*^{mut} cells while cluster 4 only contained *IDH2*^{mut} cells (Figure 40C). This led to the hypothesis that cluster 2 might confer non-responding cells to treatment whereas cells from cluster 4 might be sensitive to AG-221. Peak distribution was similar for all clusters and most peaks were found in intronic regions (Figure 40D). 22935 unique peaks were found that were visualized in a heatmap (Figure 40E) and cluster 4 showed distinct changes in accessibility compared to cluster 2. Then pairwise differential testing was performed between cluster 4 and cluster 2 alone (Figure 41A). Out of 207178 features ~1.2% were up and ~0.5% were downregulated. Enriched motifs that were found in these upregulated peaks mainly comprised TFs such as GATA, STAT and RUNX1 (Figure 41B). Especially GATA1 and GATA2 play a crucial role in erythrocyte differentiation and regulation of HSC activity [315]. Since TF-1 cells are derived from a human erythroleukemia sample and AG-221 is known to induce differentiation [316] cluster 4 might indeed represent cells responding to treatment while cells from cluster 2 chromatin state is self-sustained. The TFs downregulated in cluster 4 were JUN, FOS, BACH and CEBPA and RUNX3 which are all involved in hematopoiesis or cell differentiation as well [317-320]. Lastly, per-cell deviations across all motif annotations for all clusters were computed using chromVAR. The most variable TF motifs across the whole dataset were JUNB and FOSL, whereas the most enriched motif in cluster 4 was GATA2 (Figure 41).

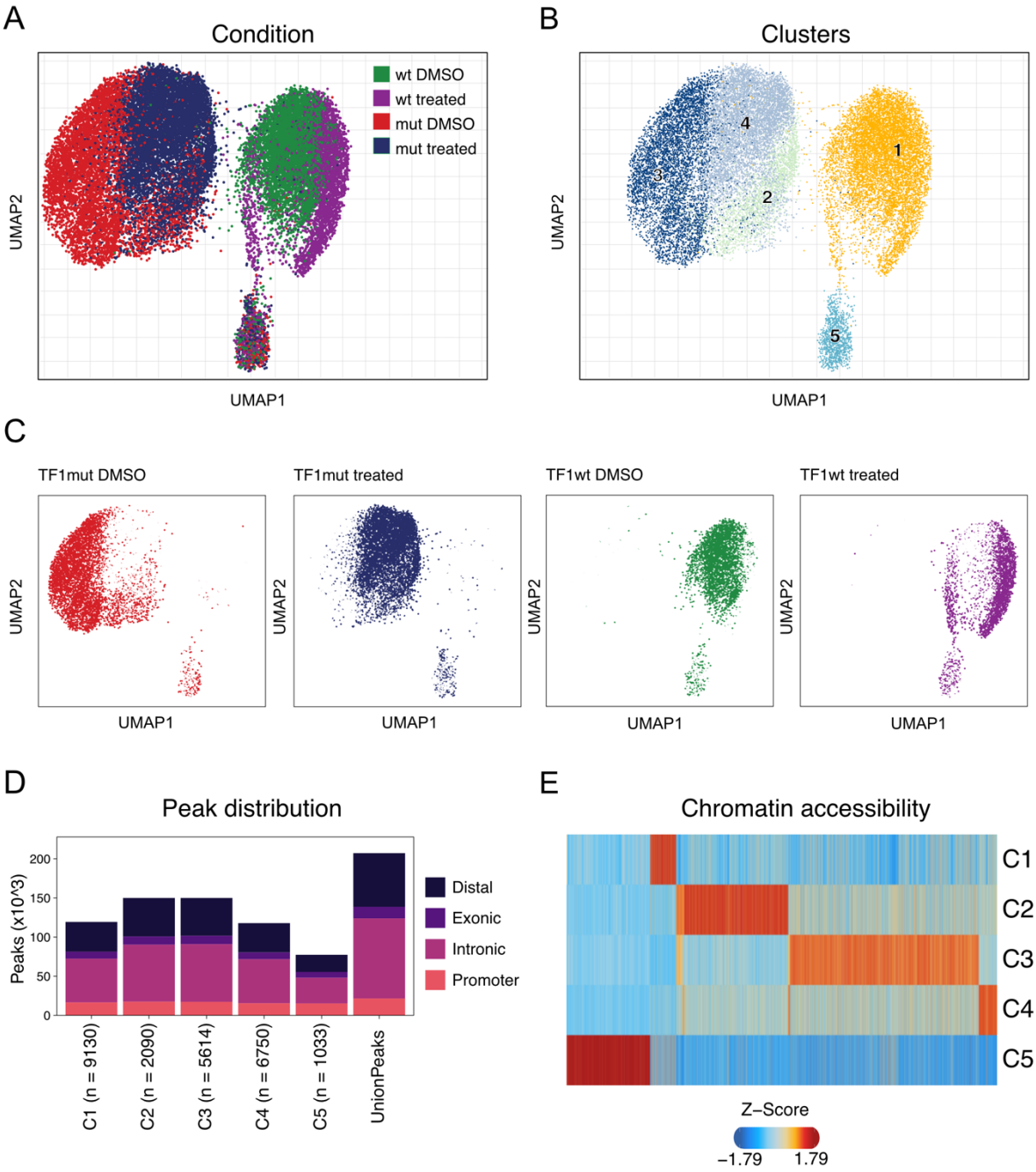


Figure 40

Chromatin changes in TF-1 cells after AG-221 treatment

(**A**) UMAP embedding colored by condition (**B**) UMAP embedding colored by clusters (**C**) UMAP embedding seperated by condition (**D**) Dodge plot of peak distribution (**E**) Heatmap of unique marker peaks

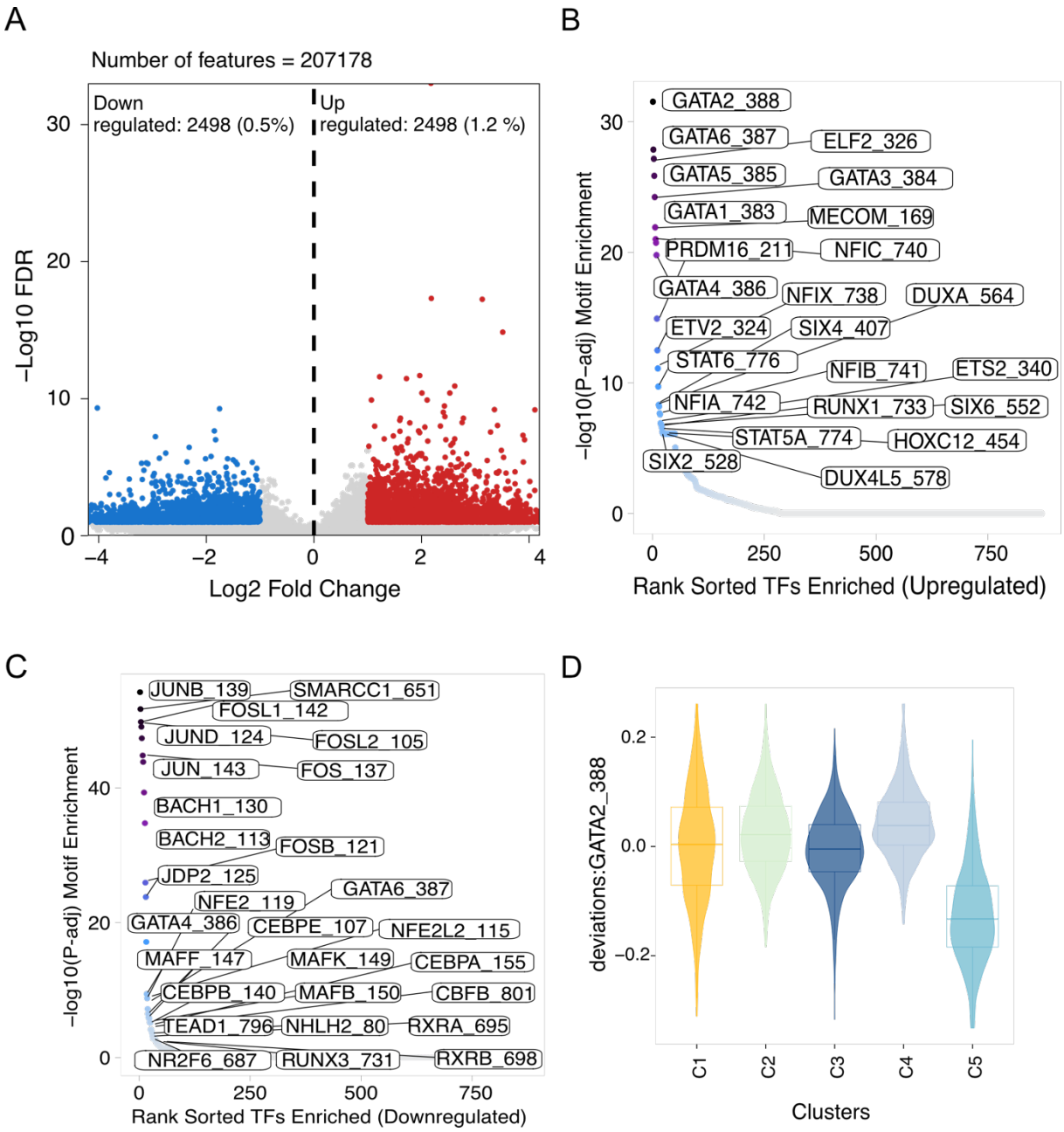


Figure 41

Motif enrichment in TF-1 *IDH2*^{mut} cells treated with AG-221

(**A**) Volcano plot of up and downregulated features in C4 vs C2 (**B**) Motif enrichment in upregulated differential peaks in C4 compared to C2 (**C**) Motif enrichment in dowregulated differential peaks in C4 compared to C2 (**D**) Violin plot depicting the distribution of chromVAR deviation scores of GATA2 for each cluster

4.3 Multi-omics response of AG-120 in one *IDH1*^{mut} AML patient

To further investigate the effects of targeted *IDH*^{mut} treatment, one patient harboring a confirmed *IDH1* mutation was examined on multi-omics levels. I assessed the effect of AG-120 on gene expression, and shifts in clonal evolution and accompanied treatment sensitivities by targeted scDNA-seq. Further, I assessed the impact of this inhibitor on chromatin accessibility patterns. Data was collected during diagnosis, remission, and relapse at the single-cell level for analysis.

4.3.1 Dynamic changes in gene expression programs caused by AG-120 treatment

At first, the effect of AG-120 concerning transcriptomic changes was studied using scRNA-seq. In total, 4870 cells distributed over three timepoints passed stringent quality control. Cells were visualized with a two-dimensional representation using UMAP. Cell types were automatically annotated via singleR [260] (Figure 42). Non-malignant cell types clustered by cell type, whereas AML cells clustered independently (Figure 42). Cell at relapse formed two distinct clusters with cells overlapping from diagnosis. To further investigate changes in transcription, I determined the 15 most essential marker genes of each tumor cluster, and their expression values were plotted onto a heatmap (Figure 43A).

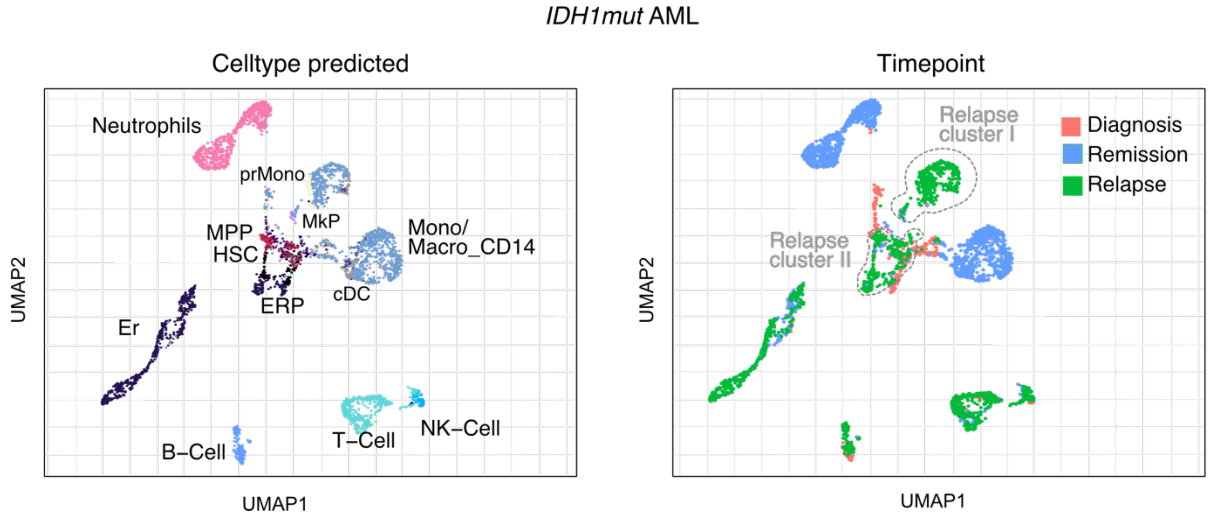


Figure 42
Annotation of *IDH1*^{mut} AML sample. UMAP embeddings are colored by predicted cell type and timepoint.

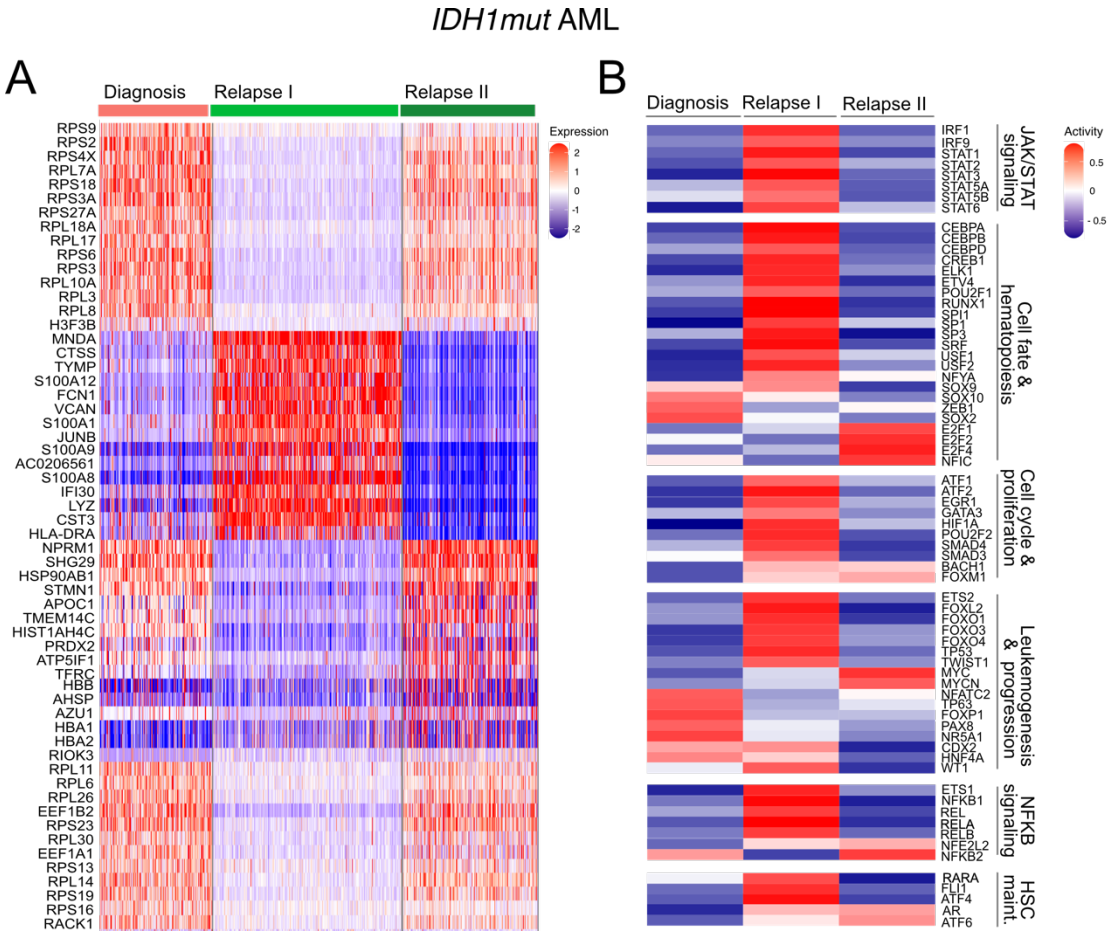


Figure 43

Differential gene expression and transcription factor activity in one AML patient. (**A**) Single-cell heatmap depicting top markers for each tumor cluster. (**B**) Transcription factor activities grouped by tumor cluster.

Differential gene expression looked similar between diagnosis and relapse cluster II, with upregulation of various ribosomal proteins. Aside from protein synthesis, ribosomal proteins are engaged in a variety of biological tasks, including cell cycle arrest and apoptosis, as well as proliferation [321]. In relapse cluster I, especially genes that are known as monocytic markers or play a role in antigen presentation or inflammation, such as LYZ, HLA-DRA [322], S100A8, S100A9, S100A12, or VCAN, were highly expressed (Figure 43A). This result is in line with predicted cell type annotation with relapse cluster II mainly composed of monocyte-like cells, while cluster I resembles more undifferentiated cells such as HSC, MPPs, or erythroblasts (Figure 42). TF analysis revealed high activity in TFs that were documented to be involved in JAK/STAT signaling, cell cycle, proliferation, NFκB signaling, HSC maintenance, and leukemogenesis (Figure 43B). Contrary, these TFs had shallow activity in relapse cluster II, and mainly, TFs of the E2F family showed high activity. These TFs play a vital role during

the G1/S transition in mammalian cells, and deregulated activity or expression of TFs from the E2F family were reported in many human cancers [323].

I then performed GSEA, which revealed significant upregulation of MYC targets at diagnosis compared to both relapse clusters (Figure 44A). As seen in TF activity analysis, JAK/STAT and NFκB signaling was downregulated in GSEA at timepoint of diagnosis (Figure 44B). In addition, a decrease in the innate immune system, inflammatory response and oxidative phosphorylation was detected.

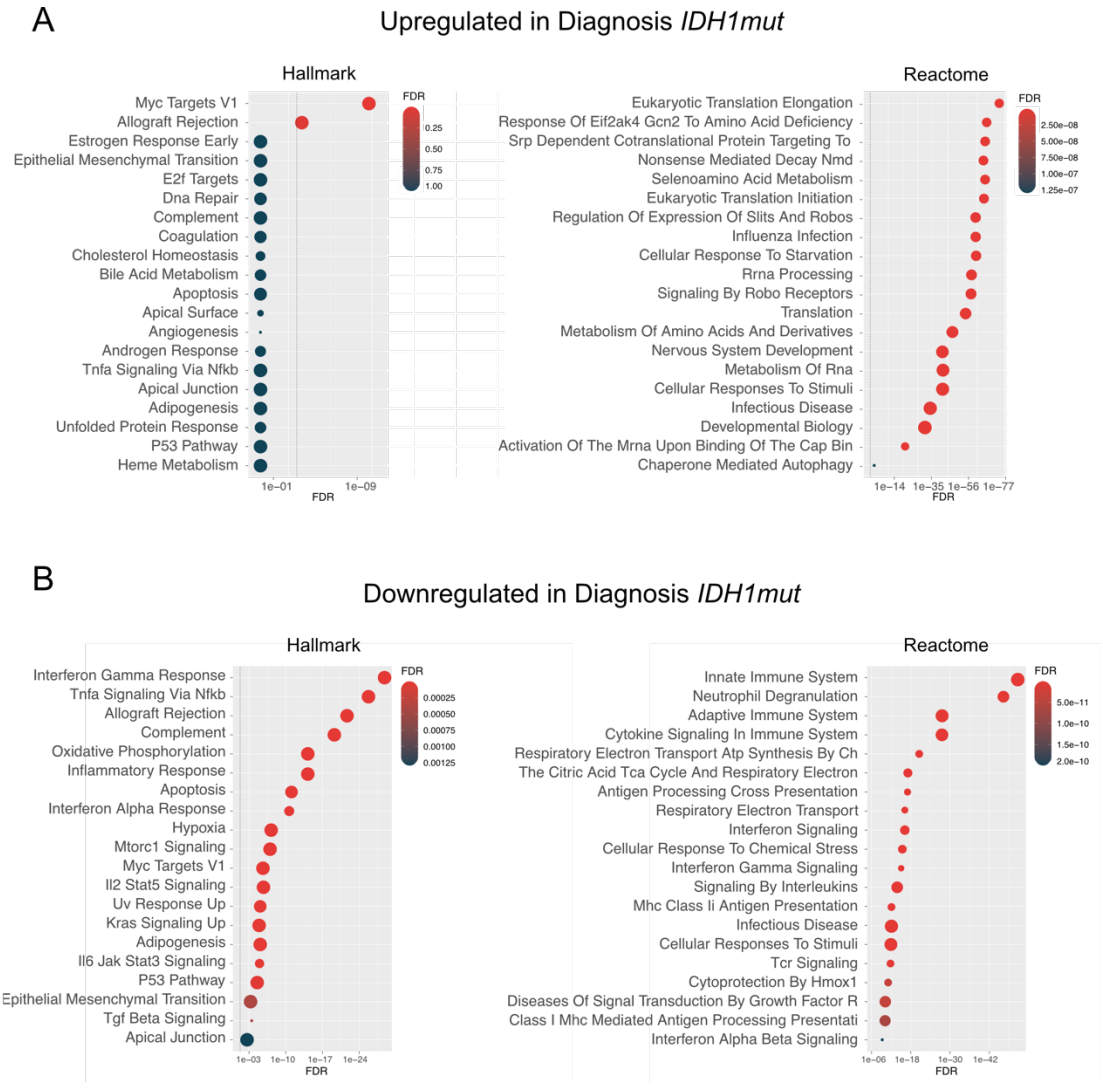


Figure 44

Gene set enrichment analysis of *IDH1*^{mut} AML. (**A**) Dot plot of up regulated gene sets of AML cells from diagnosis compared to both relapse clusters. Gene sets from Hallmark and Reactome were used. (**B**) Dot plot of down regulated gene sets of AML cells from diagnosis compared to both relapse clusters. Gene sets from Hallmark and Reactome were used. Dashed line depicts the significance threshold.

For relapse cluster II, GSEA also detected upregulation of MYC targets and E2F targets when compared to relapse cluster I (Figure 45A), confirming the finding of TF activity analysis. On the other hand, relapse cluster I, showed upregulation in interferon-gamma and inflammation response, TNFα signaling via NFκB, p53 pathway and JAK/STAT signaling (Figure 45B).

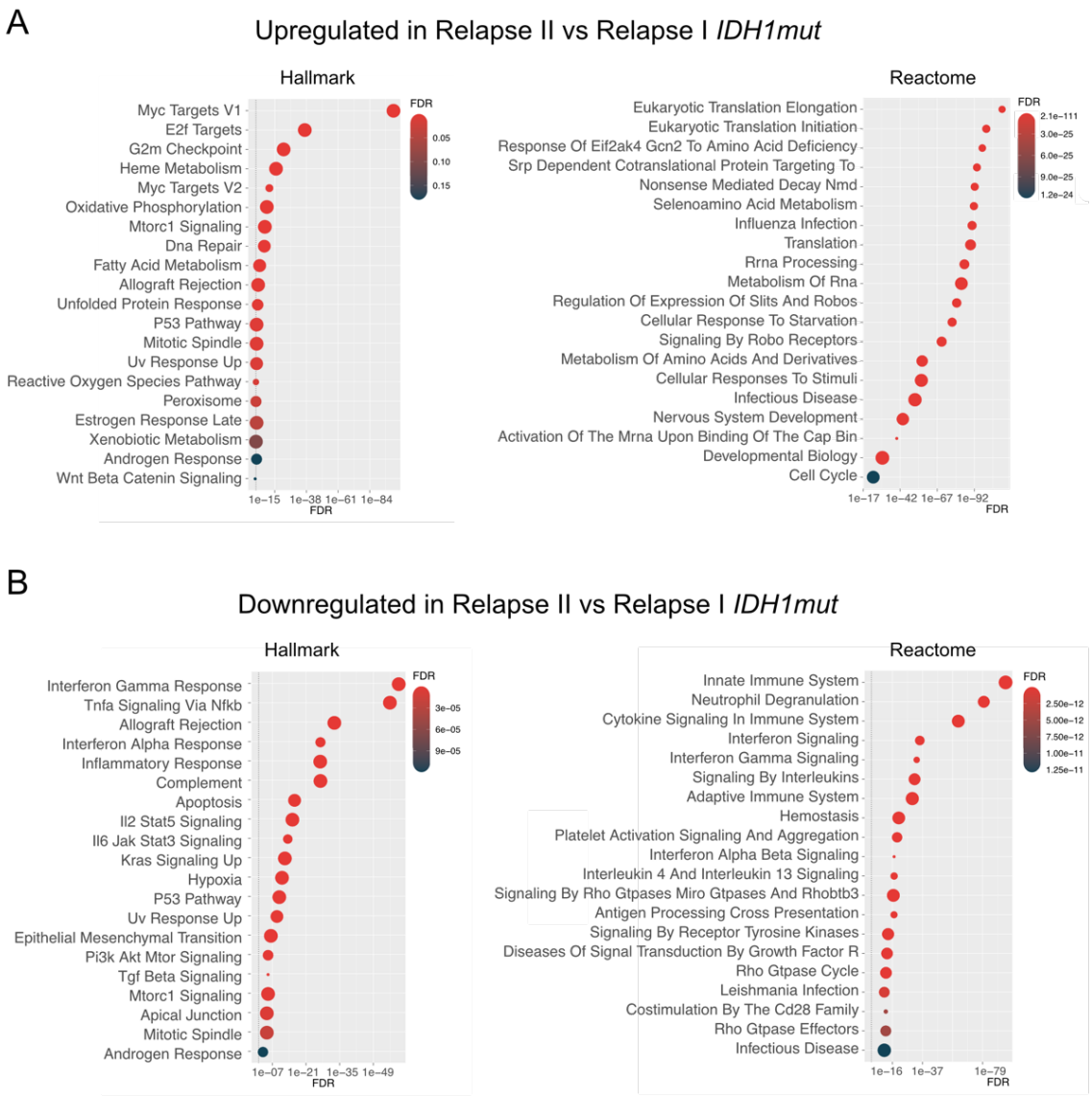


Figure 45

Gene set enrichment analysis of *IDH1*^{mut} AML. (**A**) Dot plot of up regulated gene sets of AML cells from relapse cluster II compared to relapse cluster I. Gene sets from Hallmark and Reactome were used. (**B**) Dot plot of down regulated gene sets from relapse cluster II compared to relapse cluster I. Gene sets from Hallmark and Reactome were used. Dashed line depicts significance threshold.

4.3.2 Emergence of therapy-resistant clones detected with scDNA-seq

AML is a heterogeneous disease and, AML blasts can be composed of several subclones with distinct genetic lesions. Especially during treatment, the frequency and composition of subclones can shift, or new genetic aberrations can arise. To investigate the effects of a drug targeting *IDH1*^{mut} cells and to track clonal evolution during the therapy course, I used scDNA-seq for three longitudinal samples collected from bone marrow.

At diagnosis, three distinct clones were prominent. Most likely, all clones originated from one clone with a heterozygous *CSF3R* (chr1:36933097:T/C) mutation (Figure 46). The most prominent subclone at diagnosis had an additional heterozygous *IDH1* mutation. In contrast, the third clone comprised less than 1% of all subclones and harbored a *JAK2* mutation (chr9:50773770:G/T) in addition to the *CSF3R* mutation. Subclones with *IDH1*^{mut} were shown to be very sensitive to treatment with AG-120, and no *IDH1*^{mut} cells were detected during remission. However, two clones with either a hetero- or homozygous *JAK2* mutation emerged. These clones were resistant to AG-120 and proliferated until they became dominating subclones at the time of relapse.

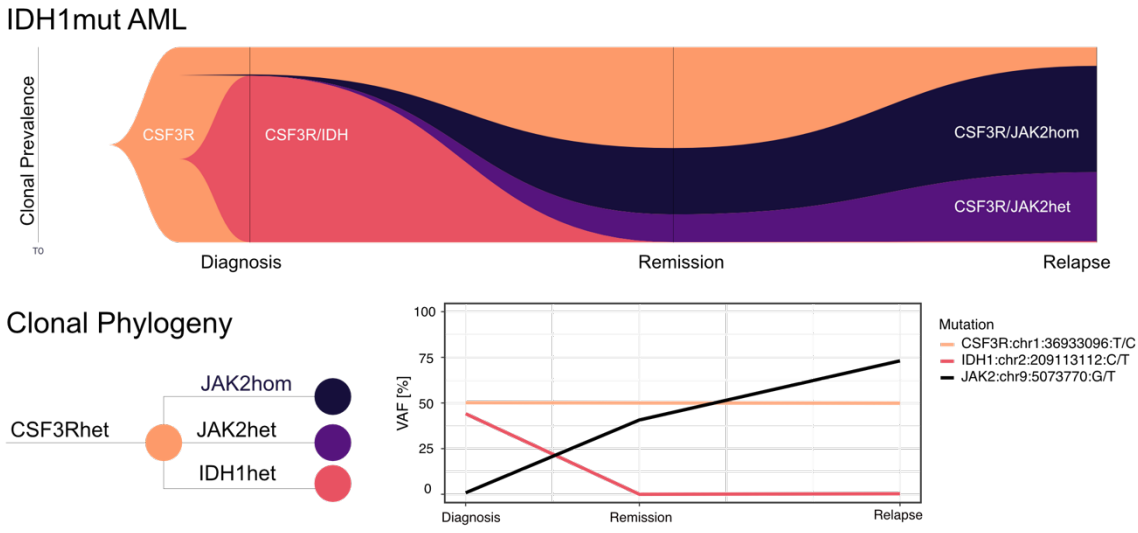


Figure 46

Clonal evolution during AG-120 treatment in *IDH1*^{mut} patient. Top: FISH-plot depicting the clonal composition of AML cells before treatment, during remission, and at relapse. Bottom left: Clonal phylogeny tree of clonal evolution. Bottom right: Aggregated variant allele frequency (VAF) calculated from single-cell data visualized as line plot.

4.3.3 Changes in chromatin state caused by *IDH1^{mut}*

To investigate the changes of chromatin accessibility caused by *IDH1^{mut}* healthy cells from remission were compared to AML blasts carrying an *IDH1^{mut}* (Figure 47A). Cluster identities were defined by integration with scRNA-seq data obtained from the same patient (Figure 47B). Cells from remission were mainly assigned as monocytes/macrophages $^{CD14+}$ and T-Cells. Cells collected at diagnosis mainly comprised erythrocyte-like AML blasts and T-Cells as well. Peaks were called with MACS2 and 628 unique peaks were plotted in a marker peak heatmap (Figure 47C). CD14 positive monocytes/macrophages showed high chromatin accessibility compared to erythrocyte-like blasts and T-Cells. Then pairwise testing of erythrocyte-like blasts to monocytes/macrophages $^{CD14+}$ was conducted. In total, ~0.1% of 21936 marker features were upregulated and ~1.5% downregulated in erythrocyte-like blasts (Figure 46D). Mutations in *IDH* have been shown to promote hypermethylation [324], inducing chromatin condensation [325]. Thus, decreased chromatin accessibility in *IDH^{mut}* blasts would be expected. Then motifs enriched in peaks found in *IDH1^{mut}* erythrocyte-like cells were investigated. Many of the upregulated motifs were part of the SOX and forkhead box (FOX) transcription factor families and the most enriched motifs was SMAD1 (Figure 47E). SMAD1 is a key mediator in TGF-β signaling and is linked to a variety of malignancies. This TF is associated in a wide range of biological processes, such as apoptosis, cell proliferation, development, and immune response [326]. SOX genes are involved in oncogenesis [327], embryonic development, and epigenetic reprogramming of stem cells [328]. In addition, they can play distinct pathogenetic roles in AML [329]. FOX TFs are also linked to differentiation, senescence, and proliferation. Further, leukemogenesis, relapse and treatment sensitivities in AML are often influenced by FOX proteins[298]. Downregulated motifs were found for TFs such as JUN, FOS, BACH and CEBPA (Figure 47F). JUN and FOS are proto-oncogenes that have been linked to promoting myeloid differentiation [319]. *CEBPA* is essential for the differentiation of myeloid progenitors and mutations in this genes are linked to acute myeloid leukemia [318]. *BACH1* is a master regulator of oxidative stress [320] and its upregulation was suggested as potential antileukemic therapy strategy[330].

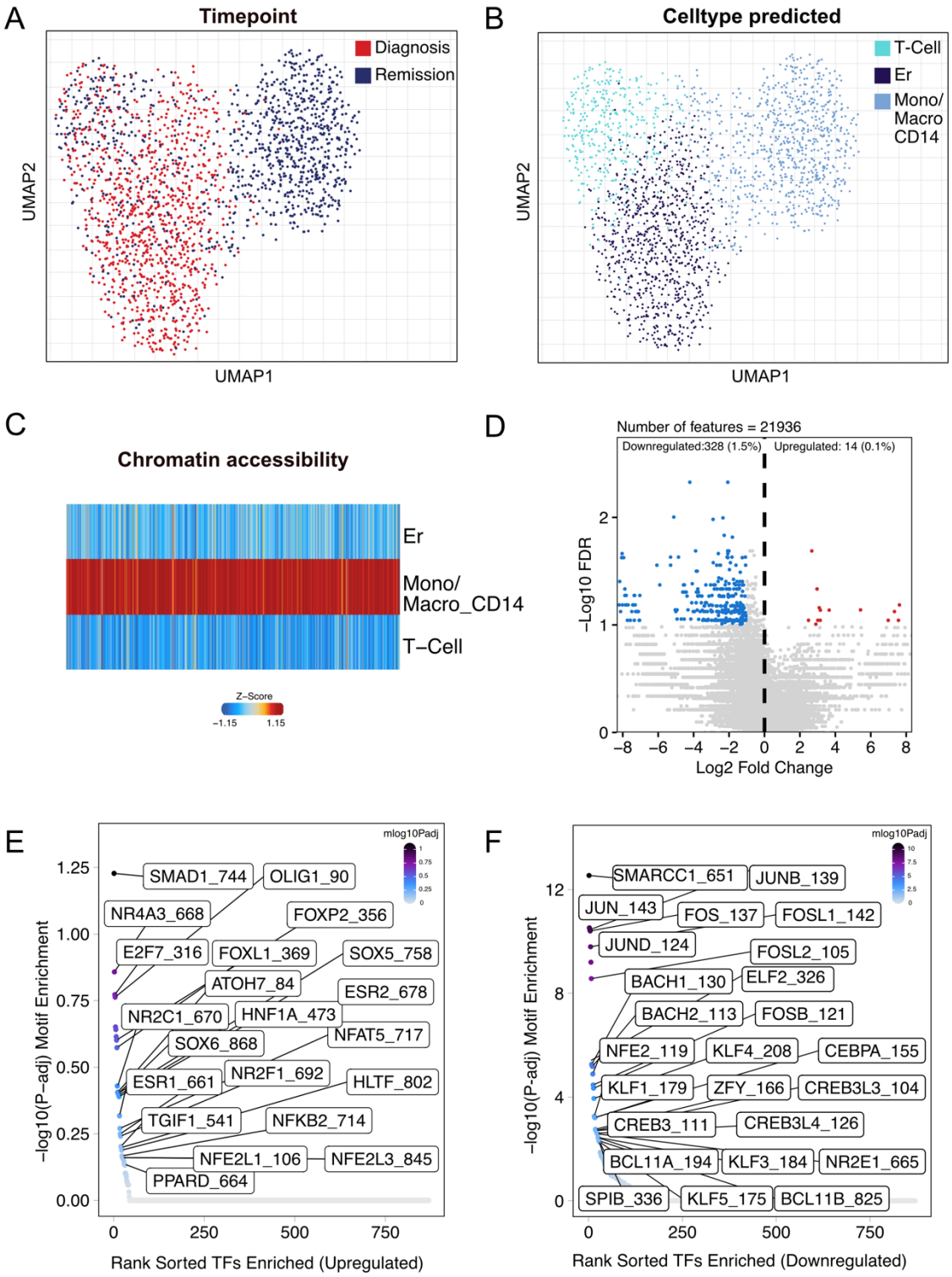


Figure 47

Chromatin accessibility in *IDH1*^{mut}

(**A**) UMAP embedding colored by timepoint (**B**) UMAP embedding colored cell type inferred from scRNA-seq (**C**) Peak matrix heatmap (**D**) Volcano plot of up and down regulated features in erythrocyte-like AML blasts compared to monocytes/macrophages (**E**) Motif enrichment in upregulated differential peaks in erythrocyte-like AML blasts compared to monocytes/macrophages (**F**) Motif enrichment in downregulated differential peaks in erythrocyte-like AML blasts compared to monocytes/macrophages

5 Stratification of leukemic blasts across three AML subgroups

Leukemic cells have a hierarchical structure that mirrors the differentiation hierarchy of non-malignant hematopoiesis [127-129]. Hence, the phenotype of a leukemic blast can resemble naïve myeloid progenitors, differentiated myeloid blood cells, or express features from distinct differentiation stages. Besides genetic factors, this may partly account for heterogeneity observed in AML. Insights on the composition of AML blasts might help explain differences in therapy responses, develop targeted therapies, and predict treatment outcomes [139].

Thus, I wanted to stratify patient-specific AML blasts along the HSC to myeloid differentiation axis across the three AML subgroups studied in the thesis (*MLL* fusions, *FLT3,* and *IDH* mutations) and compare the influence of treatment on cell type abundance. Therefore, module score and cell type prediction analysis were performed using scRNA-seq data. The investigation revealed extensive malignant cell diversity and provided detailed information on AML cell types and differentiation states (Figure 48, Figure 49). The relative abundance of distinct malignant cell types found in tumors varied; in some, just one or two identities predominated, while others were composed of a wide range of malignant cell types. *MLL* fusions conferred a highly differentiated phenotype, with monocyte/macrophage CD14-like and promonocyte-like cells predominating. *MLL-EDC4* cells were an outlier in this categorization, with elevated scores for HSC and MPP signatures (Figure 48). Automated cell type annotation also predicted a relative abundance of more than 90% in progenitor-like cells, such as HSC-, MPP-, and ERP-like cells (Figure 49). However, compared to cell composition samples harboring *FLT3*-ITDs collected at diagnosis, module scores for HSC and progenitor cells were lower in *MLL-EDC4* AML. Generally, AML samples with *FLT3*-ITDs or *IDH1* mutations were composed of a more diverse makeup of cells along the myeloid axis in comparison to *MLL* fusions. The diversity of cell types in relapse was remarkably high and displayed elevated differentiation signatures in tumoral subclusters while the corresponding other intra-tumoral subclusters maintained a high abundance of progenitor cells.

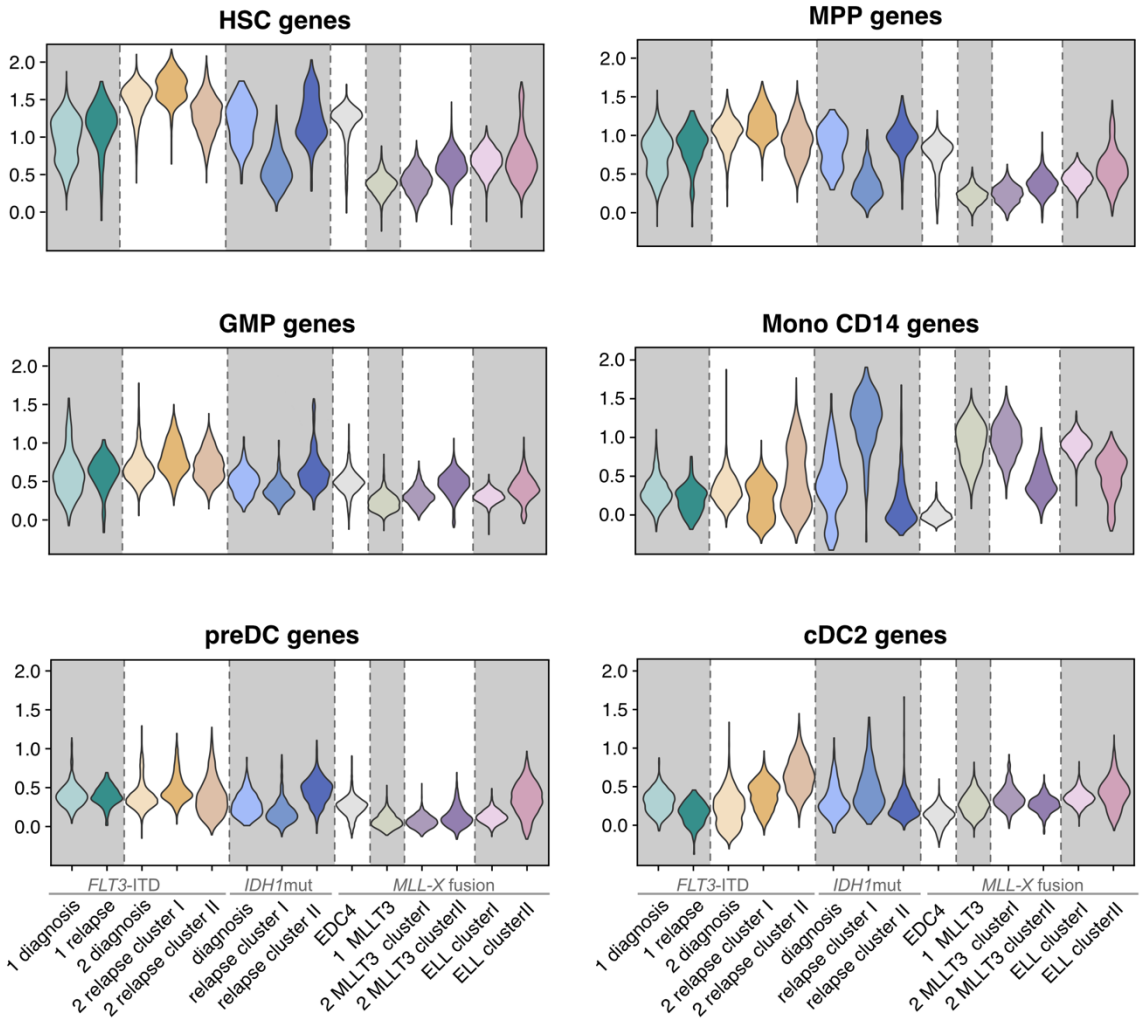


Figure 48

Gene signature scores of malignant cells across AML subgroups. Violin plots depicting module scores for HSC-, GMP, MPP, Mono CD14-, preDC- and cDC2 genes.

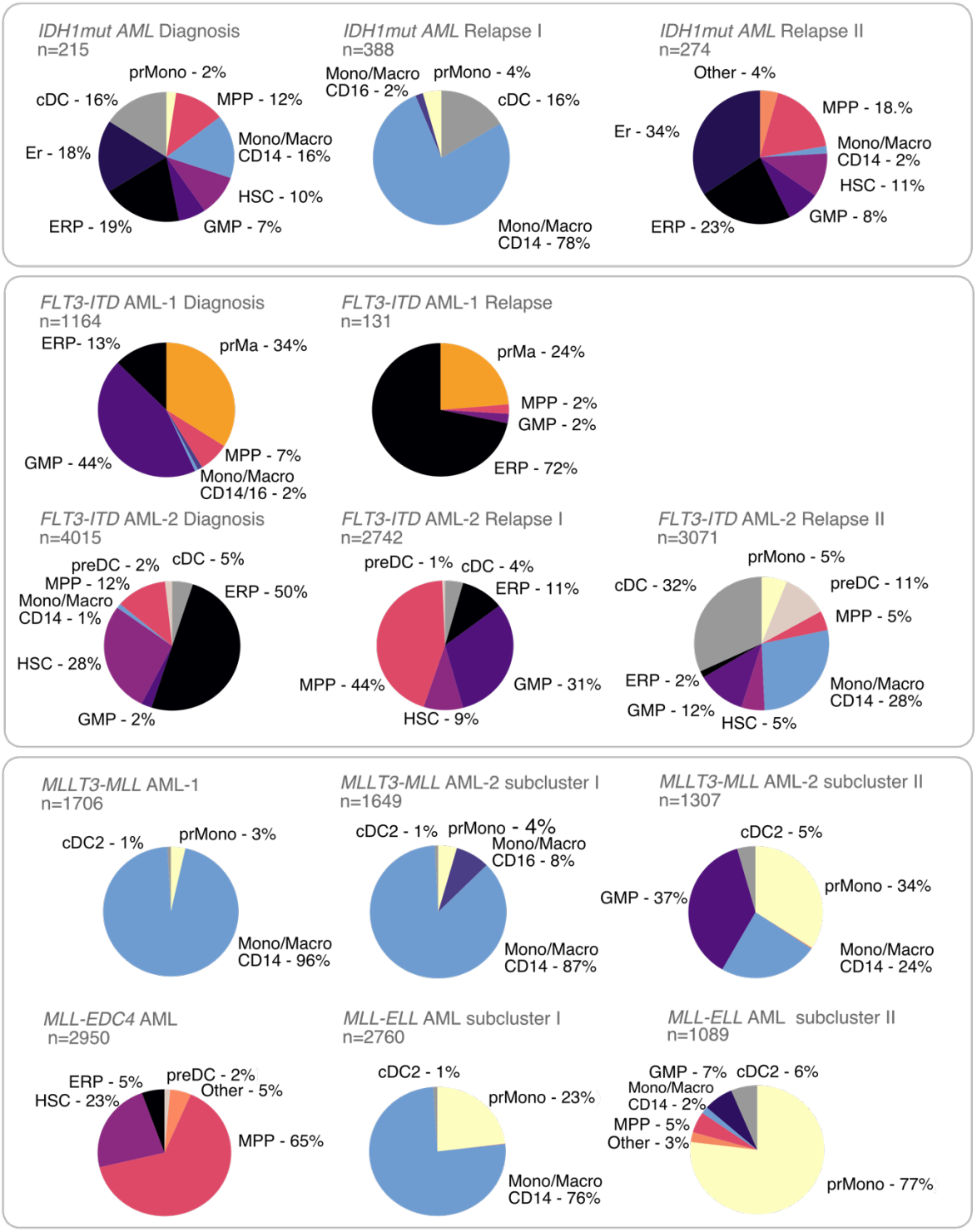


Figure 49

Cell type abundance of malignant cells across AML subgroups. Pie charts depicting the relative abundance of AML blast classification based on transcriptomic similarity to healthy myeloid cells. HSC, hematopoietic stem cell; MPP, multipotent progenitor; GMP, granulocyte monocyte progenitor; prMono, Monocyte; Mono/Macro CD14, Monocyte/Macrophage CD14^{+}; Mono/Macro CD16, Monocyte/Macrophage CD16^{+}; prMa, mast cell precursor; ERP, erythroid progenitor; Er, Erythrocyte; preDC, pre dendritic cell; cDC2, conventional dendritic cell type 2.

Discussion

A hallmark of cancer are genetic alterations that lead to the deregulation of transcriptional programs that involve oncogenes and tumor suppressors. Thus, it is crucial to identify tumor specific dysregulated transcriptional signatures to dissect the patho-mechanisms and to identify new treatment options [331]. From genomic sequencing of cancer cells alone these transcriptional dependencies can only partly be understood [332]. Furthermore, tumor cells frequently show deregulated gene expression that leads to abnormal developmental patterns and may be caused by differentiation block or epigenetic reprogramming [333,334]. Thus, the phenotypic effects of a genetic aberration may depend on the cells' underlying epigenetic landscape and vary depending on the stage of differentiation at which it occurs [191]. Single-cell approaches provide tremendous possibilities to address these issues and reveal molecular mechanisms responsible for tumor heterogeneity and drug sensitivities. Multiple single-cell sequencing readouts offer complementary information to explore tumor composition, mutational co-occurrences, and clonal evolution during treatment in thousands of cells in a particular sample. Especially the emergence of targeted epigenetic drugs and the potential reversibility of deregulated epigenetic events highlights the significance of studying these changes in AML [72,117]. This thesis explored how genetic mutations, transcription, and epigenetic programs are altered in acute myeloid leukemia (Figure 50).

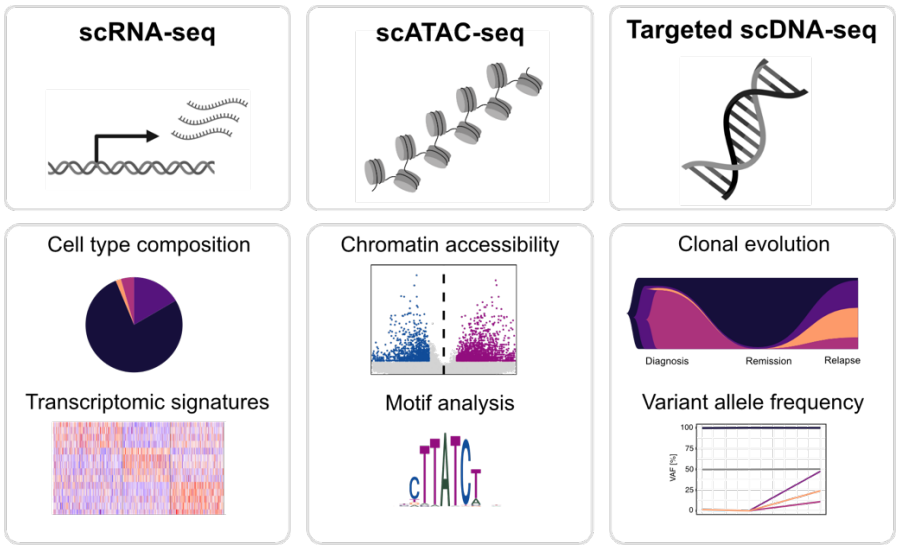


Figure 50
Different single cell sequencing readouts used to dissect heterogeneity and molecular profiles of AML subtypes.

1 Workflow for experimental approaches and data analysis methods for multi-omics analysis of primary AML cells

My thesis work illustrates how the quality of sequencing libraries is crucial for successful downstream analysis. Factors that could interfere with sample quality were identified and strategies to rescue libraries have been developed. The cell integrity and RNA quality are essential prerequisites, for single-cell transcriptome sequencing studies [335]. Sampling time has a strong effect on single-cell gene expression profiles. Significant changes were observed after 2 hours which increased in a time-dependent manner [336]. Additionally, sampling duration surpassed batch and donor for various cell types and came in third behind cell type and patient variability as the biggest drivers of variance. Thus, comparable sampling times within a dataset are necessary for reliable results. The cryoprotectant dimethyl-sulfoxide (DMSO) was shown to conserve intact and viable cells without altering transcriptional profiles. Thus, this cryopreservation method should be implemented into common single-cell workflows [335].

Furthermore, clinical samples are frequently fundamentally limited, and benchmarking or technological development can only be done on cell lines beforehand. Since every patient and sample is unique, it is important to find a balance between a standard protocol and reasonable expectations for successful experiments. First, cell viability and the extent of dissociation into a single-cell suspension should be evaluated [337]. Only samples that pass stringent quality control (QC) should move on to single cell multi-omics sequencing procedures. FACS might potentially be utilized as an extra QC before profiling.

Further, I explored distinct tools to investigate tumor populations in AML. AML blasts share antigen expression patterns found in healthy immature myeloid cells, such as the common differentiation markers CD13, CD33, and CD34 [338]. Other cell markers, including monocytic, megakaryocyte, or erythroid markers, are expressed based on the genetic AML subtype and differentiation state. This makes the disease very heterogeneous, and there is no defined marker to reliably determine AML blasts for all subtypes in scRNA-seq data sets. My study shows that *FLT3* expression is sufficient to distinguish leukemic cells from non-malignant microenvironment for patients carrying *FLT3*-ITDs. However, this method is not applicable to other genetic AML subtypes.

Previous research on intra-tumor heterogeneity has shown that single-cell copy number alterations can be inferred from scRNA-seq [339-341]. Thus, the tumor subclone composition of a given sample can be resolved from a copy number alteration (CNA) analysis. However, these tools are unsuitable for some hematopoietic diseases, especially AML, that have a low frequency of CNAs [342]. Additionally, previously used methods, such as inferCNV [342] or Honeybadger [341], were created to interpret datasets from first-generation scRNA-seq techniques with deeper sequencing and fewer cell throughput. scRNA-seq data generated with nano well or microdroplet platforms might not be suitable for these tools [243]. In this thesis, I optimized the parameters of tools designed for other sample types. My findings confirm that CopyKat [243] is applicable to distinguish normal from AML cells by inferring ploidy. Additionally, I empirically tested different tools to integrate leukemic cells with a healthy bone marrow reference and to infer pseudotime. I demonstrate that integration with custom-specified anchors works best and that the tested velocities might not be appropriate for acute myeloid leukemia samples. Establishing a framework to produce high-quality sequencing libraries and analyze the generated data sets enables a fast and reproducible analysis for further samples to come.

2 Cell type analysis of *MLL-r*

I compared the peripheral blood transcriptomic profile of a novel *MLL-EDC4* fusion in AML to *MLLT3-MLL* and *MLL-ELL* fusions representing two of the most common *MLL* fusion partners in AML. I hypothesized that investigating these transcriptomic patterns would elucidate profound differences that might be disease relevant like aberrant signalling pathways that can be used to stratify patient risk [343,344].

Subsequent analysis of these four patients revealed an overlap of upregulated gene subsets in *MLLT3-MLL* and *MLL-ELL* AML, whereas *MLL-EDC4* AML showed a distinct transcriptomic profile.

Stratification of differentiation states of all patients led to the identification that *MLL-EDC4* leukemic cells are more stem cell and progenitor-like than the other AML fusions. These cells were predicted to comprise more monocyte-like cells. Leukemic cells carrying an *MLL-EDC4* fusion might cause a more stem-like cell stage either by blocking myeloid differentiation or leading to dedifferentiation of a more developed cell of origin. Many genes upregulated in

MLL-EDC4 compared to the other *MLL*-fusions are crucial in hematopoiesis or leukemic stem cell activation. Additionally, increased expressions of the genes *HOXA9, RUNX1, MYB, and GATA2* were documented in endothelial-to-hematopoietic transition, leukemic stem cell activation, or cell-fate decision [283,284,286-288].

The analysis of the microenvironment indicated the presence of immunosuppressive monocytes in *MLL-EDC4*. The expression of CD36, cathepsins, and CLEC receptors have been described to infer immune or T-cell response [345,346]. Seto *et al*. showed that *EDC4* is vital in the posttranscriptional regulation of the pro-inflammatory cytokine IL-6 leukemic macrophage cell lines [280]. Another study linked IκB kinase-*EDC4* interaction with P-body formation and regulation of mRNAs encoding inflammatory cytokines [347]. An immunosuppressive milieu typically consists of cellular and soluble elements, supporting cancer immune escape and tumor development [348].

The aberration of *EDC4* might also be the driver of upregulation of various ribosomal proteins linked to initiation, elongation, or termination of protein translation. RPs have also been involved in embryonic development and malignant transformation of cells [349]. Various studies confirm that genes that are clinically important for *MLL*-mediated acute leukemias intervene in processes of transcription resulting in either extended or incorrect chromatin signatures at transcribed gene loci [125,350,351].

Differential gene set enrichment analysis in *MLL-EDC4* AML revealed upregulation of pathways associated with eucaryotic translation or elongation, RNA metabolism, MYC targets, or redox-related bioenergetic pathways. Upregulation of redox metabolism pathways has been demonstrated to infer hematopoiesis due to an increase in oxidative stress causing genomic instability [352,353]. Additionally, cancer cells exhibit high metabolic plasticity to cope with variable resource availability. While hypoxic tumor cells are restricted to glucose-dependent anaerobic glycolysis, oxidative cancer cells can use glutamine, lactate, and lipids in addition to glucose to fuel oxidative phosphorylation [354,355]. This biosynthetic and bioenergetic balance is fine-tuned at an enzymatic level to meet cell demands [356,357]. LHBD, a key glycolytic enzyme catalyzing the interconversion of lactate to pyruvate, was most significantly expressed in *MLL-EDC4* AML [358]. The overexpression of *LDHB* warrants metabolic plasticity and facilitates the adaption of cancer cells to specific cellular conditions [359]. So far, not many studies have investigated the role of *LDHB* in AML [291]. However, *LDHB* was shown to regulate autophagy in oxidative and glycolytic cancer cells [359]. Autophagy

promotes the proliferation and survival of cancer cells. This is achieved by recycling organelles and damaged proteins in the event of oxidative stress and providing metabolite replacement in the case of nutrient deprivation. A study showed that cancer proliferation is inhibited after selective inhibition of *LDHB* [359]. Thus, the inhibition of *LDHB* could be a promising target for treating *MLL-EDC4* AML [290,291].

The upregulation of the pathway "EIF2AK4 response to amino acid deficiency" indicates mechanisms of cancer glycolysis. Glucose starvation is known to have implications in cancer by interfering with the initiation phase of translation. Concomitantly, translationally repressed mRNAs can relocate to cytoplasmic foci such as P-bodies, stress granules, or eIF2B bodies, where they can be stabilized, decapped, or degraded [278].

The proto-oncogene *MYC* is crucial to mediate cell growth, proliferation, and tumorigenesis. It is also one of the four "Yamanaka genes" that jointly reprogrammed fibroblasts to a pluripotent stem cell state [295,360]. In addition, TF activity showed an upregulation of *POU2F1* in *MLL-EDC4* AML, which can function in cell growth control, cellular stress response, stem cell identity, and immune regulation and belongs to the same POU-family as *Oct-3,* which is also part of the Yamanaka factors [296]. A signature of high transcription factor activity in *MLL-EDC4* was detected for TFs such as E2F1, E2F4, ETS1, MYB, MYC, GATA1, and TAL1 that could potentially be linked to a more stem-like phenotype [292-294].

The emphasis on an upregulation of *PARP1* and other genes part of the MRN complex, could not be validated [276]. However, first approaches to characterize this unique patient with an *MLL-EDC4* fusion using scRNA-seq were limited to a defined set of candidate genes. Furthermore, aberrant gene expression was not set into context with other AML samples, which are required to draw conclusions regarding distinct patho-mechanisms in AML.

Since these data are based on a small number of patients carrying *MLL* fusions, further studies will be required to determine to which extent these findings can be applied in general to this newly detected fusion partner.

3 Multi-omics analysis of AML patients harboring *FLT3*-ITDs

Internal tandem duplications of the *FLT3* gene occur in around 30% of AML patients, and they generally confer a poor prognosis and enhanced risk of relapse [361,362]. The multi-kinase inhibitor midostaurin can be used therapeutically to address the ensuing constitutive *FLT3* activation, which promotes carcinogenesis. AML with *FLT3* mutations is very heterogeneous, similar to other subtypes of AML. The disease evolves by clonal selection, and the number of concurrent mutations influences the course of treatment. This poses difficulties for successful therapeutic intervention given the ongoing selection pressure towards medication resistance [363,364]. Accordingly, there is an urgent need to resolve clonal evolution patterns when using novel anti-leukemic drugs, such as the tyrosine kinase inhibitor midostaurin.

A study showed that almost 50% of *FLT3*-ITD patients treated with midostaurin acquired mutations in signaling pathways that allowed tumor clones to escape from treatment [365]. Bulk sequencing is not suitable for detecting rare subclones that may gain proliferative advantage through targeted therapy, and it is possible that these clones already existed before treatment. In contrast, the presence of different subclones and the co-occurrence of mutations on a single-cell level in a given patient sample may inform about drug sensitivities of certain subclones and direct treatment choices [366,367]. Thus, I used a targeted single-cell DNA-sequencing approach to analyze the clonal evolution of two AML patients carrying ITDs in the *FLT3* gene. Both patients were treated with midostaurin after diagnosis. Although the composition of subclones did not alter extensively after treatment for patient *FLT3*-ITD AML-1, scDNA-seq could resolve mutational co-occurrence of a mutation in *TET2* in four *FLT3*-clones, which would not be achievable with bulk sequencing. However, the possibility that these *FLT3* clones acquired additional *FLT3*-TKD mutations, not covered by the used myeloid panel, cannot be excluded. AML patients with *FLT3*-ITDs treated with selective *FLT3* inhibitors were shown to develop drug resistances caused by the emergence of *FLT3*-TKD^{+} subclones at relapse [368]. Amongst other genes, such as *BRAS*; *NRAS* or *IDH1/IDH2*, the rise of *TET2* clones in relapse samples was linked to *FLT3* inhibitor-induced selection pressure [363,369]. Treatment with midostaurin eliminated *FLT3* clones at remission for patient *FLT3*-ITD AML-2, and scDNA-seq could detect a subclone population with *BCOR*/*KDM6A*/*DNMT3A* mutations. These subclones were only present at a frequency of less than 1% at diagnosis and remission, highlighting the great sensitivity of targeted scDNA-seq. Since midostaurin is

primarily designed to eradicate *FLT3* clones, this clone might have had a clonal advantage leading to accumulation at the time of relapse [370]. The reemergence of the *FLT3* clone is probably caused by acquired resistance through further mutations in the *FLT3* genes that could not be detected via scDNA-seq. Comparing the scDNA-seq results with bulk sequencing data showed the need to optimize the used scDNA panel further to target all *FLT3*-ITDs in this patient reliably.

The presence of two distinct subclones at relapse might also explain the observed transcriptomic differences in the relapse sample obtained by scRNA-seq. scRNA-seq resolves (non-) responding cell subgroups based on their molecular transcriptomic signature, giving valuable insights into AML cell states and tumoral heterogeneity. High *FLT3* expression levels were observed in all leukemic cells, especially in relapse cluster II in patient *FLT3*-ITD AML-2. Differential gene and gene set expression analysis showed upregulation of *STAT* and *CEBP* genes, implicating enhanced *FLT3* downstream activity. For this patient, the scRNA-seq analysis revealed two different routes of *FLT3* signaling within the same tumor sample, underlining the power of single-cell sequencing. This result is in accordance with recent findings that *FLT3*$^{+}$ AML cells can either develop resistance by acquiring new mutations or bypass *FLT3* inhibition through certain molecular processes [370]. The underlying mechanisms comprise of changes in the acidity of the intracellular environment [371], elevated levels of *FLT3* ligand [372], and an increase of certain kinases, such as PIM [373]or AXL receptor tyrosine kinase [374]. Additionally, the expression of MAPK/ERK and PI3K/AKT/mTOR pathways persisted or increased in resistant *FLT3*-ITD leukemic cells, indicating *FLT3* independent signaling that drives resistance [375,376]. Interestingly, the scRNA-seq analysis did not detect broad differences in FLT3 signaling for patient *FLT3*-ITD AML-1 but upregulation of genes involved in NFκB signaling or leukemogenesis and tumor progression. Thus, In this case, plasticity in transcription might be the driver for midostaurin resistance rather than a genetic driver [370].

For patient *FLT3*-ITD AML-2, the scATAC-seq analysis yielded comparable results to the scRNA-seq. The data showed that chromatin signature, initially present at diagnosis, is maintained in relapse cluster I, and relapse cluster II has a distinct chromatin accessibility pattern. Integration of these two read-outs confirmed that cluster II in scATAC-seq is indeed the same as in scRNA-seq. A transcription factor motif analysis validated that relapse II is enriched for binding motifs linked to FLT3 downstream targets, such as CEPB, IRF, SPI1

(PU.1), and STAT [377,378]. This further strengthens the assumption, that relapse cluster II circumvents *FLT3* inhibition by upregulation of other molecular pathways while cluster I might confer resistance through acquisition of additional mutations.

In conclusion, this multi-omics approach combined the transcriptome analysis, mapping epigenetic patterns of open chromatin, and dissecting the tumor subclone structure and its evolution. By doing this, I obtained a comprehensive data set that characterized leukemic subgroups, their response to therapy, and molecular deregulation patterns that might drive resistance during midostaurin treatment.

4 Multi-omics analysis of *IDH^{mut}* in AML

Enzymatic control of epigenetic processes holds the promise of novel and intriguing therapeutic targets for cancer treatment. Drugs may directly or indirectly impact a tumor cell's dysregulated epigenetic state [379]. Here, I demonstrated that targeted therapy for AML patients/cell lines might mitigate some of the accessibility alterations brought on by these mutations. Nevertheless, only a fraction of regions displayed significant changes in reversibility, while other sustained irreversible. Therefore, the question emerges whether long exposure to the trigger (such as inhibiting histone or DNA demethylases) leads to self-sustaining regions that cannot be reverted even after trigger removal. Although epigenetic alterations in cancer are in principle reversible by drugs [379], a complex interplay of numerous factors might interfere with the reversion of aberrant activities [380,381]. Further research, such as identifying specific deregulated events and targeting these with more specific drugs, would be a possibility to answer this question. Despite encouraging preliminary results, epigenetic drugs did not meet the high expectations yet, which is ascribed to their poorly studied pleiotropic and global effects [382,383]. A deeper understanding of the molecular mechanisms behind medication activity would be necessary to overcome this issue. I could demonstrate that the epigenetically acting drug AG-221 induced alterations in chromatin accessibility and enrichment of TF motifs indicating myeloid cell differentiation when treating *IDH2^{mut}* cell lines. Further, I investigated changes in chromatin state caused by *IDH1^{mut}* in one AML patient. The most upregulated TF motif found at diagnosis was *SMAD1*, a key regulator of TGF-β signaling [384]. One study demonstrated that downregulation of *IDH1*, and subsequent increase in α-KG, inhibits TGF-β receptor degradation via the TGFBR-IDH1-

Cav1 axis. This, in turn, promoted TGF-β signaling in cancer-associated fibroblasts [385]. However, the influence of AG-120 treatment on *IDH1*^{mut} cells could not be assessed since scDNA-seq revealed successful elimination of *IDH1*^{mut} cells and emergence of new clones without this aberration. This highlights that drug therapy affects the transcriptomic and epigenetic landscape and the clonal genetic composition. Thus, clonal detection is relevant for clinical assessment to adapt treatment regimens. This conclusion is in line with other findings stating clonal evolution due to treatment pressure [386].

5 Stratification of leukemic blasts across three AML subgroups

In AML, tumorigenesis can happen at every stage of myeloid cell differentiation, and the clinical presentation is quite heterogeneous [139]. Heterogeneity in AML may result from both epigenetic memory of the cell-of-origin as well as genetic lesions that lead to a deregulated transcriptional regulatory circuitry [139,387]. Different methods have so far been employed to analyze the differentiation state. To reliably characterize molecular programs and stratify patient risk, identification of the cell-of-origin is pivotal [47,388]. This was emphasized by a study employing a model for a cytogenetically homogeneous yet clinically heterogeneous AML subgroup. The transcriptome and DNA methylation patterns were distinct in malignant cells with an introduced *MLL-AF9* fusion, depending on the cell-of-origin. More immature stem cell-derived leukemic blasts were associated with a worse prognosis and a reduced sensitivity to medication therapy[47].

Chromatin accessibility in AML revealed distinct regulatory evolution in leukemic cells with a successively higher mutation burden. Individual AML cells were shown to have unique mixed regulome patterns that reflect various maturation differentiation stages [389]. The presence of distinct open chromatin loci in leukemic blasts, which represent the transformed cell-of-origin, raises the possibility that open chromatin patterns might be used as prognostic indicators in AML [190].

Transcriptomic readouts can reveal AML hierarchies as well. A study showed that AML cells could be classified via a machine-learning approach by integrating transcriptional with genetic data [149]. Their computer-generated classifier found six malignant myeloid-like cell identities that could be projected along the axis of HSC to myeloid development and

separated leukemic from normal cells. The tool used in this thesis has the advantage of distinguishing AML subclones in a higher resolution than six subtypes and is easier to use for scientists, non-familiar with machine-learning approaches. This method aims to stratify transcriptomic profiles of various myeloid-like tumor cell types. Although leukemic cells do not express gene signatures seen during typical development but rather a combination of regulatory characteristics, the scoring system based on similarity allows for dissecting heterogeneous blast composition in AML. However, this approach is not intended for annotating intermediate cell states. While the present differentiation state of blasts provides valuable information on tumor heterogeneity and possible treatment options, the annotation of the leukemic blast might not be identical to the cell-of-origin. Tumor initiating cells can lose lineage-determining transcription factors, such as PU.1, leading to dedifferentiation [390]. Conversely, leukemic cells may continue to differentiate along the myeloid trajectory after an oncogenic event at an early developmental stage [391]. These cells might then differentiate further to various degrees [392], affecting the accurate prediction of the cell-of-origin.

6 Conclusion

I dissected transcriptomic and epigenetic deregulation in major AML subgroups by analyzing gene expression and chromatin accessibility patterns. By advancing the necessary experimental and bioinformatic methodologies I could exploit these data to gain insight on deregulated molecular features of AML on various levels. I demonstrated that leukemic cells carrying *MLL-EDC4* exert a unique transcriptomic signature and that scRNA-seq could be used to determine the developmental stages of leukemic blasts. The latter approach is generally applicable and allows to stratify patients based on a cell type assignment of blast and provides an alternative approach to machine learning tools as previously described [149]. Further, genetic lesions in AML were investigated on a sub-clonal level, and functional implications during tumorigenesis and therapy were assessed. My thesis illustrates how tumorigenic and treatment-induced features alter gene regulatory programs. Using an integrative omics approach, I resolved the subclone-specific response during *FLT3* inhibition with midostaurin. The drug treatments' effects on clonal evolution differed for the two patients investigated, which might be caused by different coping strategies to bypass *FLT3*-

inhibition. The generation of a new scDNA-panel to specifically target hundreds of distinct *FLT3*-ITDs will resolve tumor evolution in unprecedented detail for a future study which aims to dissect molecular differences between midostaurin and gilteritinib treatment in *FLT3*-ITD patients.

I further exposed that treatment with drugs targeting mutant *IDH* led to partial reversibility of chromatin accessibility. TFs such as JUN, FOS and CEBPA seemed to be involved in these mutant-specific changes in chromatin accessibility. Regions that were sensitive to treatment showed signs of differentiation indicated by the enrichment of GATA TFs. TFs maintain and shape cancer cell identity and are promising therapeutic targets. The discovery of TFs driving oncogenesis could be exploited for the design of drug candidates that inhibit aberrant TF activity and improve current treatment options [393].

Lastly, leukemic cells of three AML subtypes were classified along the myeloid hematopoietic differentiation axis. The development of cell type prediction of AML blasts based on transcriptomic expression revealed a higher variety of cell populations and a shift in differentiation in relapse compared to diagnosis samples. It has been observed that novel targeted therapeutics can cause leukemic blast maturation [118,227,394]. However, further research is necessary to determine how these differentiation patterns affects clinical outcomes [395].

The genomic, epigenomic, and transcriptomic layers, were studied as distinct readouts from the same sample. They revealed complementary information on epigenetic alterations, cellular hierarchies, and the effects on deregulated gene expression in AML after their bioinformatical integration. Recent advancements allow the simultaneous readout of transcriptome, epigenome, and genome in various combinations from the same cell [227,396,397]. This enables the unbiased exploration of regulatory relationships, such as the connections of regulatory genomic states and their causal changes in transcription. Thus, by applying single cell multi-omics analysis, it is anticipated that our understanding of cell functionalities can be further improved. For example single-cell approaches have the potential to address challenges such as characterization of pre-malignant myeloid cells and the dissection of clonal evolution in LSCs [149] or myeloid-derived suppressor cells [398] (MDSCs) within AML patients.

The implementation of the type of analysis described in this thesis, without the need to infer linkages through computer simulations, will offer insightful information on intra-tumor

heterogeneity in leukemia. Furthermore, my workflow to characterize AML blasts, their myeloid-like sub types and consequently unique TF programs holds great potential for clinical application when extended to larger sample size. While bulk DNA-sequencing is already utilized in clinical practice to identify treatment relevant mutations, the single cell methods described here, have the great potential to detect rare subclones or mutational co-occurrences and link them to regulatory programs that can differ within the same tumor. Dissecting disease- and drug-induced transcriptomic changes and linkage to transcription factor deregulation and chromatin accessibility will extend the usage and significance of sequencing-based techniques for clinical decision-making. Apart from hurdles such as the implementation of adequate preprocessing protocols, the development of analysis workflows and the storage of patient data, costs for single-cell omics experiments might prohibit the study of large patient cohorts [266]. However, it is noted that treatments with new therapeutic agents can easily reach 100 000 € per year and patient and thus quickly exceed the costs of these analysis methods. One study also demonstrated cost reductions and clinical significance for AML patients by combining clinical and genomic data [399]. Further development of combined omics approaches will decrease experimental costs in the next decade that may facilitate the implementation into the clinics and increase treatment effectiveness. The work in this thesis supports the inclusion of single cell omics in such an approach. It illustrates how a more profound understanding of AML disease pathology and drug response was obtained that could be applied in subsequent studies to enhance personalized patient therapies.

References

1 Dick, J. E. Acute myeloid leukemia stem cells. *Ann N Y Acad Sci* **1044**, 1-5, doi:10.1196/annals.1349.001 (2005).

2 Gómez-López, S., Lerner, R. G. & Petritsch, C. Asymmetric cell division of stem and progenitor cells during homeostasis and cancer. *Cell Mol Life Sci* **71**, 575-597, doi:10.1007/s00018-013-1386-1 (2014).

3 Krause, J. R. WHO Classification of Tumours of Haematopoietic and Lymphoid Tissues: An Overview. *Critical Values* **2**, 30-32, doi:10.1093/criticalvalues/2.2.30 (2015).

4 Rad, A. & Häggstöm, M. *Diagram showing the development of different blood cells from haematopoietic stem cell to mature cells*, <https://commons.wikimedia.org/wiki/File:Hematopoiesis_simple.svg> (2009).

5 Khwaja, A. *et al.* Acute myeloid leukaemia. *Nat Rev Dis Primers* **2**, 16010, doi:10.1038/nrdp.2016.10 (2016).

6 Grimwade, D. & Mrózek, K. Diagnostic and prognostic value of cytogenetics in acute myeloid leukemia. *Hematol Oncol Clin North Am* **25**, 1135-1161, vii, doi:10.1016/j.hoc.2011.09.018 (2011).

7 Juliusson, G. *et al.* Age and acute myeloid leukemia: real world data on decision to treat and outcomes from the Swedish Acute Leukemia Registry. *Blood* **113**, 4179-4187, doi:10.1182/blood-2008-07-172007 (2009).

8 Derolf, A. R. *et al.* Improved patient survival for acute myeloid leukemia: a population-based study of 9729 patients diagnosed in Sweden between 1973 and 2005. *Blood* **113**, 3666-3672, doi:10.1182/blood-2008-09-179341 (2009).

9 Burnett, A., Wetzler, M. & Löwenberg, B. Therapeutic advances in acute myeloid leukemia. *J Clin Oncol* **29**, 487-494, doi:10.1200/jco.2010.30.1820 (2011).

10 Grisolano, J. L., O'Neal, J., Cain, J. & Tomasson, M. H. An activated receptor tyrosine kinase, TEL/PDGFbetaR, cooperates with AML1/ETO to induce acute myeloid leukemia in mice. *Proc Natl Acad Sci U S A* **100**, 9506-9511, doi:10.1073/pnas.1531730100 (2003).

11 Ley, T. J. *et al.* Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. *N Engl J Med* **368**, 2059-2074, doi:10.1056/NEJMoa1301689 (2013).

12 Hughes, C. M. *et al.* Menin associates with a trithorax family histone methyltransferase complex and with the hoxc8 locus. *Mol Cell* **13**, 587-597, doi:10.1016/s1097-2765(04)00081-4 (2004).

13 Milne, T. A. *et al.* Menin and MLL cooperatively regulate expression of cyclin-dependent kinase inhibitors. *Proc Natl Acad Sci U S A* **102**, 749-754, doi:10.1073/pnas.0408836102 (2005).

14 Birke, M. *et al.* The MT domain of the proto-oncoprotein MLL binds to CpG-containing DNA and discriminates against methylation. *Nucleic Acids Res* **30**, 958-965, doi:10.1093/nar/30.4.958 (2002).

15 Milne, T. A. *et al.* MLL targets SET domain methyltransferase activity to Hox gene promoters. *Mol Cell* **10**, 1107-1117, doi:10.1016/s1097-2765(02)00741-4 (2002).

16 Tkachuk, D. C., Kohler, S. & Cleary, M. L. Involvement of a homolog of Drosophila trithorax by 11q23 chromosomal translocations in acute leukemias. *Cell* **71**, 691-700, doi:10.1016/0092-8674(92)90602-9 (1992).

17 Chen, J. *et al.* Loss of MLL PHD finger 3 is necessary for MLL-ENL-induced hematopoietic stem cell immortalization. *Cancer Res* **68**, 6199-6207, doi:10.1158/0008-5472.Can-07-6514 (2008).

18 Zeleznik-Le, N. J., Harden, A. M. & Rowley, J. D. 11q23 translocations split the "AT-hook" cruciform DNA-binding region and the transcriptional repression domain from the activation domain of the mixed-lineage leukemia (MLL) gene. *Proc Natl Acad Sci U S A* **91**, 10610-10614, doi:10.1073/pnas.91.22.10610 (1994).

19 Hsieh, J. J., Cheng, E. H. & Korsmeyer, S. J. Taspase1: a threonine aspartase required for cleavage of MLL and proper HOX gene expression. *Cell* **115**, 293-303, doi:10.1016/s0092-8674(03)00816-x (2003).

20 Hsieh, J. J., Ernst, P., Erdjument-Bromage, H., Tempst, P. & Korsmeyer, S. J. Proteolytic cleavage of MLL generates a complex of N- and C-terminal fragments that confers protein stability and subnuclear localization. *Mol Cell Biol* **23**, 186-194, doi:10.1128/mcb.23.1.186-194.2003 (2003).

21 Dou, Y. *et al.* Regulation of MLL1 H3K4 methyltransferase activity by its core components. *Nat Struct Mol Biol* **13**, 713-719, doi:10.1038/nsmb1128 (2006).

22 Li, Y. *et al.* Structural basis for activity regulation of MLL family methyltransferases. *Nature* **530**, 447-452, doi:10.1038/nature16952 (2016).

23 Mishra, B. P. *et al.* The histone methyltransferase activity of MLL1 is dispensable for hematopoiesis and leukemogenesis. *Cell Rep* **7**, 1239-1247, doi:10.1016/j.celrep.2014.04.015 (2014).

24 Winters, A. C. & Bernt, K. M. MLL-Rearranged Leukemias-An Update on Science and Clinical Approaches. *Front Pediatr* **5**, 4, doi:10.3389/fped.2017.00004 (2017).

25 Marschalek, R. Mixed lineage leukemia: roles in human malignancies and potential therapy. *Febs j* **277**, 1822-1831, doi:10.1111/j.1742-4658.2010.07608.x (2010).

26 Hanson, R. D. *et al.* Mammalian Trithorax and polycomb-group homologues are antagonistic regulators of homeotic development. *Proc Natl Acad Sci U S A* **96**, 14372-14377, doi:10.1073/pnas.96.25.14372 (1999).

27 Jude, C. D. *et al.* Unique and independent roles for MLL in adult hematopoietic stem cells and progenitors. *Cell Stem Cell* **1**, 324-337, doi:10.1016/j.stem.2007.05.019 (2007).

28 Abramovich, C. & Humphries, R. K. Hox regulation of normal and leukemic hematopoietic stem cells. *Curr Opin Hematol* **12**, 210-216, doi:10.1097/01.moh.0000160737.52349.aa (2005).

29 Borrow, J. *et al.* The t(7;11)(p15;p15) translocation in acute myeloid leukaemia fuses the genes for nucleoporin NUP98 and class I homeoprotein HOXA9. *Nat Genet* **12**, 159-167, doi:10.1038/ng0296-159 (1996).

30 Meyer, C. *et al.* The MLL recombinome of acute leukemias in 2013. *Leukemia* **27**, 2165-2176, doi:10.1038/leu.2013.135 (2013).

31 Mueller, D. *et al.* Misguided transcriptional elongation causes mixed lineage leukemia. *PLoS Biol* **7**, e1000249, doi:10.1371/journal.pbio.1000249 (2009).

32 Mohan, M. *et al.* Linking H3K79 trimethylation to Wnt signaling through a novel Dot1-containing complex (DotCom). *Genes Dev* **24**, 574-589, doi:10.1101/gad.1898410 (2010).

33 Yokoyama, A., Lin, M., Naresh, A., Kitabayashi, I. & Cleary, M. L. A higher-order complex containing AF4 and ENL family proteins with P-TEFb facilitates oncogenic and physiologic MLL-dependent transcription. *Cancer Cell* **17**, 198-212, doi:10.1016/j.ccr.2009.12.040 (2010).

34 Biswas, D. *et al.* Function of leukemogenic mixed lineage leukemia 1 (MLL) fusion proteins through distinct partner protein complexes. *Proc Natl Acad Sci U S A* **108**, 15751-15756, doi:10.1073/pnas.1111498108 (2011).

35 He, N. *et al.* Human Polymerase-Associated Factor complex (PAFc) connects the Super Elongation Complex (SEC) to RNA polymerase II on chromatin. *Proc Natl Acad Sci U S A* **108**, E636-645, doi:10.1073/pnas.1107107108 (2011).

36 Armstrong, S. A. *et al.* MLL translocations specify a distinct gene expression profile that distinguishes a unique leukemia. *Nat Genet* **30**, 41-47, doi:10.1038/ng765 (2002).

37 Yeoh, E. J. *et al.* Classification, subtype discovery, and prediction of outcome in pediatric acute lymphoblastic leukemia by gene expression profiling. *Cancer Cell* **1**, 133-143, doi:10.1016/s1535-6108(02)00032-6 (2002).

38 Rozovskaia, T. *et al.* Expression profiles of acute lymphoblastic and myeloblastic leukemias with ALL-1 rearrangements. *Proc Natl Acad Sci U S A* **100**, 7853-7858, doi:10.1073/pnas.1132115100 (2003).

39 Kroon, E. *et al.* Hoxa9 transforms primary bone marrow cells through specific collaboration with Meis1a but not Pbx1b. *Embo j* **17**, 3714-3725, doi:10.1093/emboj/17.13.3714 (1998).

40 Li, Z. *et al.* Consistent deregulation of gene expression between human and murine MLL rearrangement leukemias. *Cancer Res* **69**, 1109-1116, doi:10.1158/0008-5472.Can-08-3381 (2009).

41 Lawrence, H. J. *et al.* Mice bearing a targeted interruption of the homeobox gene HOXA9 have defects in myeloid, erythroid, and lymphoid hematopoiesis. *Blood* **89**, 1922-1930 (1997).

42 Yu, B. D., Hanson, R. D., Hess, J. L., Horning, S. E. & Korsmeyer, S. J. MLL, a mammalian trithorax-group gene, functions as a transcriptional maintenance factor in morphogenesis. *Proc Natl Acad Sci U S A* **95**, 10632-10636, doi:10.1073/pnas.95.18.10632 (1998).

43 Yu, B. D., Hess, J. L., Horning, S. E., Brown, G. A. & Korsmeyer, S. J. Altered Hox expression and segmental identity in Mll-mutant mice. *Nature* **378**, 505-508, doi:10.1038/378505a0 (1995).

44 So, C. W., Karsunky, H., Wong, P., Weissman, I. L. & Cleary, M. L. Leukemic transformation of hematopoietic progenitors by MLL-GAS7 in the absence of Hoxa7 or Hoxa9. *Blood* **103**, 3192-3199, doi:10.1182/blood-2003-10-3722 (2004).

45 Kersey, J. H., Wang, D. & Oberto, M. Resistance of t(4;11) (MLL-AF4 fusion gene) leukemias to stress-induced cell death: possible mechanism for extensive extramedullary accumulation of cells and poor prognosis. *Leukemia* **12**, 1561-1564, doi:10.1038/sj.leu.2401148 (1998).

46 Gaussmann, A. *et al.* Combined effects of the two reciprocal t(4;11) fusion proteins MLL.AF4 and AF4.MLL confer resistance to apoptosis, cell cycling capacity and growth transformation. *Oncogene* **26**, 3352-3363, doi:10.1038/sj.onc.1210125 (2007).

47 Krivtsov, A. V. *et al.* Cell of origin determines clinically relevant subtypes of MLL-rearranged AML. *Leukemia* **27**, 852-860, doi:10.1038/leu.2012.363 (2013).

48 Agnès, F. *et al.* Genomic structure of the downstream part of the human FLT3 gene: exon/intron structure conservation among genes encoding receptor tyrosine kinases (RTK) of subclass III. *Gene* **145**, 283-288, doi:10.1016/0378-1119(94)90021-3 (1994).

49 Kazi, J. U. & Rönnstrand, L. FMS-like Tyrosine Kinase 3/FLT3: From Basic Science to Clinical Implications. *Physiol Rev* **99**, 1433-1466, doi:10.1152/physrev.00029.2018 (2019).

50 Stirewalt, D. L. & Radich, J. P. The role of FLT3 in haematopoietic malignancies. *Nat Rev Cancer* **3**, 650-665, doi:10.1038/nrc1169 (2003).

51 Scholl, C., Gilliland, D. G. & Fröhling, S. Deregulation of signaling pathways in acute myeloid leukemia. *Semin Oncol* **35**, 336-345, doi:10.1053/j.seminoncol.2008.04.004 (2008).

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

53 Horiike, S. *et al.* Tandem duplications of the FLT3 receptor gene are associated with leukemic transformation of myelodysplasia. *Leukemia* **11**, 1442-1446, doi:10.1038/sj.leu.2400770 (1997).

54 Schnittger, S. *et al.* Analysis of FLT3 length mutations in 1003 patients with acute myeloid leukemia: correlation to cytogenetics, FAB subtype, and prognosis in the AMLCG study and usefulness as a marker for the detection of minimal residual disease. *Blood* **100**, 59-66, doi:10.1182/blood.v100.1.59 (2002).

55 Kiyoi, H. *et al.* Internal tandem duplication of the FLT3 gene is a novel modality of elongation mutation which causes constitutive activation of the product. *Leukemia* **12**, 1333-1337, doi:10.1038/sj.leu.2401130 (1998).

56 Lee, B. H. *et al.* FLT3 mutations confer enhanced proliferation and survival properties to multipotent progenitors in a murine model of chronic myelomonocytic leukemia. *Cancer Cell* **12**, 367-380, doi:10.1016/j.ccr.2007.08.031 (2007).

57 Thiede, C. *et al.* Analysis of FLT3-activating mutations in 979 patients with acute myelogenous leukemia: association with FAB subtypes and identification of subgroups with poor prognosis. *Blood* **99**, 4326-4335, doi:10.1182/blood.v99.12.4326 (2002).

58 Kavanagh, S. *et al.* Emerging therapies for acute myeloid leukemia: translating biology into the clinic. *JCI Insight* **2**, doi:10.1172/jci.insight.95679 (2017).

59 Willander, K. *et al.* Mutations in the isocitrate dehydrogenase 2 gene and IDH1 SNP 105C > T have a prognostic value in acute myeloid leukemia. *Biomark Res* **2**, 18, doi:10.1186/2050-7771-2-18 (2014).

60 Ogawara, Y. *et al.* IDH2 and NPM1 Mutations Cooperate to Activate Hoxa9/Meis1 and Hypoxia Pathways in Acute Myeloid Leukemia. *Cancer Res* **75**, 2005-2016, doi:10.1158/0008-5472.Can-14-2200 (2015).

61 Wang, G. *et al.* Mutation of isocitrate dehydrogenase 1 induces glioma cell proliferation via nuclear factor-κB activation in a hypoxia-inducible factor 1-α dependent manner. *Mol Med Rep* **9**, 1799-1805, doi:10.3892/mmr.2014.2052 (2014).

62 Fu, Y. *et al.* Glioma derived isocitrate dehydrogenase-2 mutations induced up-regulation of HIF-1α and β-catenin signaling: possible impact on glioma cell

metastasis and chemo-resistance. *Int J Biochem Cell Biol* **44**, 770-775, doi:10.1016/j.biocel.2012.01.017 (2012).

63 Kickingereder, P. *et al.* IDH mutation status is associated with a distinct hypoxia/angiogenesis transcriptome signature which is non-invasively predictable with rCBV imaging in human glioma. *Sci Rep* **5**, 16238, doi:10.1038/srep16238 (2015).

64 Lu, C. *et al.* IDH mutation impairs histone demethylation and results in a block to cell differentiation. *Nature* **483**, 474-478, doi:10.1038/nature10860 (2012).

65 Stein, E. M. Molecular Pathways: IDH2 Mutations-Co-opting Cellular Metabolism for Malignant Transformation. *Clin Cancer Res* **22**, 16-19, doi:10.1158/1078-0432.Ccr-15-0362 (2016).

66 Xu, W. *et al.* Oncometabolite 2-hydroxyglutarate is a competitive inhibitor of α-ketoglutarate-dependent dioxygenases. *Cancer Cell* **19**, 17-30, doi:10.1016/j.ccr.2010.12.014 (2011).

67 Solary, E., Bernard, O. A., Tefferi, A., Fuks, F. & Vainchenker, W. The Ten-Eleven Translocation-2 (TET2) gene in hematopoiesis and hematopoietic diseases. *Leukemia* **28**, 485-496, doi:10.1038/leu.2013.337 (2014).

68 Avellaneda Matteo, D. *et al.* Molecular mechanisms of isocitrate dehydrogenase 1 (IDH1) mutations identified in tumors: The role of size and hydrophobicity at residue 132 on catalytic efficiency. *J Biol Chem* **292**, 7971-7983, doi:10.1074/jbc.M117.776179 (2017).

69 Yang, B., Zhong, C., Peng, Y., Lai, Z. & Ding, J. Molecular mechanisms of "off-on switch" of activities of human IDH1 by tumor-associated mutation R132H. *Cell Res* **20**, 1188-1200, doi:10.1038/cr.2010.145 (2010).

70 Rendina, A. R. *et al.* Mutant IDH1 enhances the production of 2-hydroxyglutarate due to its kinetic mechanism. *Biochemistry* **52**, 4563-4577, doi:10.1021/bi400514k (2013).

71 Clark, O., Yen, K. & Mellinghoff, I. K. Molecular Pathways: Isocitrate Dehydrogenase Mutations in Cancer. *Clin Cancer Res* **22**, 1837-1842, doi:10.1158/1078-0432.Ccr-13-1333 (2016).

72 Chaturvedi, A. *et al.* Enantiomer-specific and paracrine leukemogenicity of mutant IDH metabolite 2-hydroxyglutarate. *Leukemia* **30**, 1708-1715, doi:10.1038/leu.2016.71 (2016).

73 Ye, D., Ma, S., Xiong, Y. & Guan, K. L. R-2-hydroxyglutarate as the key effector of IDH mutations promoting oncogenesis. *Cancer Cell* **23**, 274-276, doi:10.1016/j.ccr.2013.03.005 (2013).

74 Guo, J. U., Su, Y., Zhong, C., Ming, G. L. & Song, H. Emerging roles of TET proteins and 5-hydroxymethylcytosines in active DNA demethylation and beyond. *Cell Cycle* **10**, 2662-2668, doi:10.4161/cc.10.16.17093 (2011).

75 Deplus, R. *et al.* TET2 and TET3 regulate GlcNAcylation and H3K4 methylation through OGT and SET1/COMPASS. *Embo j* **32**, 645-655, doi:10.1038/emboj.2012.357 (2013).

76 Chen, Q., Chen, Y., Bian, C., Fujiki, R. & Yu, X. TET2 promotes histone O-GlcNAcylation during gene transcription. *Nature* **493**, 561-564, doi:10.1038/nature11742 (2013).

77 Quivoron, C. *et al.* TET2 inactivation results in pleiotropic hematopoietic abnormalities in mouse and is a recurrent event during human lymphomagenesis. *Cancer Cell* **20**, 25-38, doi:10.1016/j.ccr.2011.06.003 (2011).

78 Moran-Crusio, K. *et al.* Tet2 loss leads to increased hematopoietic stem cell self-renewal and myeloid transformation. *Cancer Cell* **20**, 11-24, doi:10.1016/j.ccr.2011.06.001 (2011).

79 Ko, M. *et al.* Ten-Eleven-Translocation 2 (TET2) negatively regulates homeostasis and differentiation of hematopoietic stem cells in mice. *Proc Natl Acad Sci U S A* **108**, 14566-14571, doi:10.1073/pnas.1112317108 (2011).

80 Ko, M. *et al.* Impaired hydroxylation of 5-methylcytosine in myeloid cancers with mutant TET2. *Nature* **468**, 839-843, doi:10.1038/nature09586 (2010).

81 Garrett-Bakelman, F. E. & Melnick, A. M. Mutant IDH: a targetable driver of leukemic phenotypes linking metabolism, epigenetics and transcriptional regulation. *Epigenomics* **8**, 945-957, doi:10.2217/epi-2016-0008 (2016).

82 Lev Maor, G., Yearim, A. & Ast, G. The alternative role of DNA methylation in splicing regulation. *Trends Genet* **31**, 274-280, doi:10.1016/j.tig.2015.03.002 (2015).

83 Papaemmanuil, E. *et al.* Genomic Classification and Prognosis in Acute Myeloid Leukemia. *N Engl J Med* **374**, 2209-2221, doi:10.1056/NEJMoa1516192 (2016).

84 Molenaar, R. J. *et al.* Clinical and biological implications of ancestral and non-ancestral IDH1 and IDH2 mutations in myeloid neoplasms. *Leukemia* **29**, 2134-2142, doi:10.1038/leu.2015.91 (2015).

85 Marcucci, G. *et al.* IDH1 and IDH2 gene mutations identify novel molecular subsets within de novo cytogenetically normal acute myeloid leukemia: a Cancer and Leukemia Group B study. *J Clin Oncol* **28**, 2348-2355, doi:10.1200/jco.2009.27.3730 (2010).

86 Patel, K. P. *et al.* Acute myeloid leukemia with IDH1 or IDH2 mutation: frequency and clinicopathologic features. *Am J Clin Pathol* **135**, 35-45, doi:10.1309/ajcpd7nr2rmnqdvf (2011).

87 DiNardo, C. D. *et al.* Characteristics, clinical outcome, and prognostic significance of IDH mutations in AML. *Am J Hematol* **90**, 732-736, doi:10.1002/ajh.24072 (2015).

88 DiNardo, C. D. *et al.* IDH1 and IDH2 mutations in myelodysplastic syndromes and role in disease progression. *Leukemia* **30**, 980-984, doi:10.1038/leu.2015.211 (2016).

89 Wagner, K. *et al.* Impact of IDH1 R132 mutations and an IDH1 single nucleotide polymorphism in cytogenetically normal acute myeloid leukemia: SNP rs11554137 is an adverse prognostic factor. *J Clin Oncol* **28**, 2356-2364, doi:10.1200/jco.2009.27.6899 (2010).

90 Montalban-Bravo, G. & DiNardo, C. D. The role of IDH mutations in acute myeloid leukemia. *Future Oncol* **14**, 979-993, doi:10.2217/fon-2017-0523 (2018).

91 Xu, Q. *et al.* Correlation Between Isocitrate Dehydrogenase Gene Aberrations and Prognosis of Patients with Acute Myeloid Leukemia: A Systematic Review and Meta-Analysis. *Clin Cancer Res* **23**, 4511-4522, doi:10.1158/1078-0432.Ccr-16-2628 (2017).

92 Gupta, R., Webb-Myers, R., Flanagan, S. & Buckland, M. E. Isocitrate dehydrogenase mutations in diffuse gliomas: clinical and aetiological implications. *J Clin Pathol* **64**, 835-844, doi:10.1136/jclinpath-2011-200227 (2011).

93 Döhner, H. *et al.* Diagnosis and management of acute myeloid leukemia in adults: recommendations from an international expert panel, on behalf of the European LeukemiaNet. *Blood* **115**, 453-474, doi:10.1182/blood-2009-07-235358 (2010).

94 Daver, N. *et al.* New directions for emerging therapies in acute myeloid leukemia: the next chapter. *Blood Cancer J* **10**, 107, doi:10.1038/s41408-020-00376-1 (2020).

95 Perry, M. C., Doll, D. C. & Freter, C. E. *The Chemotherapy Source Book*. p.80 (Lippincott Williams & Wilkins (LWW), 2012).

96 Piccart-Gebhart, M. J. Anthracyclines and the tailoring of treatment for early breast cancer. *N Engl J Med* **354**, 2177-2179, doi:10.1056/NEJMe068065 (2006).

97 de Lima, M. *et al.* Implications of potential cure in acute myelogenous leukemia: development of subsequent cancer and return to work. *Blood* **90**, 4719-4724 (1997).

98 Kadia, T. M., Ravandi, F., O'Brien, S., Cortes, J. & Kantarjian, H. M. Progress in acute myeloid leukemia. *Clin Lymphoma Myeloma Leuk* **15**, 139-151, doi:10.1016/j.clml.2014.08.006 (2015).

99 Löwenberg, B. *et al.* High-dose daunorubicin in older patients with acute myeloid leukemia. *N Engl J Med* **361**, 1235-1248, doi:10.1056/NEJMoa0901409 (2009).

100 Kantarjian, H. *et al.* Acute myeloid leukemia: current progress and future directions. *Blood Cancer J* **11**, 41, doi:10.1038/s41408-021-00425-3 (2021).

101 Grimwade, D., Ivey, A. & Huntly, B. J. Molecular landscape of acute myeloid leukemia in younger adults and its clinical relevance. *Blood* **127**, 29-41, doi:10.1182/blood-2015-07-604496 (2016).

102 Tyner, J. W. *et al.* Functional genomic landscape of acute myeloid leukaemia. *Nature* **562**, 526-531, doi:10.1038/s41586-018-0623-z (2018).

103 Shlush, L. I. *et al.* Identification of pre-leukaemic haematopoietic stem cells in acute leukaemia. *Nature* **506**, 328-333, doi:10.1038/nature13038 (2014).

104 Guryanova, O. A. *et al.* DNMT3A mutations promote anthracycline resistance in acute myeloid leukemia via impaired nucleosome remodeling. *Nat Med* **22**, 1488-1495, doi:10.1038/nm.4210 (2016).

105 Busque, L. *et al.* Recurrent somatic TET2 mutations in normal elderly individuals with clonal hematopoiesis. *Nat Genet* **44**, 1179-1181, doi:10.1038/ng.2413 (2012).

106 Genovese, G. *et al.* Clonal hematopoiesis and blood-cancer risk inferred from blood DNA sequence. *N Engl J Med* **371**, 2477-2487, doi:10.1056/NEJMoa1409405 (2014).

107 Gillies, R. J., Verduzco, D. & Gatenby, R. A. Evolutionary dynamics of carcinogenesis and why targeted therapy does not work. *Nat Rev Cancer* **12**, 487-493, doi:10.1038/nrc3298 (2012).

108 Ley, T. J. *et al.* DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome. *Nature* **456**, 66-72, doi:10.1038/nature07485 (2008).

109 Kindler, T., Lipka, D. B. & Fischer, T. FLT3 as a therapeutic target in AML: still challenging after all these years. *Blood* **116**, 5089-5102, doi:10.1182/blood-2010-04-261867 (2010).

110 Smith, C. C. *et al.* Validation of ITD mutations in FLT3 as a therapeutic target in human acute myeloid leukaemia. *Nature* **485**, 260-263, doi:10.1038/nature11016 (2012).

111 Stone, R. M. *et al.* Midostaurin plus Chemotherapy for Acute Myeloid Leukemia with a FLT3 Mutation. *N Engl J Med* **377**, 454-464, doi:10.1056/NEJMoa1614359 (2017).

112 Tamaoki, T. *et al.* Staurosporine, a potent inhibitor of phospholipid/Ca++dependent protein kinase. *Biochem Biophys Res Commun* **135**, 397-402, doi:10.1016/0006-291x(86)90008-2 (1986).

113 Propper, D. J. *et al.* Phase I and pharmacokinetic study of PKC412, an inhibitor of protein kinase C. *J Clin Oncol* **19**, 1485-1492, doi:10.1200/jco.2001.19.5.1485 (2001).

114 Yamamoto, Y. *et al.* Activating mutation of D835 within the activation loop of FLT3 in human hematologic malignancies. *Blood* **97**, 2434-2439, doi:10.1182/blood.v97.8.2434 (2001).

115 Smith, C. C., Lin, K., Stecula, A., Sali, A. & Shah, N. P. FLT3 D835 mutations confer differential resistance to type II FLT3 inhibitors. *Leukemia* **29**, 2390-2392, doi:10.1038/leu.2015.165 (2015).

116 Rydapt. *Mechanism of action of RYDAPT in AML*, <https://www.hcp.novartis.com/products/rydapt/acute-myeloid-leukemia/mechanism-of-action/> (

117 Golub, D. *et al.* Mutant Isocitrate Dehydrogenase Inhibitors as Targeted Cancer Therapeutics. *Front Oncol* **9**, 417, doi:10.3389/fonc.2019.00417 (2019).

118 Fathi, A. T. *et al.* Differentiation Syndrome Associated With Enasidenib, a Selective Inhibitor of Mutant Isocitrate Dehydrogenase 2: Analysis of a Phase 1/2 Study. *JAMA Oncol* **4**, 1106-1110, doi:10.1001/jamaoncol.2017.4695 (2018).

119 Yen, K. *et al.* AG-221, a First-in-Class Therapy Targeting Acute Myeloid Leukemia Harboring Oncogenic IDH2 Mutations. *Cancer Discov* **7**, 478-493, doi:10.1158/2159-8290.Cd-16-1034 (2017).

120 DiNardo, C. D. *et al.* Venetoclax combined with decitabine or azacitidine in treatment-naive, elderly patients with acute myeloid leukemia. *Blood* **133**, 7-17, doi:10.1182/blood-2018-08-868752 (2019).

121 DiNardo, C. D. *et al.* Durable Remissions with Ivosidenib in IDH1-Mutated Relapsed or Refractory AML. *N Engl J Med* **378**, 2386-2398, doi:10.1056/NEJMoa1716984 (2018).

122 Agios-Pharmaceuticals. *Press Release: Agios Announces Withdrawal of European Marketing Authorization Application for TIBSOVO® as a Treatment for Relapsed or Refractory IDH1-mutant Acute Myeloid Leukemia*, <https://investor.agios.com/news-releases/news-release-details/agios-announces-withdrawal-european-marketing-authorization> (2020).

123 Chaturvedi, A. *et al.* In vivo efficacy of mutant IDH1 inhibitor HMS-101 and structural resolution of distinct binding site. *Leukemia* **34**, 416-426, doi:10.1038/s41375-019-0582-x (2020).

124 Chaturvedi, A. *et al.* Pan-mutant-IDH1 inhibitor BAY1436032 is highly effective against human IDH1 mutant acute myeloid leukemia in vivo. *Leukemia* **31**, 2020-2028, doi:10.1038/leu.2017.46 (2017).

125 Bitoun, E., Oliver, P. L. & Davies, K. E. The mixed-lineage leukemia fusion partner AF4 stimulates RNA polymerase II transcriptional elongation and mediates coordinated chromatin remodeling. *Hum Mol Genet* **16**, 92-106, doi:10.1093/hmg/ddl444 (2007).

126 Universitätsklinikum-Heidelberg. *Innere Medizin V: Hämatologie, Onkologie und Rheumatologie - aktuell offene Studien*, <https://www.klinikum.uni-heidelberg.de/kliniken-institute/kliniken/zentrum-fuer-innere-medizin-krehl-klinik/innere-medizin-v-haematologie-onkologie-und-rheumatologie/behandlungsspektrum/haematologie/leukaemien/studien> (2022).

127 Thomas, D. & Majeti, R. Biology and relevance of human acute myeloid leukemia stem cells. *Blood* **129**, 1577-1585, doi:10.1182/blood-2016-10-696054 (2017).

128 Schepers, K., Campbell, T. B. & Passegué, E. Normal and leukemic stem cell niches: insights and therapeutic opportunities. *Cell Stem Cell* **16**, 254-267, doi:10.1016/j.stem.2015.02.014 (2015).

129 Tenen, D. G. Disruption of differentiation in human cancer: AML shows the way. *Nat Rev Cancer* **3**, 89-101, doi:10.1038/nrc989 (2003).

130 Austin, R., Smyth, M. J. & Lane, S. W. Harnessing the immune system in acute myeloid leukaemia. *Crit Rev Oncol Hematol* **103**, 62-77, doi:10.1016/j.critrevonc.2016.04.020 (2016).

131 Bhatia, M. *et al.* Quantitative analysis reveals expansion of human hematopoietic repopulating cells after short-term ex vivo culture. *J Exp Med* **186**, 619-624, doi:10.1084/jem.186.4.619 (1997).

132 Bonnet, D. & Dick, J. E. Human acute myeloid leukemia is organized as a hierarchy that originates from a primitive hematopoietic cell. *Nat Med* **3**, 730-737, doi:10.1038/nm0797-730 (1997).

133 Jordan, C. T. *et al.* The interleukin-3 receptor alpha chain is a unique marker for human acute myelogenous leukemia stem cells. *Leukemia* **14**, 1777-1784, doi:10.1038/sj.leu.2401903 (2000).

134 Du, W. *et al.* Cluster of differentiation 96 as a leukemia stem cell-specific marker and a factor for prognosis evaluation in leukemia. *Mol Clin Oncol* **3**, 833-838, doi:10.3892/mco.2015.552 (2015).

135 Hosen, N. *et al.* CD96 is a leukemic stem cell-specific marker in human acute myeloid leukemia. *Proc Natl Acad Sci U S A* **104**, 11008-11013, doi:10.1073/pnas.0704271104 (2007).

136 Morsink, L. M., Walter, R. B. & Ossenkoppele, G. J. Prognostic and therapeutic role of CLEC12A in acute myeloid leukemia. *Blood Rev* **34**, 26-33, doi:10.1016/j.blre.2018.10.003 (2019).

137 Kikushige, Y. *et al.* TIM-3 is a promising target to selectively kill acute myeloid leukemia stem cells. *Cell Stem Cell* **7**, 708-717, doi:10.1016/j.stem.2010.11.014 (2010).

138 van Rhenen, A. *et al.* The novel AML stem cell associated antigen CLL-1 aids in discrimination between normal and leukemic stem cells. *Blood* **110**, 2659-2666, doi:10.1182/blood-2007-03-083048 (2007).

139 Chopra, M. & Bohlander, S. K. The cell of origin and the leukemia stem cell in acute myeloid leukemia. *Genes Chromosomes Cancer* **58**, 850-858, doi:10.1002/gcc.22805 (2019).

140 Taussig, D. C. *et al.* Leukemia-initiating cells from some acute myeloid leukemia patients with mutated nucleophosmin reside in the CD34(-) fraction. *Blood* **115**, 1976-1984, doi:10.1182/blood-2009-02-206565 (2010).

141 Taussig, D. C. *et al.* Anti-CD38 antibody-mediated clearance of human repopulating cells masks the heterogeneity of leukemia-initiating cells. *Blood* **112**, 568-575, doi:10.1182/blood-2007-10-118331 (2008).

142 Sarry, J. E. *et al.* Human acute myelogenous leukemia stem cells are rare and heterogeneous when assayed in NOD/SCID/IL2Rγc-deficient mice. *J Clin Invest* **121**, 384-395, doi:10.1172/jci41495 (2011).

143 Goardon, N. *et al.* Coexistence of LMPP-like and GMP-like leukemia stem cells in acute myeloid leukemia. *Cancer Cell* **19**, 138-152, doi:10.1016/j.ccr.2010.12.012 (2011).

144 Pollyea, D. A. & Jordan, C. T. Therapeutic targeting of acute myeloid leukemia stem cells. *Blood* **129**, 1627-1635, doi:10.1182/blood-2016-10-696039 (2017).

145 Ho, T. C. *et al.* Evolution of acute myelogenous leukemia stem cell properties after treatment and progression. *Blood* **128**, 1671-1678, doi:10.1182/blood-2016-02-695312 (2016).

146 Lapidot, T. *et al.* A cell initiating human acute myeloid leukaemia after transplantation into SCID mice. *Nature* **367**, 645-648, doi:10.1038/367645a0 (1994).

147 Eppert, K. *et al.* Stem cell gene expression programs influence clinical outcome in human leukemia. *Nat Med* **17**, 1086-1093, doi:10.1038/nm.2415 (2011).

148 Ng, S. W. *et al.* A 17-gene stemness score for rapid determination of risk in acute leukaemia. *Nature* **540**, 433-437, doi:10.1038/nature20598 (2016).

149 van Galen, P. *et al.* Single-Cell RNA-Seq Reveals AML Hierarchies Relevant to Disease Progression and Immunity. *Cell* **176**, 1265-1281.e1224, doi:10.1016/j.cell.2019.01.031 (2019).

150 Lichtenegger, F. S., Krupka, C., Haubner, S., Köhnke, T. & Subklewe, M. Recent developments in immunotherapy of acute myeloid leukemia. *J Hematol Oncol* **10**, 142, doi:10.1186/s13045-017-0505-0 (2017).

151 Pyzer, A. R. *et al.* MUC1-mediated induction of myeloid-derived suppressor cells in patients with acute myeloid leukemia. *Blood* **129**, 1791-1801, doi:10.1182/blood-2016-07-730614 (2017).

152 Milne, T. A. Mouse models of MLL leukemia: recapitulating the human disease. *Blood* **129**, 2217-2223, doi:10.1182/blood-2016-10-691428 (2017).

153 So, C. W. *et al.* MLL-GAS7 transforms multipotent hematopoietic progenitors and induces mixed lineage leukemias in mice. *Cancer Cell* **3**, 161-171, doi:10.1016/s1535-6108(03)00019-9 (2003).

154 Cozzio, A. *et al.* Similar MLL-associated leukemias arising from self-renewing stem cells and short-lived myeloid progenitors. *Genes Dev* **17**, 3029-3035, doi:10.1101/gad.1143403 (2003).

155 Huntly, B. J. *et al.* MOZ-TIF2, but not BCR-ABL, confers properties of leukemic stem cells to committed murine hematopoietic progenitors. *Cancer Cell* **6**, 587-596, doi:10.1016/j.ccr.2004.10.015 (2004).

156 Velten, L. *et al.* Identification of leukemic and pre-leukemic stem cells by clonal tracking from single-cell transcriptomics. *Nat Commun* **12**, 1366, doi:10.1038/s41467-021-21650-1 (2021).

157 Jan, M. *et al.* Clonal evolution of preleukemic hematopoietic stem cells precedes human acute myeloid leukemia. *Sci Transl Med* **4**, 149ra118, doi:10.1126/scitranslmed.3004315 (2012).

158 Abelson, S. *et al.* Prediction of acute myeloid leukaemia risk in healthy individuals. *Nature* **559**, 400-404, doi:10.1038/s41586-018-0317-6 (2018).

159 Desai, P. *et al.* Somatic mutations precede acute myeloid leukemia years before diagnosis. *Nat Med* **24**, 1015-1023, doi:10.1038/s41591-018-0081-z (2018).

160 Schuurhuis, G. J. *et al.* Minimal/measurable residual disease in AML: a consensus document from the European LeukemiaNet MRD Working Party. *Blood* **131**, 1275-1291, doi:10.1182/blood-2017-09-801498 (2018).

161 Saygin, C. *et al.* Genomic analysis of cellular hierarchy in acute myeloid leukemia using ultrasensitive LC-FACSeq. *Leukemia* **35**, 3406-3420, doi:10.1038/s41375-021-01295-1 (2021).

162 Wen, L. & Tang, F. Boosting the power of single-cell analysis. *Nat Biotechnol* **36**, 408-409, doi:10.1038/nbt.4131 (2018).

163 Shema, E., Bernstein, B. E. & Buenrostro, J. D. Single-cell and single-molecule epigenomics to uncover genome regulation at unprecedented resolution. *Nat Genet* **51**, 19-25, doi:10.1038/s41588-018-0290-x (2019).

164 Shalek, A. K. *et al.* Single-cell RNA-seq reveals dynamic paracrine control of cellular variation. *Nature* **510**, 363-369, doi:10.1038/nature13437 (2014).

165 Li, L. & Clevers, H. Coexistence of quiescent and active adult stem cells in mammals. *Science* **327**, 542-545, doi:10.1126/science.1180794 (2010).

166 Huang, S. Non-genetic heterogeneity of cells in development: more than just noise. *Development* **136**, 3853-3862, doi:10.1242/dev.035139 (2009).

167 Maamar, H., Raj, A. & Dubnau, D. Noise in gene expression determines cell fate in Bacillus subtilis. *Science* **317**, 526-529, doi:10.1126/science.1140818 (2007).

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

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

170 Shalek, A. K. *et al.* Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells. *Nature* **498**, 236-240, doi:10.1038/nature12172 (2013).

171 Petropoulos, S. *et al.* Single-Cell RNA-Seq Reveals Lineage and X Chromosome Dynamics in Human Preimplantation Embryos. *Cell* **165**, 1012-1026, doi:10.1016/j.cell.2016.03.023 (2016).

172 Trapnell, C. *et al.* The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat Biotechnol* **32**, 381-386, doi:10.1038/nbt.2859 (2014).

173 Stubbington, M. J. T. *et al.* T cell fate and clonality inference from single-cell transcriptomes. *Nat Methods* **13**, 329-332, doi:10.1038/nmeth.3800 (2016).

174 Nachman, I., Regev, A. & Friedman, N. Inferring quantitative models of regulatory networks from expression data. *Bioinformatics* **20 Suppl 1**, i248-256, doi:10.1093/bioinformatics/bth941 (2004).

175 Liang, S., Fuhrman, S. & Somogyi, R. Reveal, a general reverse engineering algorithm for inference of genetic network architectures. *Pac Symp Biocomput*, 18-29 (1998).

176 Kamimoto, K., Hoffmann, C. M. & Morris, S. A. CellOracle: dissecting cell identity via network inference and
in silico gene perturbation. *bioRxiv* (2020).

177 Hwang, B., Lee, J. H. & Bang, D. Single-cell RNA sequencing technologies and bioinformaticspipelines. *Experimental & Molecular Medicine* **50**, 1-14, doi:10.1038/s12276-018-0071-8 (2018).

178 Kim, K. T. *et al.* Single-cell mRNA sequencing identifies subclonal heterogeneity in anti-cancer drug responses of lung adenocarcinoma cells. *Genome Biol* **16**, 127, doi:10.1186/s13059-015-0692-3 (2015).

179 Emert, B. L. *et al.* Variability within rare cell states enables multiple paths toward drug resistance. *Nat Biotechnol* **39**, 865-876, doi:10.1038/s41587-021-00837-3 (2021).

180 Bar-Joseph, Z., Gitter, A. & Simon, I. Studying and modelling dynamic biological processes using time-series gene expression data. *Nat Rev Genet* **13**, 552-564, doi:10.1038/nrg3244 (2012).

181 Marjanovic, N. D. *et al.* Emergence of a High-Plasticity Cell State during Lung Cancer Evolution. *Cancer Cell* **38**, 229-246.e213, doi:10.1016/j.ccell.2020.06.012 (2020).

182 Young, M. D. *et al.* Single-cell transcriptomes from human kidneys reveal the cellular identity of renal tumors. *Science* **361**, 594-599, doi:10.1126/science.aat1699 (2018).

183 Hanahan, D. Hallmarks of Cancer: New Dimensions. *Cancer Discovery* **12**, 31-46, doi:10.1158/2159-8290.Cd-21-1059 (2022).

184 Smith, Z. D. & Meissner, A. DNA methylation: roles in mammalian development. *Nat Rev Genet* **14**, 204-220, doi:10.1038/nrg3354 (2013).

185 Meissner, A. Epigenetic modifications in pluripotent and differentiated cells. *Nat Biotechnol* **28**, 1079-1088, doi:10.1038/nbt.1684 (2010).

186 Massagué, J. & Ganesh, K. Metastasis-Initiating Cells and Ecosystems. *Cancer Discov* **11**, 971-994, doi:10.1158/2159-8290.Cd-21-0010 (2021).

187 Valencia, A. M. & Kadoch, C. Chromatin regulatory mechanisms and therapeutic opportunities in cancer. *Nat Cell Biol* **21**, 152-161, doi:10.1038/s41556-018-0258-1 (2019).

188 Blanpain, C. Tracing the cellular origin of cancer. *Nat Cell Biol* **15**, 126-134, doi:10.1038/ncb2657 (2013).

189 Ferone, G., Lee, M. C., Sage, J. & Berns, A. Cells of origin of lung cancers: lessons from mouse studies. *Genes Dev* **34**, 1017-1032, doi:10.1101/gad.338228.120 (2020).

190 George, J. *et al.* Leukaemia cell of origin identified by chromatin landscape of bulk tumour cells. *Nat Commun* **7**, 12166, doi:10.1038/ncomms12166 (2016).

191 Visvader, J. E. Cells of origin in cancer. *Nature* **469**, 314-322, doi:10.1038/nature09781 (2011).

192 Gonzalez Castro, L. N., Tirosh, I. & Suvà, M. L. Decoding Cancer Biology One Cell at a Time. *Cancer Discov* **11**, 960-970, doi:10.1158/2159-8290.Cd-20-1376 (2021).

193 Gola, A. & Fuchs, E. Environmental control of lineage plasticity and stem cell memory. *Curr Opin Cell Biol* **69**, 88-95, doi:10.1016/j.ceb.2020.12.015 (2021).

194 Buenrostro, J. D. *et al.* Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature* **523**, 486-490, doi:10.1038/nature14590 (2015).

195 Cusanovich, D. A. *et al.* Multiplex single cell profiling of chromatin accessibility by combinatorial cellular indexing. *Science* **348**, 910-914, doi:10.1126/science.aab1601 (2015).

196 Kaya-Okur, H. S. *et al.* CUT&Tag for efficient epigenomic profiling of small samples and single cells. *Nat Commun* **10**, 1930, doi:10.1038/s41467-019-09982-5 (2019).

197 Kelsey, G., Stegle, O. & Reik, W. Single-cell epigenomics: Recording the past and predicting the future. *Science* **358**, 69-75, doi:10.1126/science.aan6826 (2017).

198 Rotem, A. *et al.* Single-cell ChIP-seq reveals cell subpopulations defined by chromatin state. *Nat Biotechnol* **33**, 1165-1172, doi:10.1038/nbt.3383 (2015).

199 Tata, P. R. *et al.* Developmental History Provides a Roadmap for the Emergence of Tumor Plasticity. *Dev Cell* **44**, 679-693.e675, doi:10.1016/j.devcel.2018.02.024 (2018).

200 LaFave, L. M. *et al.* Epigenomic State Transitions Characterize Tumor Progression in Mouse Lung Adenocarcinoma. *Cancer Cell* **38**, 212-228.e213, doi:10.1016/j.ccell.2020.06.006 (2020).

201 Buenrostro, J. D. *et al.* Integrated Single-Cell Analysis Maps the Continuous Regulatory Landscape of Human Hematopoietic Differentiation. *Cell* **173**, 1535-1548.e1516, doi:10.1016/j.cell.2018.03.074 (2018).

202 Granja, J. M. *et al.* ArchR is a scalable software package for integrative single-cell chromatin accessibility analysis. *Nat Genet* **53**, 403-411, doi:10.1038/s41588-021-00790-6 (2021).

203 Ma, S. *et al.* Chromatin Potential Identified by Shared Single-Cell Profiling of RNA and Chromatin. *Cell* **183**, 1103-1116.e1120, doi:10.1016/j.cell.2020.09.056 (2020).

204 Hnisz, D. *et al.* Super-enhancers in the control of cell identity and disease. *Cell* **155**, 934-947, doi:10.1016/j.cell.2013.09.053 (2013).

205 Breuss, M. W. *et al.* Somatic mosaicism in the mature brain reveals clonal cellular distributions during cortical development. *bioRxiv*, 2020.2008.2010.244814, doi:10.1101/2020.08.10.244814 (2020).

206 Sloan, D. B., Broz, A. K., Sharbrough, J. & Wu, Z. Detecting Rare Mutations and DNA Damage with Sequencing-Based Methods. *Trends Biotechnol* **36**, 729-740, doi:10.1016/j.tibtech.2018.02.009 (2018).

207 Gawad, C., Koh, W. & Quake, S. R. Single-cell genome sequencing: current state of the science. *Nat Rev Genet* **17**, 175-188, doi:10.1038/nrg.2015.16 (2016).

208 Baslan, T. & Hicks, J. Unravelling biology and shifting paradigms in cancer with single-cell sequencing. *Nat Rev Cancer* **17**, 557-569, doi:10.1038/nrc.2017.58 (2017).

209 Bell, A. D. *et al.* Insights into variation in meiosis from 31,228 human sperm genomes. *Nature* **583**, 259-264, doi:10.1038/s41586-020-2347-0 (2020).

210 Ren, X., Kang, B. & Zhang, Z. Understanding tumor ecosystems by single-cell sequencing: promises and limitations. *Genome Biol* **19**, 211, doi:10.1186/s13059-018-1593-z (2018).

211 Dagogo-Jack, I. & Shaw, A. T. Tumour heterogeneity and resistance to cancer therapies. *Nat Rev Clin Oncol* **15**, 81-94, doi:10.1038/nrclinonc.2017.166 (2018).

212 Bakker, B. *et al.* Single-cell sequencing reveals karyotype heterogeneity in murine and human malignancies. *Genome Biol* **17**, 115, doi:10.1186/s13059-016-0971-7 (2016).

213 Stepanauskas, R. Single cell genomics: an individual look at microbes. *Curr Opin Microbiol* **15**, 613-620, doi:10.1016/j.mib.2012.09.001 (2012).

214 Hou, Y. *et al.* Single-cell exome sequencing and monoclonal evolution of a JAK2-negative myeloproliferative neoplasm. *Cell* **148**, 873-885, doi:10.1016/j.cell.2012.02.028 (2012).

215 Hughes, A. E. *et al.* Clonal architecture of secondary acute myeloid leukemia defined by single-cell sequencing. *PLoS Genet* **10**, e1004462, doi:10.1371/journal.pgen.1004462 (2014).

216 Baslan, T. *et al.* Novel insights into breast cancer copy number genetic heterogeneity revealed by single-cell genome sequencing. *Elife* **9**, doi:10.7554/eLife.51480 (2020).

217 Morita, K. *et al.* Publisher Correction: Clonal evolution of acute myeloid leukemia revealed by high-throughput single-cell genomics. *Nat Commun* **11**, 5996, doi:10.1038/s41467-020-19902-7 (2020).

218 Zhang, J. *et al.* Intratumor heterogeneity in localized lung adenocarcinomas delineated by multiregion sequencing. *Science* **346**, 256-259, doi:10.1126/science.1256930 (2014).

219 Wang, Y. *et al.* Clonal evolution in breast cancer revealed by single nucleus genome sequencing. *Nature* **512**, 155-160, doi:10.1038/nature13600 (2014).

220 Navin, N. *et al.* Tumour evolution inferred by single-cell sequencing. *Nature* **472**, 90-94, doi:10.1038/nature09807 (2011).

221 Liu, M. *et al.* Multi-region and single-cell sequencing reveal variable genomic heterogeneity in rectal cancer. *BMC Cancer* **17**, 787, doi:10.1186/s12885-017-3777-4 (2017).

222 Gawad, C., Koh, W. & Quake, S. R. Dissecting the clonal origins of childhood acute lymphoblastic leukemia by single-cell genomics. *Proc Natl Acad Sci U S A* **111**, 17947-17952, doi:10.1073/pnas.1420822111 (2014).

223 Gao, R. *et al.* Punctuated copy number evolution and clonal stasis in triple-negative breast cancer. *Nat Genet* **48**, 1119-1130, doi:10.1038/ng.3641 (2016).

224 Leung, M. L. *et al.* Single-cell DNA sequencing reveals a late-dissemination model in metastatic colorectal cancer. *Genome Res* **27**, 1287-1299, doi:10.1101/gr.209973.116 (2017).

225 Navin, N. E. The first five years of single-cell cancer genomics and beyond. *Genome Res* **25**, 1499-1507, doi:10.1101/gr.191098.115 (2015).

226 Kim, C. *et al.* Chemoresistance Evolution in Triple-Negative Breast Cancer Delineated by Single-Cell Sequencing. *Cell* **173**, 879-893.e813, doi:10.1016/j.cell.2018.03.041 (2018).

227 Demaree, B. *et al.* Joint profiling of DNA and proteins in single cells to dissect genotype-phenotype associations in leukemia. *Nature Communications* **12**, 1583, doi:10.1038/s41467-021-21810-3 (2021).

228 Nam, A. S., Chaligne, R. & Landau, D. A. Integrating genetic and non-genetic determinants of cancer evolution by single-cell multi-omics. *Nat Rev Genet* **22**, 3-18, doi:10.1038/s41576-020-0265-5 (2021).

229 Lareau, C. A. *et al.* Massively parallel single-cell mitochondrial DNA genotyping and chromatin profiling. *Nat Biotechnol* **39**, 451-461, doi:10.1038/s41587-020-0645-6 (2021).

230 Miles, L. A. *et al.* Single-cell mutation analysis of clonal evolution in myeloid malignancies. *Nature* **587**, 477-482, doi:10.1038/s41586-020-2864-x (2020).

231 Bian, S. *et al.* Single-cell multiomics sequencing and analyses of human colorectal cancer. *Science* **362**, 1060-1063, doi:10.1126/science.aao3791 (2018).

232 Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-842, doi:10.1093/bioinformatics/btq033 (2010).

233 Biorender.com. <https://biorender.com/> (

234 Ross-Innes, C. S. *et al.* Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. *Nature* **481**, 389-393, doi:10.1038/nature10730 (2012).

235 Andrew, S. FastQC: A Quality Control Tool for High Throughput Sequence Data. *http://www.bioinformatics.babraham.ac.uk/projects/fastqc/* (2010).

236 Robinson, J. T. *et al.* Integrative genomics viewer. *Nat Biotechnol* **29**, 24-26, doi:10.1038/nbt.1754 (2011).

237 Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9**, R137, doi:10.1186/gb-2008-9-9-r137 (2008).

238 Ewels, P., Magnusson, M., Lundin, S. & Käller, M. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* **32**, 3047-3048, doi:10.1093/bioinformatics/btw354 (2016).

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

240 Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114-2120, doi:10.1093/bioinformatics/btu170 (2014).

241 Gentleman, R. C. *et al.* Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol* **5**, R80, doi:10.1186/gb-2004-5-10-r80 (2004).

242 Gu, Z., Eils, R. & Schlesner, M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* **32**, 2847-2849, doi:10.1093/bioinformatics/btw313 (2016).

243 Gao, R. *et al.* Delineating copy number and clonal substructure in human tumors from single-cell transcriptomes. *Nat Biotechnol* **39**, 599-608, doi:10.1038/s41587-020-00795-2 (2021).

244 Gustavsen, J. A., Pai, S., Isserlin, R., Demchak, B. & Pico, A. R. RCy3: Network biology using Cytoscape from within R. *F1000Res* **8**, 1774, doi:10.12688/f1000research.20887.2 (2019).

245 Dowle, M. *et al.* Package `data.table´. https://github.com/Rdatatable/data.table (2021).

246 Garcia-Alonso, L., Holland, C. H., Ibrahim, M. M., Turei, D. & Saez-Rodriguez, J. Benchmark and integration of resources for the estimation of human transcription factor activities. *Genome Res* **29**, 1363-1375, doi:10.1101/gr.240663.118 (2019).

247 Borcherding, N. & Andrews, J. escape: Easy single cell analysis platform for enrichment. https://bioconductor.org/packages/release/bioc/html/escape.html, doi: 10.18129/B9.bioc.escape (2021).

248 Wickham, H. ggplot2: Elegant Graphics for Data Analysis. https://ggplot2.tidyverse.org (2016).

249 Kremer, L. & Anders, S. ggpointdensity. https://github.com/LKremer/ggpointdensity (2019).

250 Korsunsky, I. *et al.* Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat Methods* **16**, 1289-1296, doi:10.1038/s41592-019-0619-0 (2019).

251 Federico, A. & Monti, S. hypeR: an R package for geneset enrichment workflows. *Bioinformatics* **36**, 1307-1308, doi:10.1093/bioinformatics/btz700 (2020).

252 Harris, C. R. *et al.* Array programming with NumPy. *Nature* **585**, 357-362, doi:10.1038/s41586-020-2649-2 (2020).

253 Wolf, F. A. *et al.* PAGA: graph abstraction reconciles clustering with trajectory inference through a topology preserving map of single cells. *Genome Biol* **20**, 59, doi:10.1186/s13059-019-1663-x (2019).

254 Pandas-Development-Team. pandas-dev/pandas: Pandas. *Zenodo*, doi:10.5281/zenodo.3509134 (2020).

255 Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol* **19**, 15, doi:10.1186/s13059-017-1382-0 (2018).

256 Wolock, S. L., Lopez, R. & Klein, A. M. Scrublet: Computational Identification of Cell Doublets in Single-Cell Transcriptomic Data. *Cell Syst* **8**, 281-291.e289, doi:10.1016/j.cels.2018.11.005 (2019).

257 Hafemeister, C. & Satija, R. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biol* **20**, 296, doi:10.1186/s13059-019-1874-1 (2019).

258 Bergen, V., Lange, M., Peidli, S., Wolf, F. A. & Theis, F. J. Generalizing RNA velocity to transient cell states through dynamical modeling. *Nat Biotechnol* **38**, 1408-1414, doi:10.1038/s41587-020-0591-3 (2020).

259 Satija, R., Farrell, J. A., Gennert, D., Schier, A. F. & Regev, A. Spatial reconstruction of single-cell gene expression data. *Nat Biotechnol* **33**, 495-502, doi:10.1038/nbt.3192 (2015).

260 Aran, D. *et al.* Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. *Nat Immunol* **20**, 163-172, doi:10.1038/s41590-018-0276-y (2019).

261 Street, K. *et al.* Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. *BMC Genomics* **19**, 477, doi:10.1186/s12864-018-4772-0 (2018).

262 La Manno, G. *et al.* RNA velocity of single cells. *Nature* **560**, 494-498, doi:10.1038/s41586-018-0414-6 (2018).

263 Garnier, S., Ross, N., Rudis, R. & Camargo, P. S., M; Scherer, C; . viridis - Colorblind-Friendly Color Maps for R. https://sjmgarnier.github.io/viridis/, doi:10.5281/zenodo.4679424 (2021).

264 Stosch, J. M. *et al.* Gene mutations and clonal architecture in myelodysplastic syndromes and changes upon progression to acute myeloid leukaemia and under treatment. *Br J Haematol* **182**, 830-842, doi:10.1111/bjh.15461 (2018).

265 Kitamura, T. *et al.* Establishment and characterization of a unique human cell line that proliferates dependently on GM-CSF, IL-3, or erythropoietin. *J Cell Physiol* **140**, 323-334, doi:10.1002/jcp.1041400219 (1989).

266 Regev, A. *et al.* The Human Cell Atlas. *Elife* **6**, doi:10.7554/eLife.27041 (2017).

267 Alvarez, M. J. *et al.* Functional characterization of somatic mutations in cancer using network-based inference of protein activity. *Nat Genet* **48**, 838-847, doi:10.1038/ng.3593 (2016).

268 Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* **13**, 2498-2504, doi:10.1101/gr.1239303 (2003).

269 Illumina, I. Nextera library validation and cluster density optimization. (2014).

270 Mamanova, L. *et al.* Target-enrichment strategies for next-generation sequencing. *Nat Methods* **7**, 111-118, doi:10.1038/nmeth.1419 (2010).

271 Kreso, A. & Dick, J. E. Evolution of the cancer stem cell model. *Cell Stem Cell* **14**, 275-291, doi:10.1016/j.stem.2014.02.006 (2014).

272 Tran, T. N. & Bader, G. D. Tempora: Cell trajectory inference using time-series single-cell RNA sequencing data. *PLoS Comput Biol* **16**, e1008205, doi:10.1371/journal.pcbi.1008205 (2020).

273 Stassen, S. V., Yip, G. G. K., Wong, K. K. Y., Ho, J. W. K. & Tsia, K. K. Generalized and scalable trajectory inference in single-cell omics data with VIA. *Nature Communications* **12**, 5528, doi:10.1038/s41467-021-25773-3 (2021).

274 Fan, J., Slowikowski, K. & Zhang, F. Single-cell transcriptomics in cancer: computational challenges and opportunities. *Experimental & Molecular Medicine* **52**, 1452-1465, doi:10.1038/s12276-020-0422-0 (2020).

275 Friedmann-Morvinski, D. & Verma, I. M. Dedifferentiation and reprogramming: origins of cancer stem cells. *EMBO Rep* **15**, 244-253, doi:10.1002/embr.201338254 (2014).

276 Becker, H. *et al.* Identification of enhancer of mRNA decapping 4 as a novel fusion partner of MLL in acute myeloid leukemia. *Blood Adv* **3**, 761-765, doi:10.1182/bloodadvances.2018023879 (2019).

277 Yu, J. H., Yang, W. H., Gulick, T., Bloch, K. D. & Bloch, D. B. Ge-1 is a central component of the mammalian cytoplasmic mRNA processing body. *Rna* **11**, 1795-1802, doi:10.1261/rna.2142405 (2005).

278 Janapala, Y., Preiss, T. & Shirokikh, N. E. Control of Translation at the Initiation Phase During Glucose Starvation in Yeast. *Int J Mol Sci* **20**, doi:10.3390/ijms20164043 (2019).

279 Teixeira, D. & Parker, R. Analysis of P-body assembly in Saccharomyces cerevisiae. *Mol Biol Cell* **18**, 2274-2287, doi:10.1091/mbc.e07-03-0199 (2007).

280 Seto, E., Yoshida-Sugitani, R., Kobayashi, T. & Toyama-Sorimachi, N. The Assembly of EDC4 and Dcp1a into Processing Bodies Is Critical for the Translational Regulation of IL-6. *PLoS One* **10**, e0123223, doi:10.1371/journal.pone.0123223 (2015).

281 Mikuda, N. *et al.* The IkappaB kinase complex is a regulator of mRNA stability. *EMBO J* **37**, doi:10.15252/embj.201798658 (2018).

282 Levine, J. H. *et al.* Data-Driven Phenotypic Dissection of AML Reveals Progenitor-like Cells that Correlate with Prognosis. *Cell* **162**, 184-197, doi:10.1016/j.cell.2015.05.047 (2015).

283 Menendez-Gonzalez, J. B. *et al.* Inhibition of GATA2 restrains cell proliferation and enhances apoptosis and chemotherapy mediated apoptosis in human GATA2 overexpressing AML cells. *Sci Rep* **9**, 12212, doi:10.1038/s41598-019-48589-0 (2019).

284 Argiropoulos, B. & Humphries, R. K. Hox genes in hematopoiesis and leukemogenesis. *Oncogene* **26**, 6766-6776, doi:10.1038/sj.onc.1210760 (2007).

285 Tsapogas, P., Mooney, C. J., Brown, G. & Rolink, A. The Cytokine Flt3-Ligand in Normal and Malignant Hematopoiesis. *Int J Mol Sci* **18**, doi:10.3390/ijms18061115 (2017).

286 Li, B. E. & Ernst, P. Two decades of leukemia oncoprotein epistasis: the MLL1 paradigm for epigenetic deregulation in leukemia. *Exp Hematol* **42**, 995-1012, doi:10.1016/j.exphem.2014.09.006 (2014).

287 Lee, J. W. *et al.* Regulation of HOXA9 activity by predominant expression of DACH1 against C/EBPα and GATA-1 in myeloid leukemia with MLL-AF9. *Biochem Biophys Res Commun* **426**, 299-305, doi:10.1016/j.bbrc.2012.08.048 (2012).

288 Ahmadi, S. E., Rahimi, S., Zarandi, B., Chegeni, R. & Safa, M. MYC: a multipurpose oncogene with prognostic and therapeutic implications in blood malignancies. *J Hematol Oncol* **14**, 121, doi:10.1186/s13045-021-01111-4 (2021).

289 Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* **102**, 15545-15550, doi:10.1073/pnas.0506580102 (2005).

290 Urbańska, K. & Orzechowski, A. Unappreciated Role of LDHA and LDHB to Control Apoptosis and Autophagy in Tumor Cells. *Int J Mol Sci* **20**, doi:10.3390/ijms20092085 (2019).

291 Du, Y. *et al.* ATPR triggers acute myeloid leukaemia cells differentiation and cycle arrest via the RARα/LDHB/ERK-glycolysis signalling axis. *J Cell Mol Med* **24**, 6952-6965, doi:10.1111/jcmm.15353 (2020).

292 Pellicano, F. *et al.* hsa-mir183/EGR1-mediated regulation of E2F1 is required for CML stem/progenitor cell survival. *Blood* **131**, 1532-1544, doi:10.1182/blood-2017-05-783845 (2018).

293 Saint-André, V. *et al.* Models of human core transcriptional regulatory circuitries. *Genome Res* **26**, 385-396, doi:10.1101/gr.197590.115 (2016).

294 Lulli, V. *et al.* Transcriptional silencing of the ETS1 oncogene contributes to human granulocytic differentiation. *Haematologica* **95**, 1633-1641, doi:10.3324/haematol.2010.023267 (2010).

295 Dang, C. V. MYC on the path to cancer. *Cell* **149**, 22-35, doi:10.1016/j.cell.2012.03.003 (2012).

296 Vázquez-Arreguín, K. & Tantin, D. The Oct1 transcription factor and epithelial malignancies: Old protein learns new tricks. *Biochim Biophys Acta* **1859**, 792-804, doi:10.1016/j.bbagrm.2016.02.007 (2016).

297 Moser, B., Edtmayer, S., Witalisz-Siepracka, A. & Stoiber, D. The Ups and Downs of STAT Inhibition in Acute Myeloid Leukemia. *Biomedicines* **9**, doi:10.3390/biomedicines9081051 (2021).

298 Gurnari, C., Falconi, G., De Bellis, E., Voso, M. T. & Fabiani, E. The Role of Forkhead Box Proteins in Acute Myeloid Leukemia. *Cancers (Basel)* **11**, doi:10.3390/cancers11060865 (2019).

299 Cho, E. C., Mitton, B. & Sakamoto, K. M. CREB and leukemogenesis. *Crit Rev Oncog* **16**, 37-46, doi:10.1615/critrevoncog.v16.i1-2.50 (2011).

300 Jiang, Y. *et al.* CEBPG promotes acute myeloid leukemia progression by enhancing EIF4EBP1. *Cancer Cell International* **21**, 598, doi:10.1186/s12935-021-02305-z (2021).

301 Aryal, S., Zhang, Y., Wren, S., Li, C. & Lu, R. Molecular regulators of HOXA9 in acute myeloid leukemia. *Febs j*, doi:10.1111/febs.16268 (2021).

302 Tian, T., Li, X. & Zhang, J. mTOR Signaling in Cancer and mTOR Inhibitors in Solid Tumor Targeting Therapy. *Int J Mol Sci* **20**, doi:10.3390/ijms20030755 (2019).

303 Miscia, S. *et al.* Tumor necrosis factor alpha (TNF-alpha) activates Jak1/Stat3-Stat5B signaling through TNFR-1 in human B cells. *Cell Growth Differ* **13**, 13-18 (2002).

304 Amaya, M. L. *et al.* The STAT3-MYC axis promotes survival of leukemia stem cells by regulating SLC1A5 and oxidative phosphorylation. *Blood* **139**, 584-596, doi:10.1182/blood.2021013201 (2022).

305 Cauchy, P. *et al.* Chronic FLT3-ITD Signaling in Acute Myeloid Leukemia Is Connected to a Specific Chromatin Signature. *Cell Rep* **12**, 821-836, doi:10.1016/j.celrep.2015.06.069 (2015).

306 Spitzer, B. *et al.* Characterizing Transcriptional and Epigenetic Signatures Induced By FLT3-ITD Activation. *Blood* **124**, 2186, doi:https://doi.org/10.1182/blood.V124.21.2186.2186 (2014).

307 Wu, X. *et al.* Bcl11a controls Flt3 expression in early hematopoietic progenitors and is required for pDC development in vivo. *PLoS One* **8**, e64800, doi:10.1371/journal.pone.0064800 (2013).

308 Sandén, C. *et al.* Clonal competition within complex evolutionary hierarchies shapes AML over time. *Nat Commun* **11**, 579, doi:10.1038/s41467-019-14106-0 (2020).

309 MissionBio. *Interpreting missed targets*, <https://support.missionbio.com/hc/en-us/articles/4404883911575-Interpreting-missed-targets-using-Tapestri-Designer> (

310 Mason, R. E., Likar, I., Biern, R. O. & Ross, R. S. Multiple-lead exercise electrocardiography. Experience in 107 normal subjects and 67 patients with angina pectoris, and comparison with coronary cinearteriography in 84 patients. *Circulation* **36**, 517-525 (1967).

311 Mondesir, J., Willekens, C., Touat, M. & de Botton, S. IDH1 and IDH2 mutations as novel therapeutic targets: current perspectives. *J Blood Med* **7**, 171-180, doi:10.2147/JBM.S70716 (2016).

312 Pusch, S. *et al.* Pan-mutant IDH1 inhibitor BAY 1436032 for effective treatment of IDH1 mutant astrocytoma in vivo. *Acta Neuropathol* **133**, 629-644, doi:10.1007/s00401-017-1677-y (2017).

313 Ernst, J. & Kellis, M. ChromHMM: automating chromatin-state discovery and characterization. *Nat Methods* **9**, 215-216, doi:10.1038/nmeth.1906 (2012).

314 An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57-74, doi:10.1038/nature11247 (2012).

315 Fu, Y.-K. *et al.* Gata2-L359V impairs primitive and definitive hematopoiesis and blocks cell differentiation in murine chronic myelogenous leukemia model. *Cell Death & Disease* **12**, 568, doi:10.1038/s41419-021-03826-1 (2021).

316 Fathi, A. T. *et al.* Differentiation Syndrome Associated With Enasidenib, a Selective Inhibitor of Mutant Isocitrate Dehydrogenase 2: Analysis of a Phase 1/2 Study. *JAMA Oncology* **4**, 1106-1110, doi:10.1001/jamaoncol.2017.4695 (2018).

317 Menezes, A. C. *et al.* Increased expression of RUNX3 inhibits normal human myeloid development. *Leukemia* **36**, 1769-1780, doi:10.1038/s41375-022-01577-2 (2022).

318 Bullinger, L. CEBPA mutations in AML: site matters. *Blood* **139**, 6-7, doi:10.1182/blood.2021013557 (2022).

319 Lord, K. A., Abdollahi, A., Hoffman-Liebermann, B. & Liebermann, D. A. Proto-oncogenes of the fos/jun family of transcription factors are positive regulators of myeloid differentiation. *Mol Cell Biol* **13**, 841-851, doi:10.1128/mcb.13.2.841-851.1993 (1993).

320 NandyMazumdar, M. *et al.* BACH1, the master regulator of oxidative stress, has a dual effect on CFTR expression. *Biochemical Journal* **478**, 3741-3756, doi:10.1042/bcj20210252 (2021).

321 Xu, X., Xiong, X. & Sun, Y. The role of ribosomal proteins in the regulation of cell proliferation, tumorigenesis, and genomic integrity. *Sci China Life Sci* **59**, 656-672, doi:10.1007/s11427-016-0018-0 (2016).

322 Tillinger, W., Jilch, R., Waldhoer, T., Reinisch, W. & Junger, W. Monocyte human leukocyte antigen-DR expression-a tool to distinguish intestinal bacterial infections from inflammatory bowel disease? *Shock* **40**, 89-94, doi:10.1097/SHK.0b013e318299ebdd (2013).

323 Chen, H. Z., Tsai, S. Y. & Leone, G. Emerging roles of E2Fs in cancer: an exit from cell cycle control. *Nat Rev Cancer* **9**, 785-797, doi:10.1038/nrc2696 (2009).

324 Unruh, D. *et al.* Methylation and transcription patterns are distinct in IDH mutant gliomas compared to other IDH mutant cancers. *Sci Rep* **9**, 8946, doi:10.1038/s41598-019-45346-1 (2019).

325 Jimenez-Useche, I. & Yuan, C. The effect of DNA CpG methylation on the dynamic conformation of a nucleosome. *Biophys J* **103**, 2502-2512, doi:10.1016/j.bpj.2012.11.012 (2012).

326 Wu, J. *et al.* SMAD1 as a biomarker and potential therapeutic target in drug-resistant multiple myeloma. *Biomark Res* **9**, 48, doi:10.1186/s40364-021-00296-7 (2021).

327 Grimm, D. *et al.* The role of SOX family members in solid tumours and metastasis. *Semin Cancer Biol* **67**, 122-153, doi:10.1016/j.semcancer.2019.03.004 (2020).

328 Julian, L. M., McDonald, A. C. & Stanford, W. L. Direct reprogramming with SOX factors: masters of cell fate. *Curr Opin Genet Dev* **46**, 24-36, doi:10.1016/j.gde.2017.06.005 (2017).

329 Leung, R. K. C., Leung, H. C. & Leung, A. Y. H. Diverse pathogenetic roles of SOX genes in acute myeloid leukaemia and their therapeutic implications. *Seminars in Cancer Biology* **67**, 24-29, doi:https://doi.org/10.1016/j.semcancer.2019.11.001 (2020).

330 Miyazaki, T. *et al.* Expression of heme oxygenase-1 in human leukemic cells and its regulation by transcriptional repressor Bach1. *Cancer Sci* **101**, 1409-1416, doi:10.1111/j.1349-7006.2010.01550.x (2010).

331 Bradner, J. E., Hnisz, D. & Young, R. A. Transcriptional Addiction in Cancer. *Cell* **168**, 629-643, doi:10.1016/j.cell.2016.12.013 (2017).

332 Garraway, L. A. & Lander, E. S. Lessons from the cancer genome. *Cell* **153**, 17-37, doi:10.1016/j.cell.2013.03.002 (2013).

333 Suva, M. L., Riggi, N. & Bernstein, B. E. Epigenetic reprogramming in cancer. *Science* **339**, 1567-1570, doi:10.1126/science.1230184 (2013).

334 Laugesen, A. & Helin, K. Chromatin repressive complexes in stem cells, development, and cancer. *Cell Stem Cell* **14**, 735-751, doi:10.1016/j.stem.2014.05.006 (2014).

335 Guillaumet-Adkins, A. *et al.* Single-cell transcriptome conservation in cryopreserved cells and tissues. *Genome Biol* **18**, 45, doi:10.1186/s13059-017-1171-9 (2017).

336 Massoni-Badosa, R. *et al.* Sampling time-dependent artifacts in single-cell genomics studies. *Genome Biol* **21**, 112, doi:10.1186/s13059-020-02032-0 (2020).

337 Slyper, M. *et al.* A single-cell and single-nucleus RNA-Seq toolbox for fresh and frozen human tumors. *Nat Med* **26**, 792-802, doi:10.1038/s41591-020-0844-1 (2020).

338 Saultz, J. N. & Garzon, R. Acute Myeloid Leukemia: A Concise Review. *J Clin Med* **5**, doi:10.3390/jcm5030033 (2016).

339 Brady, S. W. *et al.* Combating subclonal evolution of resistant cancer phenotypes. *Nat Commun* **8**, 1231, doi:10.1038/s41467-017-01174-3 (2017).

340 Venteicher, A. S. *et al.* Decoupling genetics, lineages, and microenvironment in IDH-mutant gliomas by single-cell RNA-seq. *Science* **355**, doi:10.1126/science.aai8478 (2017).

341 Fan, J. *et al.* Linking transcriptional and genetic tumor heterogeneity through allele analysis of single-cell RNA-seq data. *Genome Res* **28**, 1217-1227, doi:10.1101/gr.228080.117 (2018).

342 Patel, A. P. *et al.* Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science* **344**, 1396-1401, doi:10.1126/science.1254257 (2014).

343 Wang, J. *et al.* Identification of Potential Novel Prognosis-Related Genes Through Transcriptome Sequencing, Bioinformatics Analysis, and Clinical Validation in Acute Myeloid Leukemia. *Front Genet* **12**, 723001, doi:10.3389/fgene.2021.723001 (2021).

344 Docking, T. R. *et al.* A clinical transcriptome approach to patient stratification and therapy selection in acute myeloid leukemia. *Nature Communications* **12**, 2474, doi:10.1038/s41467-021-22625-y (2021).

345 Jakoš, T., Pišlar, A., Jewett, A. & Kos, J. Cysteine Cathepsins in Tumor-Associated Immune Cells. *Front Immunol* **10**, 2037, doi:10.3389/fimmu.2019.02037 (2019).

346 Wang, J. & Li, Y. CD36 tango in cancer: signaling pathways and functions. *Theranostics* **9**, 4893-4908, doi:10.7150/thno.36037 (2019).

347 Mikuda, N. *et al.* The IκB kinase complex is a regulator of mRNA stability. *Embo j* **37**, doi:10.15252/embj.201798658 (2018).

348 Dunn, G. P., Bruce, A. T., Ikeda, H., Old, L. J. & Schreiber, R. D. Cancer immunoediting: from immunosurveillance to tumor escape. *Nat Immunol* **3**, 991-998, doi:10.1038/ni1102-991 (2002).

349 Warner, J. R. & McIntosh, K. B. How common are extraribosomal functions of ribosomal proteins? *Mol Cell* **34**, 3-11, doi:10.1016/j.molcel.2009.03.006 (2009).

350 Corden, J. L. & Patturajan, M. A CTD function linking transcription to splicing. *Trends Biochem Sci* **22**, 413-416, doi:10.1016/s0968-0004(97)01125-0 (1997).

351 Estable, M. C. *et al.* MCEF, the newest member of the AF4 family of transcription factors involved in leukemia, is a positive transcription elongation factor-b-associated protein. *J Biomed Sci* **9**, 234-245, doi:10.1007/bf02256070 (2002).

352 Pagano, G. *et al.* Bone marrow cell transcripts from Fanconi anaemia patients reveal in vivo alterations in mitochondrial, redox and DNA repair pathways. *Eur J Haematol* **91**, 141-151, doi:10.1111/ejh.12131 (2013).

353 Richardson, C., Yan, S. & Vestal, C. G. Oxidative stress, bone marrow failure, and genome instability in hematopoietic stem cells. *Int J Mol Sci* **16**, 2366-2385, doi:10.3390/ijms16022366 (2015).

354 Porporato, P. E., Dhup, S., Dadhich, R. K., Copetti, T. & Sonveaux, P. Anticancer targets in the glycolytic metabolism of tumors: a comprehensive review. *Front Pharmacol* **2**, 49, doi:10.3389/fphar.2011.00049 (2011).

355 Hensley, C. T., Wasti, A. T. & DeBerardinis, R. J. Glutamine and cancer: cell biology, physiology, and clinical opportunities. *J Clin Invest* **123**, 3678-3684, doi:10.1172/jci69600 (2013).

356 Warburg, O. On the origin of cancer cells. *Science* **123**, 309-314, doi:10.1126/science.123.3191.309 (1956).

357 Mullen, A. R. *et al.* Reductive carboxylation supports growth in tumour cells with defective mitochondria. *Nature* **481**, 385-388, doi:10.1038/nature10642 (2011).

358 Sola-Penna, M. Metabolic regulation by lactate. *IUBMB Life* **60**, 605-608, doi:10.1002/iub.97 (2008).

359 Brisson, L. *et al.* Lactate Dehydrogenase B Controls Lysosome Activity and Autophagy in Cancer. *Cancer Cell* **30**, 418-431, doi:https://doi.org/10.1016/j.ccell.2016.08.005 (2016).

360 Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**, 663-676, doi:10.1016/j.cell.2006.07.024 (2006).

361 Kottaridis, P. D. *et al.* The presence of a FLT3 internal tandem duplication in patients with acute myeloid leukemia (AML) adds important prognostic information to cytogenetic risk group and response to the first cycle of chemotherapy: analysis of 854 patients from the United Kingdom Medical Research Council AML 10 and 12 trials. *Blood* **98**, 1752-1759, doi:10.1182/blood.v98.6.1752 (2001).

362 Scholl, S. *et al.* Clinical impact of nucleophosmin mutations and Flt3 internal tandem duplications in patients older than 60 yr with acute myeloid leukaemia. *Eur J Haematol* **80**, 208-215, doi:10.1111/j.1600-0609.2007.01019.x (2008).

363 McMahon, C. M. *et al.* Clonal Selection with RAS Pathway Activation Mediates Secondary Clinical Resistance to Selective FLT3 Inhibition in Acute Myeloid Leukemia. *Cancer Discov* **9**, 1050-1063, doi:10.1158/2159-8290.Cd-18-1453 (2019).

364 Foo, J. & Michor, F. Evolution of acquired resistance to anti-cancer therapy. *J Theor Biol* **355**, 10-20, doi:10.1016/j.jtbi.2014.02.025 (2014).

365 Schmalbrock, L. K. *et al.* Clonal evolution of acute myeloid leukemia with FLT3-ITD mutation under treatment with midostaurin. *Blood* **137**, 3093-3104, doi:10.1182/blood.2020007626 (2021).

366 Morita, K. *et al.* Clonal evolution of acute myeloid leukemia revealed by high-throughput single-cell genomics. *Nature Communications* **11**, 5327, doi:10.1038/s41467-020-19119-8 (2020).

367 Tirier, S. M. *et al.* Subclone-specific microenvironmental impact and drug response in refractory multiple myeloma revealed by single-cell transcriptomics. *Nat Commun* **12**, 6960, doi:10.1038/s41467-021-26951-z (2021).

368 Smith, C. C. *et al.* Heterogeneous resistance to quizartinib in acute myeloid leukemia revealed by single-cell analysis. *Blood* **130**, 48-58, doi:10.1182/blood-2016-04-711820 (2017).

369 Zhang, H. *et al.* Clinical resistance to crenolanib in acute myeloid leukemia due to diverse molecular mechanisms. *Nat Commun* **10**, 244, doi:10.1038/s41467-018-08263-x (2019).

370 Lam, S. S. Y. & Leung, A. Y. H. Overcoming Resistance to FLT3 Inhibitors in the Treatment of FLT3-Mutated AML. *Int J Mol Sci* **21**, doi:10.3390/ijms21041537 (2020).

371 Man, C. H. *et al.* A novel tescalcin-sodium/hydrogen exchange axis underlying sorafenib resistance in FLT3-ITD+ AML. *Blood* **123**, 2530-2539, doi:10.1182/blood-2013-07-512194 (2014).

372 Sato, T. *et al.* FLT3 ligand impedes the efficacy of FLT3 inhibitors in vitro and in vivo. *Blood* **117**, 3286-3293, doi:10.1182/blood-2010-01-266742 (2011).

373 Green, A. S. *et al.* Pim kinases modulate resistance to FLT3 tyrosine kinase inhibitors in FLT3-ITD acute myeloid leukemia. *Sci Adv* **1**, e1500221, doi:10.1126/sciadv.1500221 (2015).

374 Park, I. K. *et al.* Receptor tyrosine kinase Axl is required for resistance of leukemic cells to FLT3-targeted therapy in acute myeloid leukemia. *Leukemia* **29**, 2382-2389, doi:10.1038/leu.2015.147 (2015).

375 Lindblad, O. *et al.* Aberrant activation of the PI3K/mTOR pathway promotes resistance to sorafenib in AML. *Oncogene* **35**, 5119-5131, doi:10.1038/onc.2016.41 (2016).

376 Piloto, O. *et al.* Prolonged exposure to FLT3 inhibitors leads to resistance via activation of parallel signaling pathways. *Blood* **109**, 1643-1652, doi:10.1182/blood-2006-05-023804 (2007).

377 Takahashi, S. Downstream molecular pathways of FLT3 in the pathogenesis of acute myeloid leukemia: biology and therapeutic implications. *J Hematol Oncol* **4**, 13, doi:10.1186/1756-8722-4-13 (2011).

378 Volpe, G. *et al.* C/EBPα and MYB regulate FLT3 expression in AML. *Leukemia* **27**, 1487-1496, doi:10.1038/leu.2013.23 (2013).

379 Morel, D., Jeffery, D., Aspeslagh, S., Almouzni, G. & Postel-Vinay, S. Combining epigenetic drugs with other therapies for solid tumours - past lessons and future promise. *Nat Rev Clin Oncol* **17**, 91-107, doi:10.1038/s41571-019-0267-4 (2020).

380 Struhl, K. Is DNA methylation of tumour suppressor genes epigenetic? *Elife* **3**, e02475, doi:10.7554/eLife.02475 (2014).

381 Turcan, S. *et al.* Mutant-IDH1-dependent chromatin state reprogramming, reversibility, and persistence. *Nat Genet* **50**, 62-72, doi:10.1038/s41588-017-0001-z (2018).

382 West, A. C. & Johnstone, R. W. New and emerging HDAC inhibitors for cancer treatment. *J Clin Invest* **124**, 30-39, doi:10.1172/jci69738 (2014).

383 Heuser, M. *et al.* Safety and efficacy of BAY1436032 in IDH1-mutant AML: phase I study results. *Leukemia* **34**, 2903-2913, doi:10.1038/s41375-020-0996-5 (2020).

384 Moustakas, A., Souchelnytskyi, S. & Heldin, C. H. Smad regulation in TGF-beta signal transduction. *J Cell Sci* **114**, 4359-4369, doi:10.1242/jcs.114.24.4359 (2001).

385 Hou, X., Zhang, J., Wang, Y., Xiong, W. & Mi, J. TGFBR-IDH1-Cav1 axis promotes TGF-β signalling in cancer-associated fibroblast. *Oncotarget* **8**, 83962-83974, doi:10.18632/oncotarget.20861 (2017).

386 Misund, K. *et al.* Clonal evolution after treatment pressure in multiple myeloma: heterogenous genomic aberrations and transcriptomic convergence. *Leukemia* **36**, 1887-1897, doi:10.1038/s41375-022-01597-y (2022).

387 Lee, T. I. & Young, R. A. Transcriptional regulation and its misregulation in disease. *Cell* **152**, 1237-1251, doi:10.1016/j.cell.2013.02.014 (2013).

388 Zeller, C. *et al.* Adverse stem cell clones within a single patient's tumor predict clinical outcome in AML patients. *J Hematol Oncol* **15**, 25, doi:10.1186/s13045-022-01232-4 (2022).

389 Corces, M. R. *et al.* Lineage-specific and single-cell chromatin accessibility charts human hematopoiesis and leukemia evolution. *Nat Genet* **48**, 1193-1203, doi:10.1038/ng.3646 (2016).

390 McKenzie, M. D. *et al.* Interconversion between Tumorigenic and Differentiated States in Acute Myeloid Leukemia. *Cell Stem Cell* **25**, 258-272.e259, doi:https://doi.org/10.1016/j.stem.2019.07.001 (2019).

391 Vicente-Dueñas, C., Gutiérrez de Diego, J., Rodríguez, F. D., Jiménez, R. & Cobaleda, C. The role of cellular plasticity in cancer development. *Curr Med Chem* **16**, 3676-3685, doi:10.2174/092986709789105019 (2009).

392 Jögi, A., Vaapil, M., Johansson, M. & Påhlman, S. Cancer cell differentiation heterogeneity and aggressive behavior in solid tumors. *Ups J Med Sci* **117**, 217-224, doi:10.3109/03009734.2012.659294 (2012).

393 Bushweller, J. H. Targeting transcription factors in cancer - from undruggable to reality. *Nat Rev Cancer* **19**, 611-624, doi:10.1038/s41568-019-0196-7 (2019).

394 Wang, F. *et al.* Targeted inhibition of mutant IDH2 in leukemia cells induces cellular differentiation. *Science* **340**, 622-626, doi:10.1126/science.1234769 (2013).

395 Yun, H. D. *et al.* Erythroid differentiation of myeloblast induced by gilteritinib in relapsed FLT3-ITD–positive acute myeloid leukemia. *Blood Advances* **3**, 3709-3712, doi:10.1182/bloodadvances.2019000775 (2019).

396 Cao, J. *et al.* Joint profiling of chromatin accessibility and gene expression in thousands of single cells. *Science* **361**, 1380-1385, doi:10.1126/science.aau0730 (2018).

397 Liu, L. *et al.* Deconvolution of single-cell multi-omics layers reveals regulatory heterogeneity. *Nat Commun* **10**, 470, doi:10.1038/s41467-018-08205-7 (2019).

398 Veglia, F., Perego, M. & Gabrilovich, D. Myeloid-derived suppressor cells coming of age. *Nat Immunol* **19**, 108-119, doi:10.1038/s41590-017-0022-x (2018).

399 Gerstung, M. *et al.* Precision oncology for acute myeloid leukemia using a knowledge bank approach. *Nat Genet* **49**, 332-340, doi:10.1038/ng.3756 (2017).

Suppl. table 1 Gene list for integration

Genes

ABCA9	CD14	CST3	EMP1	HIST1H2BK	JUN	MRPL33	PIWIL4	S100A8	SRSF6
ABCB1	CD163	CSTA	EP400NL	HLA-DMA	KBTBD11	MRPL57	PKIB	S100A9	ST3GAL1
ACSL1	CD300E	CTSC	EPS8	HLA-DMB	KIF2A	MS4A10	PLAC8	S100B	TAPT1-AS1
ACTB	CD68	CTSG	F13A1	HLA-DPA1	KMT2A	MS4A3	PLBD1	SAMHD1	TFDP2
AK2	CD70	CTSS	FAM107B	HLA-DPB1	LGALS2	MSI2	PLCB4	SCRN1	THBS1
ALDH2	CD74	CXCL2	FAM30A	HLA-DPB2	LILRB3	NACA2	PLD3	SELENOP	TLR4
ANKRD28	CDK6	CXCL8	FAM74A4	HLA-DQA1	LINC01623	NAMPT	PRDX6	SELL	TMEM25
APOBEC3A	CDKN2A	CYP1B1	FCER1A	HLA-DQA2	LINC01770	NAP1L1	PRLR	SERPINA1	TMEM70
AQP9	CEBPB	DBI	FCN1	HLA-DQB1	LOC107984974	NAPSB	PROK2	SERPINB10	TMEM74
ATP5G2	CFAP61	DDIT3	FEZ1	HLA-DQB2	LOC643802	NCF1	PSAP	SERPINB2	TMSB10
ATP8B4	CLEC10A	DEFB1	FPR1	HLA-DRA	LOC644936	NCF2	PSME1	SESN3	TNFAIP2
B4GALT6	CLEC4A	DHRS9	FRMD3	HLA-DRB1	LPL	NEAT1	PTGFR	SH3BGRL3	TP53INP2
BCL2A1	CLEC4E	DSE	FTH1P3	HLA-DRB5	LRP1	NFKBIA	PTPRCAP	SLC11A1	TPM4
BCL6	CLEC7A	DST	FTL	HLA-DRB6	LRRC75A-AS1	NOP53	PYCARD	SLC24A4	TPSB2
C20orf203	CLNK	DUSP1	FUT4	HNRNPA1	LRRK2	NPDC1	RACK1	SLC44A1	TPSD1
C5AR1	CLU	EAF2	GABARAP	HOPX	LY86	NPTX2	RBPMS	SLC6A13	TPT1
CACNB4	CMTM2	EBPL	GABRA4	HOXB-AS3	MAFB	NRIP1	RETN	SLPI	TSC22D1
CALCRL	COTL1	ECRP	GNG11	HSH2D	MECOM	PAK1	RFLNB	SMIM24	TTLL10
CAP1	CPA3	EEF1A1	GNPTAB	HSP90AB1	MEF2C	PALLD	RFX8	SNHG6	UMODL1-AS1
CCDC144NL-AS1	CPVL	EEF1B2	GPR183	HSPA5	MEFV	PARP1	RNASE2	SNHG8	VCAN
CCDC152	CR1	EEF1G	GPX1	IFNGR1	MEGF9	PCNP	RNASE3	SOX4	VNN2
CCDC18-AS1	CRHBP	EGFL7	GRB2	IGSF6	MEIS1	PDLIM1	RNF217	SPINK2	XIRP2
CCL23	CRIP1	EIF4B	H1F0	IL31RA	MEST	PDZRN4	RNVU1-6	SPN	ZBTB20
CCR2	CRISPLD2	ELF1	H2AFY	ITGB7	MPPED2	PEBP1	S100A10	SPON1	ZFR
CCS	CSF1R	EMB	HINT1	JAML	MRC1	PHACTR3	S100A12	SRGN	ZPBP

Suppl. table 2 Gene list for module score calculation

HSC		MPP		GMP	Mono	preDC		cDC	
ANKRD28	RPS5	BTF3	RPSA	AC020656.1	CD14	ACTG1	SPINK2	ACTB	RGS10
AVP	RPS6	C1QTNF4	SMIM24	AREG	CSTA	C12orf75	SRP14	ACTG1	S100A10
C6orf48	RPS8	DUT	SPINK2	AZU1	CTSS	CCDC50	STMN1	ARPC2	SNX3
CD164	RPSA	EEF1B2	STMN1	CALR	CXCL8	CD74	TCF4	C1orf162	TAGLN2
EEF1B2	SNHG8	EEF2	TUBA1B	CLEC11A	DUSP1	H2AFZ	TUBA1B	C1orf54	TMSB4X
EEF2	SPINK2	EIF3E	TUBB	CST7	FCN1	HIST1H4C	TUBB	CD74	TUBA1B
EIF3E	ZFAS1	ENO1	UBB	CTSG	FOS	HMGB1	UBB	CLEC9A	TXN
FAM30A		GYPC		DUT	FTL	HMGB2		CPNE3	
HINT1		HINT1		ELANE	G0S2	HMGN1		CPVL	
HNRNPA1		HIST1H4C		FABP5	NAMPT	HMGN2		CST3	
HOPX		HMGA1		H2AFZ	NEAT1	HNRNPA1		DNASE1L3	
HSP90AB1		HMGB1		HMGN1	NFKBIA	HNRNPA2B1		EEF1B2	
LDHB		HNRNPA1		HSP90B1	RGS2	HSP90AA1		GSTP1	
NOP53		HSP90AA1		HSPB1	S100A12	IGLL1		HLA-C	
NPM1		HSP90AB1		IGLL1	S100A4	IRF8		HLA-DMA	
PRDX1		IGLL1		LDHB	S100A6	ITM2C		HLA-DMB	
RACK1		LDHB		MPO	S100A8	LDHB		HLA-DPA1	
RPL10A		NPM1		MS4A3	S100A9	NPM1		HLA-DPB1	
RPL15		NUCB2		NPM1	SLC2A3	NUCB2		HLA-DQA1	
RPL3		PRDX1		NUCB2	TYROBP	NUCKS1		HLA-DQA2	
RPL30		PRSS57		PLAC8	VCAN	PCLAF		HLA-DQB1	
RPL31		RPL10A		PRSS57	ZFP36L1	PCNA		HLA-DRA	
RPL4		RPL3				PLAC8		HLA-DRB1	
RPL5		RPL4				PLD4		HLA-DRB5	
RPL7A		RPL5				PLP2		ID2	
RPLP0		RPL7A				PPIB		IRF8	
RPS12		RPLP0				RPSA		LMNA	
RPS18		RPS18				SCT		LSP1	
RPS2		RPS3				SEC61B		NAP1L1	
RPS23		RPS4X				SEPT6		PPA1	
RPS3		RPS5				SLC25A5		PPT1	
RPS4X		RPS6				SOX4		PSMB9	

Suppl. table 3 List of unique ITDs

ITD length [bp]	Insertion site chr 13 [bp]	Insertion site domain
48	28608292	exon 14 JMD switch motif
33	28608291	exon 14 JMD switch motif
27	28608288	exon 14 JMD switch motif
45	28608288	exon 14 JMD switch motif
30	28608286	exon 14 JMD switch motif
36	28608286	exon 14 JMD switch motif
39	28608286	exon 14 JMD switch motif
42	28608286	exon 14 JMD switch motif
54	28608286	exon 14 JMD switch motif
42	28608279	exon 14 JMD zipper motif
30	28608278	exon 14 JMD zipper motif
30	28608277	exon 14 JMD zipper motif
36	28608277	exon 14 JMD zipper motif
39	28608277	exon 14 JMD zipper motif
51	28608277	exon 14 JMD zipper motif
60	28608277	exon 14 JMD zipper motif
66	28608277	exon 14 JMD zipper motif
33	28608276	exon 14 JMD zipper motif
36	28608276	exon 14 JMD zipper motif
63	28608276	exon 14 JMD zipper motif
24	28608275	exon 14 JMD zipper motif
33	28608275	exon 14 JMD zipper motif
60	28608275	exon 14 JMD zipper motif
27	28608274	exon 14 JMD zipper motif
39	28608274	exon 14 JMD zipper motif
45	28608274	exon 14 JMD zipper motif
36	28608273	exon 14 JMD zipper motif
54	28608273	exon 14 JMD zipper motif
69	28608273	exon 14 JMD zipper motif
42	28608272	exon 14 JMD zipper motif
9	28608271	exon 14 JMD zipper motif
36	28608271	exon 14 JMD zipper motif
39	28608269	exon 14 JMD zipper motif
12	28608268	exon 14 JMD zipper motif
15	28608268	exon 14 JMD zipper motif
18	28608268	exon 14 JMD zipper motif
30	28608268	exon 14 JMD zipper motif
33	28608268	exon 14 JMD zipper motif
36	28608268	exon 14 JMD zipper motif
42	28608268	exon 14 JMD zipper motif
57	28608268	exon 14 JMD zipper motif
66	28608268	exon 14 JMD zipper motif
72	28608268	exon 14 JMD zipper motif
21	28608267	exon 14 JMD zipper motif
27	28608267	exon 14 JMD zipper motif
30	28608267	exon 14 JMD zipper motif
33	28608267	exon 14 JMD zipper motif
54	28608267	exon 14 JMD zipper motif
24	28608266	exon 14 JMD zipper motif
48	28608266	exon 14 JMD zipper motif
18	28608265	exon 14 JMD zipper motif
21	28608265	exon 14 JMD zipper motif
33	28608265	exon 14 JMD zipper motif
18	28608264	exon 14 JMD zipper motif
21	28608264	exon 14 JMD zipper motif
57	28608264	exon 14 JMD zipper motif
57	28608263	exon 14 JMD zipper motif
15	28608262	exon 14 JMD zipper motif
18	28608262	exon 14 JMD zipper motif
21	28608262	exon 14 JMD zipper motif
24	28608262	exon 14 JMD zipper motif
27	28608262	exon 14 JMD zipper motif
30	28608262	exon 14 JMD zipper motif

39	28608262	exon 14 JMD zipper motif
45	28608262	exon 14 JMD zipper motif
48	28608262	exon 14 JMD zipper motif
51	28608262	exon 14 JMD zipper motif
54	28608262	exon 14 JMD zipper motif
60	28608262	exon 14 JMD zipper motif
63	28608262	exon 14 JMD zipper motif
78	28608262	exon 14 JMD zipper motif
21	28608261	exon 14 JMD zipper motif
54	28608261	exon 14 JMD zipper motif
63	28608261	exon 14 JMD zipper motif
9	28608260	exon 14 JMD zipper motif
21	28608260	exon 14 JMD zipper motif
27	28608260	exon 14 JMD zipper motif
36	28608260	exon 14 JMD zipper motif
48	28608260	exon 14 JMD zipper motif
51	28608260	exon 14 JMD zipper motif
27	28608259	exon 14 JMD zipper motif
54	28608258	exon 14 JMD zipper motif
69	28608258	exon 14 JMD zipper motif
24	28608257	exon 14 JMD zipper motif
45	28608257	exon 14 JMD zipper motif
51	28608257	exon 14 JMD zipper motif
57	28608257	exon 14 JMD zipper motif
54	28608256	exon 14 JMD zipper motif
66	28608256	exon 14 JMD zipper motif
78	28608256	exon 14 JMD zipper motif
21	28608255	exon 14 JMD zipper motif
36	28608255	exon 14 JMD zipper motif
42	28608255	exon 14 JMD zipper motif
39	28608254	exon 14 JMD zipper motif
63	28608254	exon 14 JMD zipper motif
72	28608254	exon 14 JMD zipper motif
18	28608253	exon 14 JMD zipper motif
36	28608253	exon 14 JMD zipper motif
66	28608253	exon 14 JMD zipper motif
69	28608253	exon 14 JMD zipper motif
75	28608253	exon 14 JMD zipper motif
30	28608252	exon 14 JMD zipper motif
33	28608252	exon 14 JMD zipper motif
63	28608252	exon 14 JMD zipper motif
66	28608252	exon 14 JMD zipper motif
78	28608252	exon 14 JMD zipper motif
21	28608251	exon 14 JMD zipper motif
24	28608251	exon 14 JMD zipper motif
21	28608250	exon 14 JMD zipper motif
27	28608250	exon 14 JMD zipper motif
33	28608250	exon 14 JMD zipper motif
36	28608250	exon 14 JMD zipper motif
18	28608249	exon 14 JMD zipper motif
21	28608249	exon 14 JMD zipper motif
39	28608249	exon 14 JMD zipper motif
60	28608249	exon 14 JMD zipper motif
75	28608249	exon 14 JMD zipper motif
18	28608248	exon 14 JMD zipper motif
27	28608247	exon 14 JMD zipper motif
66	28608247	exon 14 JMD zipper motif
27	28608246	exon 14 JMD hinge region
69	28608246	exon 14 JMD hinge region
27	28608245	exon 14 JMD hinge region
30	28608245	exon 14 JMD hinge region
60	28608245	exon 14 JMD hinge region
66	28608245	exon 14 JMD hinge region
24	28608244	exon 14 JMD hinge region
42	28608244	exon 14 JMD hinge region
96	28608244	exon 14 JMD hinge region
66	28608243	exon 14 JMD hinge region

75	28608243	exon 14 JMD hinge region
24	28608242	exon 14 JMD hinge region
30	28608242	exon 14 JMD hinge region
78	28608242	exon 14 JMD hinge region
27	28608241	exon 14 JMD hinge region
72	28608241	exon 14 JMD hinge region
69	28608240	exon 14 JMD hinge region
78	28608238	exon 14 JMD hinge region
90	28608238	exon 14 JMD hinge region
33	28608237	exon 14 JMD hinge region
66	28608237	exon 14 JMD hinge region
72	28608237	exon 14 JMD hinge region
30	28608236	exon 14 JMD hinge region
36	28608236	exon 14 JMD hinge region
72	28608236	exon 14 JMD hinge region
78	28608236	exon 14 JMD hinge region
81	28608235	exon 14 JMD hinge region
45	28608234	exon 14 JMD hinge region
87	28608233	exon 14 JMD hinge region
24	28608232	exon 14 JMD hinge region
27	28608232	exon 14 JMD hinge region
45	28608232	exon 14 JMD hinge region
51	28608232	exon 14 JMD hinge region
30	28608231	exon 14 JMD hinge region
36	28608231	exon 14 JMD hinge region
48	28608231	exon 14 JMD hinge region
54	28608231	exon 14 JMD hinge region
57	28608231	exon 14 JMD hinge region
72	28608231	exon 14 JMD hinge region
90	28608231	exon 14 JMD hinge region
45	28608230	exon 14 JMD hinge region
81	28608230	exon 14 JMD hinge region
84	28608230	exon 14 JMD hinge region
54	28608229	exon 14 JMD hinge region
51	28608228	exon 14 TKD1 beta 1 sheet
69	28608228	exon 14 TKD1 beta 1 sheet
75	28608228	exon 14 TKD1 beta 1 sheet
81	28608228	exon 14 TKD1 beta 1 sheet
87	28608228	exon 14 TKD1 beta 1 sheet
84	28608227	exon 14 TKD1 beta 1 sheet
90	28608227	exon 14 TKD1 beta 1 sheet
48	28608226	exon 14 TKD1 beta 1 sheet
60	28608226	exon 14 TKD1 beta 1 sheet
66	28608226	exon 14 TKD1 beta 1 sheet
96	28608226	exon 14 TKD1 beta 1 sheet
36	28608225	exon 14 TKD1 beta 1 sheet
42	28608225	exon 14 TKD1 beta 1 sheet
51	28608225	exon 14 TKD1 beta 1 sheet
66	28608225	exon 14 TKD1 beta 1 sheet
72	28608225	exon 14 TKD1 beta 1 sheet
90	28608225	exon 14 TKD1 beta 1 sheet
48	28608224	exon 14 TKD1 beta 1 sheet
84	28608224	exon 14 TKD1 beta 1 sheet
90	28608224	exon 14 TKD1 beta 1 sheet
96	28608224	exon 14 TKD1 beta 1 sheet
33	28608223	exon 14 TKD1 beta 1 sheet
54	28608223	exon 14 TKD1 beta 1 sheet
60	28608223	exon 14 TKD1 beta 1 sheet
75	28608223	exon 14 TKD1 beta 1 sheet
87	28608223	exon 14 TKD1 beta 1 sheet
93	28608222	exon 14 TKD1 beta 1 sheet
69	28608221	exon 14 TKD1 beta 1 sheet
81	28608221	exon 14 TKD1 beta 1 sheet
93	28608221	exon 14 TKD1 beta 1 sheet
93	28608221	exon 14 TKD1 beta 1 sheet
48	28608220	exon 14 TKD1 beta 1 sheet
66	28608220	exon 14 TKD1 beta 1 sheet

93	28608220	exon 14 TKD1 beta 1 sheet
108	28608220	exon 14 TKD1 beta 1 sheet
21	28608219	exon 14 TKD1 beta 1 sheet
48	28608219	exon 14 TKD1 beta 1 sheet
57	28608219	exon 14 TKD1 beta 1 sheet
60	28608219	exon 14 TKD1 beta 1 sheet
78	28608219	exon 14 TKD1 beta 1 sheet
36	28608218	intron 14 splice donor
42	28608218	intron 14 splice donor
48	28608218	intron 14 splice donor
51	28608218	intron 14 splice donor
54	28608218	intron 14 splice donor
57	28608218	intron 14 splice donor
60	28608218	intron 14 splice donor
66	28608218	intron 14 splice donor
84	28608218	intron 14 splice donor
87	28608218	intron 14 splice donor
80	28608218	intron 14 splice donor
81	28608218	intron 14 splice donor
137	28608217	intron 14 splice donor
179	28608217	intron 14 splice donor
186	28608216	intron 14
197	28608216	intron 14
188	28608216	intron 14
240	28608216	intron 14
34	28608216	intron 14
176	28608216	intron 14

Suppl. table 4 Overview of amplicons for custom scDNA-seq panel

Amp ID	Chr	Amp start	Insert start	Insert end	Amp end	Forw sequence	Rev sequence
137462	1	36932094	36932118	36932342	36932362	AACCAGAGGTTCTCATAGGACTTG	GAAAAGAAGCCGGTGCCCTG
137463	1	36932363	36932384	36932529	36932552	ATCCTCCTCCAGCACTGTGAG	CCCTTTGTGTTCCACCAGTAACA
121260	1	36933031	36933051	36933262	36933282	GGGCTGGGACTCTCAGACAT	TAACTTCTTCCGGCCCTGCC
222768	1	36933313	36933334	36933539	36933559	CCTCCGACCAGGGGATTCAAA	CATCCTGAATGCCTCCTCCC
3592	1	43814903	43814923	43815083	43815107	GGGCCGAAGTCTGACCCTTT	CGAACCAAGAATGCCTGTTTACAG
112922	1	115256292	115256318	115256536	115256561	TCAAACAACCTAAAACCAACTCTTCC	CTGTTTGTTGGACATACTGGATACA
113413	1	115258503	115258534	115258748	115258768	AAAATGAACTGTTCTCTATAAACACGTTAAG	AACTGGTGGTGGTTGGAGCA
54970	2	25457110	25457133	25457344	25457363	TAACTTTGTGTCGCTACCTCAGT	GTGTGGTTAGACGGCTTCC
121710	2	25458456	25458476	25458697	25458724	CCCACTGTTCCCAGGACGTT	GCTTATTCCTCTTTTCTCCTCTTCATC
121267	2	25459597	25459617	25459840	25459863	CGACACTCACCAGGGAGGAA	CCACTGTGAATGATAAGCTGGAG
124395	2	25461831	25461853	25462077	25462100	CTGACTCCTCTTCCTCCTCAAG	GTCTCCTGTTTTGTAGTCCAACC
198115	2	25463085	25463107	25463317	25463339	TCCACAATGCAGATGAGACAGG	TTTCTCTTCCGACCTCTCAGAG
137474	2	25463346	25463366	25463588	25463611	AGCTGGGGCTGTCTGCATAG	TTATCCTCCCAGATCCAGGAGTG
130080	2	25464311	25464332	25464554	25464575	TGGGGCAAAGGGTGAAGAGAA	GCTCCTGGTGCTGAAGGACTT
113843	2	25466552	25466572	25466801	25466821	GGCTGCTTTACCACCTGTTT	GCTGAGAAGAGGAAGCCCAT
124400	2	25467008	25467028	25467194	25467213	CAGGCCCAGCACTCACAAAT	TTCCAGGTGCTTTTGCGTG
198120	2	25467264	25467284	25467495	25467515	GGTGTGCTACCTGGAATGGA	TTTCTGGAGTGTGCGTACCA
130084	2	25467961	25467981	25468195	25468217	GCCCGTGTCATCAGGACTCT	TCCCTCTGCTTTCCAGACATCT
137480	2	25468677	25468697	25468892	25468911	GGAGTCCCACACCCTGAAGA	GCAGAAGTGCCGGAACATT
222784	2	25469005	25469025	25469173	25469195	CCTCTCCAGAAGCAGGCCAA	CTGCTTTCCTCCTCCAGAAGAG
113724	2	25469465	25469485	25469661	25469681	AGAAAGCTGGGTGCCCTCAT	TCCTGACAACCCCAACCCTG
113725	2	25469766	25469786	25470011	25470032	GGAGGAGCGGGATGTGCATA	CCCAGGTGTGTGTTGAGAAGC
10420	2	25470404	25470425	25470621	25470641	CCCTGGGATCAAGAACCTTCC	ATTTCTGCTCCTTGGGGCTC
121281	2	25470886	25470905	25471109	25471130	CTGGTGAAGAAGCCGCTCA	TTTCCCCAGGCTGAGAAGAAA
124408	2	25472308	25472327	25472558	25472577	CAGACGCCATGCTGGAGTT	TGTGAGAAGGAATGGGCGC
121282	2	25505239	25505259	25505485	25505508	GAGCCAAGTCCCTGACTCTC	GCCTCAGAGCTATTACCCAATGG
137487	2	25536656	25536676	25536899	25536922	CACACCCTGTCGTGAGCACT	ATAATTCCTTCCCCAAAGCCCAG
222792	2	25961484	25961509	25961714	25961735	CCCCATTATAGCAAATCACTTCACC	AAGAGCTGGCAAATGGGAAAC
222793	2	25963429	25963460	25963673	25963698	GTTCCTTAACCATGACCTAAATATTATGTAC	CACACACACACAATACAAGACATTT
222794	2	25966876	25966896	25967087	25967107	GGGGTTTGCTGTAGTTGTGC	GTCCAGGAGAGGGTGGTGAA
222795	2	25967126	25967146	25967293	25967315	CAGGTCCTGGAATGGTCCCT	TTTCATCCATCGCAGGTCTCTC
222796	2	25967325	25967347	25967489	25967514	GGAGATTCTGGAGACCGGGATC	AACTAGTATTGCTGTCTCCAAGGTT
113138	2	25972337	25972360	25972585	25972606	CCAGGTACCTTCCTTATGCAAAC	AGAGGGGACAGAATCCAGGTG

192463	2	25972892	25972915	25973138	25973161	ATTAGAAGCTGTGGTG AGATGGT	ATAGTTCCAGTAG TCTCCCAGTC
222799	2	25982274	25982304	25982519	25982543	ACTTAGTACCATTAATA CCACTTTACAAGT	GCCCAAACGATTC TTCTCTGTTTT
222800	2	26014395	26014428	26014571	26014604	TGTAAATAATAAGACT ATCACTAAAGGTTAAC C	TAAGATACTATTC TTCTCTGGAGAGT TTAATCA
222801	2	26024007	26024027	26024229	26024260	AGGCTGGTCCTGGATT TGAA	TATAATGTCATAC GGATGTCAGTTTT ATGTA
222802	2	26063315	26063340	26063540	26063572	GCAACTGTCACCATCA TCCATCTTT	TATAAAATTCTGA TACATGCTACAGT ATGGAT
222803	2	26092073	26092095	26092312	26092342	CCTGTAGTCCCAAGCTT CTCCA	TCCTTTAAGACAG ACTTTCTTTTAAAG TGT
222804	2	26099445	26099472	26099682	26099714	GGTCCAGGTCTGCTTA CCTATAATTAA	CAATTTGTATCTCT CTTAGTTGGAAAT GTAAA
222805	2	26101054	26101075	26101301	26101321	CAGGTCCTGCCCTTCTT CCTA	CAATATGGCAGCG GCGGTAG
130091	2	198265978	198266007	198266224	198266247	TGACTAAAATCCATCTC CTTTCATAATCA	GGTAAAACAGTGT TGTGGGACAG
222807	2	198266413	198266440	198266631	198266664	TGTTAGAACCATGAAA CATATCCAGTT	GGCGACATAAATC TAAATTACTAAAG TACATAT
130093	2	198266669	198266702	198266836	198266858	CTTTAATGAAGATAAA TCAAAAGGTAATTGGT G	GTAGGTCTTGTGG ATGAGCAGC
130094	2	198267299	198267326	198267543	198267568	TAAACTTCTAAGATGT GGCAAGATGGC	TCTTTGTTTACATT TTAGGCTGCTG
222810	2	198267571	198267601	198267771	198267797	GGACAGTCATGAGTTG GTAATATTAATCTT	ATGTGAAAGTGTA GCTTCTTCTCTTT
113430	2	209113073	209113100	209113322	209113342	TTATTGCCAACATGACT TACTTGATCC	AGTCACCAAGGAT GCTGCAG
124414	3	128199949	128199969	128200116	128200138	GATGTGTCCGGAGTGG CTGA	GAAGGAAGGGAT CCAGACTCGG
130097	3	128200484	128200504	128200732	128200751	CACCTCCTGAGCAGAG GCAA	TGTCAGACGACAA CCACCA
130098	3	128202566	128202589	128202816	128202835	GACATCCCAGTGCTTTT CATGAT	TGTCAACTGTGGG GCCACA
222815	4	55561577	55561605	55561801	55561826	TATTGTAGAGTACACA GAAGATGGAACT	ATCCAGGATCTCA AAAGTCCATTTG
124418	4	55569682	55569709	55569927	55569951	AATGACTGTCTTTCAAC ATAATCCTGT	GCCTGACGTTCAT AATTGAAGTCA
113736	4	55589556	55589575	55589778	55589797	AGCCTTTCTGGGTTGG ACC	GGAGCATGCCATT CACGAG
137500	4	55592005	55592032	55592226	55592255	TTGTTTTCTTCCCTTTA GATGCTCTGC	CCTAAACATCCCC TTAAATTGGATTA AAA
121297	4	55593449	55593475	55593691	55593713	ATGTGCATTATTGTGAT GATTCTGAC	CTGACCAAAACTC AGCCTGTTT
121298	4	55594013	55594032	55594262	55594282	TCGGGAAGGTTGTTGA GGC	ATGGTGCAGGCTC CAAGTAG
137503	4	55599194	55599226	55599360	55599383	AAATGAATTTAAATGG TTTTCTTTTCTCCTCC	TGTCAAGCAGAGA ATGGGTACTC
124423	4	55602443	55602468	55602688	55602712	ATAGTGCTCACATCTTA AAATGGGG	CACAGTTGAAAAT GCTTTCAGGTG
130114	4	106155020	106155049	106155236	106155260	AACACATTTTAATTTTT GTTTCCATGCTC	CACTTGGTGTCTC CATTTACTTCT
124425	4	106155310	106155332	106155557	106155579	AATAGTCGTGTGAGTC CTGACT	AACTGCATTTTCTT GGGCTACA
222825	4	106155609	106155629	106155793	106155822	CTGCAGTGGGCCTGAA AATC	CGCAATGGAAACA CAATCTGGATAAT ATT
121302	4	106156002	106156026	106156246	106156271	TGATAATGCCAGTAAA CTAGCTGC	GAATCCTTAGTGA ACACTGAGCTTT
121303	4	106156303	106156325	106156547	106156572	TCACAATTGCTTCTTTC TCCCC	CTGTATGTCATTCC TGTTCACACAA

222828	4	106156630	106156650	106156870	106156895	CACCTCAAGCATAACCCACC	TGATTGGAGAGATTGGGTTGATACT
222829	4	106156903	106156930	106157147	106157170	TCCAAACAATACACTGGAAATTCCAAC	TTTGGGAATCTGCTCTTTGTTGA
113866	4	106157305	106157327	106157539	106157565	TCACATCTCCCTCAAAACCAGC	GGTCTGACTATAAGGGGAATTTCTAC
113867	4	106157746	106157769	106157995	106158015	TTTCAAGAACAGGAGCAGAAGTC	AGACTCAGTTTGGGGTTGCT
222832	4	106158026	106158047	106158275	106158295	AGATGCACAGGCCAATTAAGG	TCAAAACTGTGACTGGCCCT
222833	4	106158311	106158333	106158473	106158500	CTGCAGAACTTGATAGCCACAC	CTGCAAGATGGGAAATCATATTGAGTC
222834	4	106158501	106158525	106158665	106158690	ATGTGTAGGTAAGTGCCAGAAATG	TCTCTTTCACAAGACACAAGCATCG
130126	4	106162340	106162370	106162586	106162609	AGATCAGTCCATCAATCTACTCATTTTAAA	CCCTGTGCCTTTGCGTTAATTAC
133942	4	106163861	106163889	106164086	106164111	AATAATAAACCGTTCATTTCTCAGGATG	CCAAAGGCTTTATCAAGTCACACTT
113752	4	106164694	106164714	106164916	106164937	GGGGTGTTTGGGATGGAATG	ACTCTTCATTCAAGGCACACC
198161	4	106180686	106180718	106180925	106180954	ACTCATTTTGCATATAGACACCTATAATATCA	GCGATTATACATCAGGAAGTAAACAAACC
115267	4	106182737	106182770	106182981	106183006	TACTAGCACATATGAAATTAAATGATAGTCATG	CCTGATTATTATATGCATCAGGTGC
137518	4	106190675	106190708	106190877	106190898	TTTAAAGTTCTAAATGGTCTAAATACTAGTGAAG	GCTGCCATTCTGCATGTTGTG
130132	4	106193686	106193714	106193931	106193955	GTAGTTGAGGCTGTAATGTCTTACTTCC	TCTTGGCTTCTAGTTTCCTTTGTC
113876	4	106193962	106193983	106194128	106194153	CAGCTGAAAAGCTTTCCTCCC	GGGGCAAAACCAAAATAATTTTCAT
130134	4	106196066	106196091	106196309	106196330	TGCTCTTATCTTTGCTTAATGGGTG	CTGTCTGAGGGTGATGTGGCT
121317	4	106196395	106196422	106196644	106196664	TAGTCCTTATCCAAACTCTTCACACAC	GGGTCTTGGCTTGGATACCT
130136	4	106196734	106196764	106196952	106196975	TACATCTAAATACTTAGGTTATGGAAACCA	GAGCTGCACTGTAGTTATGGATT
130137	4	106196976	106196996	106197185	106197206	CGGGCATGTTCAACAGCTCT	AGACCTCATCGTTGTCCTCTG
222847	4	106197218	106197240	106197444	106197469	GAGCAGAGCTTTCTGGATCCTG	CTTTTCACACTCTTCCTCTTTCTCA
222848	4	106197470	106197490	106197682	106197706	TATGGCCCAGACTATGTGCC	AAGTGAGGTAACCAACAAAAGGGG
130140	5	170837363	170837396	170837595	170837627	AAATTACATCTGAGTATAAATTTTCTTGGAGTC	CTGTTACAGAAATGAAATAAGACGGAAAATTT
222850	5	176934677	176934697	176934922	176934942	CCCCAGGCTCACTTCCTTTT	GTTGTGAACAGAGGGAGTCC
222851	5	176936136	176936158	176936308	176936328	GCTCAGCCCCTCACCCTATTTA	GAATGGAGTGCGGGTGTCAG
222852	5	176936350	176936371	176936596	176936616	ATCCTCCACCCTGACTCCCTT	TGTGAGGAGTTCCCGTCCAG
222853	5	176939280	176939300	176939530	176939549	GGGACCCAGGGAACAGCTAA	TGCCATCCAGCACGTCATC
195719	5	176940146	176940164	176940384	176940403	ATGGCCCGTCATCTGGAC	TGACCATCAATGTGGGGCG
232466	5	176942746	176942769	176942989	176943010	CCAGGATGTGGTATTTCTTCCGC	AGCCATGTCTCTCTCTTCAGC
222855	5	176948486	176948506	176948735	176948755	CCACGATGTTCACTCAGGGC	CTATTACTCCTCCCTGGGCC
137612	7	140448975	140449001	140449194	140449218	AAAATCAAGAAATCTATCCTTCACGC	GCACCAGAAGTCATCAGAATGCAA
115641	7	140452951	140452979	140453193	140453220	TGATTTTTGTGAATACTGGGAACTATGA	TGTTTTCCTTTACTTACTACACCTCAG
121337	7	140476580	140476607	140476829	140476849	CATTAGTTAGCATCCTTATGTTCCTGG	TTCCACAAAGCCACAACTGG

117899	7	140477615	140477646	140477861	140477883	TTGCATACTACTTAAAAGAATGTGGTTAAAG	CCTTTTAGGTGATGTGGCAGTG
121339	7	140481174	140481194	140481419	140481443	ATGGTAGGAGTCCCGACTGC	GCAGATTACAGTGGGACAAAGAAT
121340	7	140482699	140482731	140482929	140482948	TATTTAATTATGGGAATTCTGTGTCACATATG	CAGGTTTGTCTGCTACCCC
137618	7	140487130	140487159	140487368	140487399	CCTTTTATGTCCAGTCATCTTTTCTACTC	ACTATTTATTTGTAGGACTTGATTAGAGACC
121342	7	140500014	140500047	140500261	140500283	GTAAATATCTGAGTGGTATGATAAGTTATTTGG	AGTTTGCTGTTTGTCTCCAAGT
222864	7	140501161	140501194	140501389	140501419	AATTTAAGTGTAAAATGGTAGGTAGAAAAGAGA	CAGTTTCTAGAAAGTTTTCTTGTGAGTTTT
137621	7	140534372	140534401	140534616	140534641	AAGAAACAGCAAAATGGTGATATTAAAAC	CACTCCAACAAAGAGAACAACAGTT
137622	7	140549735	140549764	140549972	140549998	TTAACAAGTGACACAAAACAAGATTAAGA	ACAAATGATTAAGTTGACACAGGAAC
130141	7	148504522	148504549	148504766	148504788	GCATTATTGCAAAAATTCACTGGTACA	GCTGATGCCCTGAAGTATGTCG
222868	7	148505936	148505961	148506186	148506205	GGGCTTTCTGCATGGATTTTACTCT	AGAGAGCCATCCAGACTGG
112729	7	148506244	148506267	148506491	148506513	ATAACTGCAAAGAGACACACTGG	CACTGGGCTGTGCTTACTTTTT
113193	7	148507295	148507316	148507535	148507559	CCCCATGCCTCTAAGGAGATC	CAGGCTTGATCACCTTTATCCAAA
113766	7	148508540	148508564	148508788	148508809	GCACAATCCAGTTACTAAGCATGC	CTATTGCTGGCACCATCTGAC
137629	7	148510962	148510987	148511212	148511231	CCCCAGCTAAATCATCTAAGGCAAT	AGGTCAAAACCGCTTTCCG
137630	7	148511851	148511871	148512065	148512084	GGCCAGGTATGGGGCTCAAT	CGGCAGCCTTGTGACAGTT
121353	7	148512377	148512399	148512626	148512646	CTTCAACTGACTCGGACTCACA	TTTTGCAGGTTGTGGGCTGC
137632	7	148513553	148513586	148513791	148513815	TTTATGACTCTTAACATACCAAATATACTGAAG	ATGTGGATACTCCTCCAAGGAAAA
121355	7	148514226	148514252	148514467	148514492	ACCAAGAATTTTCTTTGTTTGGACAA	GTTTTGTAGAAGCAAATTCTCGGTG
121356	7	148514887	148514918	148515137	148515156	TATTCAATGCATTATACATCCTTAATCCTCA	ACCAAAACGTCCAGGAGGC
117902	7	148516501	148516528	148516722	148516746	CATTAATGAAGGACTAAAATGTGCAGT	AGAACACAGAAACAGCTCTAGACA
121358	7	148523532	148523561	148523772	148523801	AAATGAAAACGTACAATAATTGCACTTAC	GGAAACCTTTTAGAAACTGTTTTCAAAAA
121359	7	148524108	148524134	148524347	148524377	CAAGTACACATGTTGCTTCAAGTATC	TTCATATTCTCCTGTTTAGATAAAGAAAGC
137638	7	148525664	148525687	148525904	148525931	GCCACCCTACCTGGCCATAATAT	AATGCCCTTGGTCAATATAATGATGAT
222882	7	148526679	148526702	148526918	148526946	GGCCCAGGTTCAGTCCCTTATAG	TTTTAGGTGGAAGATGAAACTGTTTTAC
222883	7	148529540	148529565	148529784	148529809	AAGAGTAATACTGCACAGGCCTTAA	CAAGTCATCCCATTAAAGACTCTGA
137641	7	148543401	148543427	148543643	148543670	GTGATCTACAGCAGTCATTAACAGTT	TCAGAAAATTTTGGAAAGAACGGAAAT
130154	7	148544197	148544223	148544407	148544440	AAAAACTTATTGAACTTAGGAGGGGA	GCTGGTTAGATTAGTGATTTTAATATGAAACCA
121365	8	117863977	117864000	117864217	117864238	AAAAGAGTCAGGAAAGACTGCAC	TCAGGGAGTTAAGCGAAAAGC
137644	8	117864772	117864796	117864992	117865024	TTCCAACAGAACAAACCGATAACA	GAAGATAGAAATCAGTGGTGACTTTTTAATTA
222888	8	117868792	117868824	117869003	117869033	AAACTACAACTTTTAGT	CTAACTTACATTCC

						TTGTCATCTTCTATG	CTATTTCAGATGACAA
222889	8	117869492	117869519	117869728	117869761	TCATTCCCACATACCTAATATTCCACC	GTGCTATTTCTAATCTTTCTAAACTAATGACAA
113199	8	128750505	128750524	128750715	128750733	GACGATGCCCCTCAACGTT	AACGTAGGAGGGCGAGCA
121370	8	128750823	128750843	128751059	128751077	GCTGGGAGGAGACATGGTGA	TCGATGCACTCTGAGGCG
121371	8	128751085	128751105	128751254	128751274	GTGGTCTTCCCCTACCCTCT	TTCGCTTACCAGAGTCGCTG
121372	8	128752577	128752610	128752821	128752842	ATCTGGTAATTGATTATTTTAATGTAACCTTGC	GATAGTCCTTCCGAGTGGAGG
55727	8	128752843	128752863	128753016	128753035	CTGCTGCCAAGAGGGTCAAG	AACTCCGGGATCTGGTCAC
121374	9	5069814	5069846	5070049	5070080	GAGTTATTAAGCATTTCTTATACGTAGAACAC	AGACAGTAATGAGTATCTAATGACTTACAAA
121375	9	5073530	5073549	5073773	5073799	TGATGGCAGTTGCAGGTCC	GTAGTTTTACTTACTCTCGTCTCCAC
222897	10	27276651	27276684	27276900	27276920	AAGCAAATCTTCATAGAAAGTTAATCTATTGAC	AAGTGGAGAAACGGCACCTG
222898	10	27279712	27279744	27279927	27279948	CATACCAAATTAAAAACCTTAACAAAACAAGC	GGAAGTCGAGCTTGTCTGCTT
222899	10	27284890	27284919	27285127	27285157	TATATCATGTATATTGCCATTCTCTCCAG	GATATTCTCTCTTCTTGATCTCCCAAAGTG
222900	10	27286683	27286708	27286928	27286952	ATGTGTTTCGGAGAAAATGCTAAAC	GCAATTTGTTGGGGTAATCAGACC
222901	10	27287313	27287334	27287563	27287582	TTTGAGCCTGTCTTCTGGACT	TCAAAGGCTGAGCTGCGAA
222902	10	27288201	27288225	27288449	27288470	GACAATTCGCCATAAGCATTTTGC	CCACATCTTAAGCCCCTGACA
222903	10	27288840	27288871	27289088	27289109	AGAACTCGAATTGTAAGTAGAAATTAAGACC	TCTATCTGGGTCCTCTGCCCA
222904	10	27289370	27289397	27289570	27289603	CCAGGAATTGGAAAGGATACAATACTG	GATTTATTCCCAGTTATTTGATACTTTTGATTC
222905	10	27289945	27289974	27290164	27290184	ACAAACAAACAAACAAAGACAGAATAAC	ACTCCAGCCCCTAACTCTGG
222906	10	27290520	27290543	27290712	27290739	TGCATAGTGCAGAAAATCTGGCC	CTGACATAGATAAGCAAGCTGAAACCT
222907	10	27291759	27291787	27292002	27292028	TGCAACATTTATTCCTTGATTTAACACA	ATGACATATAACAGAACCAATTTGGC
222908	10	27293144	27293173	27293366	27293391	ATGTTAACATTTGGGGAATAATATCCAAA	GGGAGGTAATTTGAGACTATGTACA
222909	10	27297417	27297450	27297656	27297681	TGACTTGTATTAAGGAACTACTAATGTAATGTA	ACCCTTTAAGTGTTAGGCGAGAAAA
222910	10	27298986	27299005	27299211	27299244	GAGCCCCTACCTGTCCTTT	AGGTGTTTATTTTTAGTATTTCTAGATAAGCTG
222911	10	27299380	27299413	27299615	27299641	TTTTATATTTGCATCATGCTTAGTAAGTTACAG	AACATAGTGAGGCATAGTGAGATCTC
232524	10	27303450	27303483	27303699	27303719	CTAACTTTTAATTCTTCTAGATAGAGTTGCTTA	GCCCTTAGACACCGAGTAGG
232525	10	27303732	27303758	27303899	27303921	AACCCAAACTCCAAAAAGAGAAAGAT	TGGGTTGCCAGTCAGCCTTTAT
222913	10	27304365	27304390	27304609	27304634	TTTAACCTTTCCCTAGGTCTCTTCA	GTTGTGAGAGTAGAAAGTCTTCACT
222914	10	27308735	27308763	27308942	27308961	TTTTTCTGCTTATTGCTTCCTAACAAAG	TCCTTGCCGTTGGAACCCA
222915	10	27309324	27309349	27309568	27309593	CTAGCTCTAATTTGGTAGCCATCAC	AGATGTGGTGTATATACAGGATGGA
222916	10	27309878	27309898	27310046	27310067	CCGTACCCAGCCACCAAATC	AGACGCTGAGGTTTCCAAGAA
222917	10	27310091	27310121	27310268	27310296	GTATACACCATATTTTC	AATGTGGAGTTTC

						TTTTCCATTGTCT	TTCAGAAAACTAAAA
222918	10	27310929	27310954	27311177	27311197	TACTCTGTTGATACTTCCTTTGCTC	TCCTCTAGGAGGCCCCATTG
222919	10	27312417	27312447	27312659	27312686	TTCAAAGTAGATAAGAGACAGGAAAGAAAC	GCTTTTCTGTTTATTTGCTGAGTACTT
222920	10	27315013	27315037	27315249	27315282	TGTTTGCCTTTGCTTTCTTTTGAC	CAACATCTCAGAAATGGAAAATTTAGTAAAAGT
222921	10	27315338	27315369	27315584	27315607	GGAGACTTCATATCAATATTGTCATGTCA	AACAGGATTCTCTGGTCATTCAC
222922	10	27316006	27316034	27316175	27316199	AGGAACTGTTAAGGTTTTGAAAGACTGA	TACAAGCATTCTTTTCCCCCAAAG
222923	10	27316787	27316812	27317026	27317056	GACTTTATAATGGCGTGAAACCATC	GATTACATCCTCATAAACACATTGTAAGTT
222924	10	27317534	27317562	27317693	27317724	AATTACATAAAAGTAACTTCAGGCTACA	GACAAGTTGAGTATCCCTTATTTAAAATGCT
222925	10	27317725	27317757	27317916	27317949	CTACTAAAGCAATAAAATTCCTTTTCAGTACG	GGTGTTTGCTAAGAAATTACACAGTATTTTAAA
222926	10	27318127	27318159	27318371	27318396	ATGATCATAGCTGATATAATTCTCATACTACA	ATATTGAAGGGGTATGACTGTGTTC
222927	10	27320010	27320030	27320235	27320260	GGGGAATTGTGTGCCAGGTC	TGGCTATACCAGTTATCCATTCATC
222928	10	27322063	27322090	27322303	27322332	AAGAAGGCTACCCACTAAATTAGTATG	AAATCAGTAACAAAATGTCTTATCAGGTT
222929	10	27322788	27322808	27323025	27323057	GTAACCTGGCCTCCACTTGC	CCCCAGTGGTATTTATAATTTACTTTGAATAT
222930	10	27324167	27324200	27324409	27324434	CTTTCAGTTTTAAATAGTTGTTGAGAAAGAATC	CAGCTGAGAATGCAATGCTAAATTC
222931	10	27325201	27325226	27325447	27325470	ATCTAGGCATTGTACTAAGCACTTC	TGCGACAACAAATAATCTCGTGA
222932	10	27326800	27326829	27327043	27327067	CTGATTCAAATTACTTTTTACAGTCCTCA	ACCAACATACAAAGCGGCATATCA
222933	10	27329424	27329444	27329660	27329681	CTGAGGGAGAGGTGGCTACA	CGGCCACTTCCTTTGTTTCTT
222934	10	27330868	27330901	27331103	27331123	ACCATATATGATGTGGTATGATATTATTTGAAG	ACTGCATTTGCCATTGTGCC
222935	10	27331850	27331877	27332038	27332067	CCAGCCTGATTGACAGAGTGAAATTTT	TTAATGGTATCTTTTGAAGAGCAGAAGTT
222936	10	27332774	27332797	27333012	27333039	GCAGTATCGTGACCATAGCTCTC	TTATAAAAGATAAAGTGGGCTGGGCAC
222937	10	27334804	27334837	27335046	27335070	AACTACATTAATAGAAGTTCTAAGGACAATTAG	GTTTGAGAGAACTTGGACTGAGAG
222938	10	27335072	27335092	27335241	27335271	CCGTCTGTGCCCATCCACTA	TGATTTCTTTATTTTTCTATGGAGGAAGGT
222939	10	27337094	27337125	27337315	27337343	AGTATACTGAAATGTTTACATGTACTACAGA	CACACACACAAGTTAAGGACATTTATTA
222940	10	27338607	27338636	27338851	27338876	ACCTCATTATCTACTGATTAAATGATCCA	CCCGGCCAATATTCTTCTTTAAAAG
222941	10	27339266	27339296	27339509	27339530	AAATACTAAATTTCTGGCTGAGGATTTATG	TCCCAGCCAAAGGGATCTTTT
222942	10	27342549	27342582	27342785	27342817	AAGAAATACTTCTAAGTATCAACTTAATGAGGT	CCTTGTTAAAAAGATACTGTTATAGAGAACAT
222943	10	27343600	27343621	27343844	27343869	GGTAAAGGGCAGGAAAGGGTA	TCCAATGCAAATAGGACCTTGAGTC
222944	10	27344447	27344469	27344697	27344716	GATCACATGATTAAGGGCACCG	ATAGAGTGGACTGCCCCGT
222945	10	27348204	27348237	27348430	27348452	AAAATGGAATACTATA	GAGGTAACCAACA

						AAGAACTTAAACCAGAG	CGCTGAATT
222946	10	27349147	27349176	27349316	27349345	TTTTTAAATCCTTCCAACAAATTTGGAGA	CTACCATTGAAATGAAAGATTCTGTTCCA
222947	10	27349507	27349539	27349725	27349753	AATATGACAAAGTCAATCATAAATTCATGCAG	CACATTGGTATATGTGCACTTTTCTATT
222948	10	27349822	27349847	27350072	27350091	AATGTCAAGGCTGATGGTAAAATGC	ATGCCAGTAGCCCACATGG
222949	10	27350547	27350570	27350753	27350774	GGCAAAATCATGGCTCATTGCTC	TTGCCTGTCACTCAGCACTTT
222950	10	27352817	27352846	27353053	27353086	TAATAGGTGATTTTCTTCGGCTATTAGAA	GAAAGTGATTAGAAAATCTTTAGCAATTCATTC
222951	10	27353115	27353144	27353293	27353320	AGCAATTAGGCTTAAAAATAAGATAAGCA	TGATATTAGTTCCTTGTTGAACTGTGC
222952	10	27353439	27353467	27353679	27353708	AATACACTTGCTATAGTACAGCGTTAAA	TGCTTTTCTTTAGACATAACATAAAGCAA
222953	10	27354082	27354111	27354324	27354351	CCATTTTGATCCCCATCCTAACTACAAAC	AGGAATATAGTGAATTGACGCTTGAGA
222954	10	27356712	27356736	27356958	27356979	AATATGAGATTTGGGGCAGGACAC	AGCTTCCCAGGTGGAACTTAC
222955	10	27356980	27357004	27357187	27357212	CAGGAAATGATGCAAAGATAGGGC	GTATGAGGAACTACCAAACACTTGT
222956	10	27358941	27358961	27359178	27359204	GAAGCTGAAGCCAGACCCAG	AGTGTGTACTACCAAACATGGCTAAT
222957	10	27359334	27359363	27359535	27359562	ACAAAAGAAAAGAAAGCTATGATTATCCA	CCAGCAATGAGGACATTTTATATTCTC
222958	10	27360661	27360681	27360910	27360930	GCCTCTAGAGTGTGAACGGG	TCTCCACCTGTGATGGCAGC
222959	10	27364577	27364601	27364819	27364846	TAGTACAGGACTTCAACACTCCAC	TGCTGTGGTCTGAGAAGATACTTCATA
222960	10	27366873	27366906	27367100	27367125	TACTATGATATCTGTTCGATTATTTTCTTAACC	TGACAGAATAGAAGCAGCCTATACT
222961	10	27367137	27367170	27367379	27367404	ATAAATGATTGTATAAATTCAGCTCCTTACAAG	ACTCGAATGGTCATGCCTCTATTTT
222962	10	27367405	27367432	27367588	27367618	GTGAGCTTAAGTCAAGGAAGAAAAGGA	AGCTTTCAGGTAGTAGTAGATGATCATTTT
222963	10	27369456	27369481	27369691	27369713	AAATCTTTACAGGCACTCAGATACC	GCTTGTTTTCCAGGAGCAGAGA
222964	10	27369741	27369762	27369987	27370010	AGCACAGGCATGTTAACAACT	AGTCTCTACAGCCAGTAGGAATG
222965	10	27371215	27371241	27371463	27371484	GTCTGCTTATGTAAAATGTATGGCCC	AATCCCTTCTCTTGCCCCTCA
222966	10	27375255	27375282	27375427	27375458	TTTTTCAGATAGTACCATTTGGCATCA	CCAATAGTGAATTTCTCTTGGTAATAGTACC
222967	10	27375459	27375490	27375631	27375660	AAGTCTTACCTGAATTACTATTTTGAGAAGA	GGGATAATTAAAATTCTGCTGTTAGCTCT
222968	10	27375661	27375691	27375842	27375862	ATACACTTCAAATATGTAAGCTCTACAAAC	TGAGCGTAGTGACGGGTACC
222969	10	27376372	27376400	27376619	27376641	CCAAAGGTTCTAATATTCAAATGTTGCC	AGTTTGAGGAGGTAGAGAAGCA
222970	10	27377154	27377186	27377372	27377398	TTTTGTCATGAAAATCACAATTTCAATACTGG	GTTCCATGACTCATCAGTATCCATGT
222971	10	27378150	27378176	27378394	27378419	TGTCCCCTCCTGATTTATAAAAGAGG	CTCAGGACTTTTACGAACTTAATCC
222972	10	27378425	27378449	27378628	27378649	AAAATTCTATACCCGGTTGGACAC	CCGTGGGCTAAGGAGATTCTC
222973	10	27378828	27378856	27379044	27379069	AAATTACCTGACCCAAACTATGAAATCT	CCCTCATGCCCACTATATTTTATGA
222974	10	27382344	27382368	27382588	27382613	ATAGACAGCATAGTGAAGAGCAGT	GTATATAGTAGCCAGCTCTTTCAGC

222975	10	27382619	27382640	27382864	27382888	TCAGAGCTGTCCTGTTTTCGT	TACAATTTGGGATTTGGTTGGGGA
222976	10	27383358	27383385	27383592	27383625	AACTTGGGGTTGAATCCTACTTTAAGC	CTGTCAACTAGGTATAATTGTATCATTTCATTT
222977	10	27384669	27384690	27384915	27384938	TCTCAACTTGCTCAGGCTGGG	GTCAAGTATGAGAGCTGTTGGTG
222978	10	27386426	27386453	27386674	27386695	GCTACATAACAGGTATTCAGGTTATGA	CAGGCAGTAGAATAGCGTGCA
222979	10	27387086	27387111	27387326	27387355	GCAAGAACAAAGATCTGCTTTACCT	AGAGGAATATGAATCTTGCACAGATAAAA
222980	10	27389545	27389564	27389776	27389800	AGCCCTGCGTTCCTGTCAA	TGTTAAGTTCTCATGGAGACCAGA
222981	10	27390910	27390930	27391075	27391099	CTTGGATGGAGACCACGACT	CTACAACTTACCTGGAGAGTGGTG
222982	10	27392491	27392512	27392737	27392760	TCTGCCAGGTTTCAATCAGGA	AATCTCTGCTGGGATTACAGGTG
117766	11	32413366	32413392	32413613	32413635	ATTTAAAGATAGCCACGCACTATTCC	GGCTAGACCTTCTCTGTCCATT
130157	11	32414016	32414041	32414263	32414285	AATGTGGGGTGTTTCCTTTTCTTTC	GTGTGACTTCAAGGACTGTGAA
117768	11	32417725	32417746	32417953	32417974	GGTCCTTAGCAGTGTGAGAGC	CCTCTTACTCTCTGCCTGCAG
124469	11	32421306	32421327	32421542	32421561	CTCTCTAGGAGAGGACACAGC	ACACAACGCCCATCCTCTG
117771	11	32438969	32438990	32439209	32439238	CCAACTAGGGGAAGGAGGAAA	TGGTTATGTGTTTCTAACTCTAGATGTTT
137665	11	118343131	118343152	118343381	118343400	GCGGATCATTAAGACCCCTCG	ACTTCAGGGGTATCGCTCC
137666	11	118343758	118343780	118343997	118344027	GCCACAATACTTTTCCTCAGCA	GGCAAGGTTACAGACTCAAATATTCTAGAG
121434	11	118344032	118344055	118344280	118344301	AATCGAACTTCTGCTGGAACATC	AGATTCCCCAGACTGAGTCAG
137668	11	118344359	118344384	118344593	118344614	TCAAGTAGTCCTACTCCTCTCTTCC	CTTTGGAAACCCTACCCACAG
222992	11	118352344	118352373	118352550	118352572	CTGGATCATATGGTGGCTCTGTAATTCTA	ACTACTGCTTTTCTTTGGGGCA
137670	11	118354812	118354841	118354997	118355018	AATTTGTCATTTGCATTATTATCTGTTGC	CTCTGATCCTGTGGACTCCAT
137671	11	118355395	118355420	118355644	118355664	GAGCCTTTTAATAGTCCGTGTCTGA	GAAAGCAAACCACCCTGGGT
137672	11	118373676	118373703	118373919	118373945	CCCTCTTCAGTGTCGTTTTCTTCTAAA	ACTGGAATGCTTTTCTTTGAAAGTTT
121440	11	118373962	118373982	118374206	118374231	CTGGTCAGGTGACAACTGGT	GGGCATTTTTGGATTGACTCTCATT
121441	11	118374232	118374256	118374411	118374431	TGAAAGAAAGTAGTCCTGCTTCCC	TTTGGGGCCATCTGGAAGCA
121442	11	118374670	118374695	118374913	118374936	CTAGAACAGTGATTTCTTCAGGTGG	GGCCAACAGAACTCTTAGTGACA
121443	11	118374941	118374966	118375125	118375150	AAATGAGCCAAAGATGGATAACTGC	AGGATATTCCCACAGTCATCACTGT
121444	11	118375375	118375397	118375621	118375642	AGTTTGAGTTGCCTCTAGAGCT	CAAGGAGGGCTAGAGATGTGG
223001	11	118375643	118375664	118375867	118375891	TGGTTCAGTAGAGCAAGGTCA	TGGAGAACATAAAGTGGCTGCATG
121446	11	118376077	118376098	118376322	118376346	TACCCATGTCTCATCACCAGC	CAATGTTTGAAGGGGTACTAGAGG
121447	11	118376959	118376982	118377208	118377228	AACACTATCAGCTTCAGCATGTG	GTTTGGGTTTAGTGGGACCC
137681	11	119148329	119148362	119148514	119148535	TTTTGAAGTAAGATTGATCTTTATACTTACACC	TCTTGGGGAGTTGGTTCACAT
121449	11	119148751	119148779	119148999	119149019	CCACTGTTGTGACATTTTTATATAAGCA	TTAGATCCGTACCTGCCAGG
121450	11	119149043	119149074	119149290	119149312	AATAACCTTGGAAAATTCGGTATTATATAGC	GCCACTCCCTCTAGGATCAAAC

137684	12	11802933	11802953	11803100	11803123	CCGGGAGAGATGCTG GAAGA	CACGTAGAAGGA GGGGAGAAGAT
121452	12	11905300	11905320	11905513	11905532	CCACCCCGAGATGGTC TCAT	TCGGGTCACGAAC ACTCAC
121453	12	11991980	11992007	11992223	11992246	TTGTTAACTATTCAGAG ACCTTCTCCC	ACTCTCACCTGAA TGAGGAGATC
137687	12	12006217	12006239	12006464	12006486	CATGAGCTTGGTGAGG GTAGGA	GGTTCTGATGCAG TATGACCTC
223011	12	12022185	12022214	12022377	12022398	AAAAATATGCACCCTT CAAATTGTTAAGC	ATTATCCACGGAT GGCCTGGG
121456	12	12022445	12022467	12022682	12022700	CACCTATCACGACAAA TCACCG	ATCAGCTGGATCA CGCGT
137690	12	12022707	12022727	12022878	12022900	CCCCATCATGCACCCTC TGA	CTATTCTCCCAATG GGCATGGC
121458	12	12037272	12037292	12037518	12037541	AACCCAAGCTAGGCAG AAGC	GATATCTGCTGCC CTTTTACCTT
113093	12	12038758	12038786	12038972	12038997	GCTGAAGAGCTTTTAT TTTAATAGCTCC	CTCCCAAAAGAAG GAAAGGAGAAAA
137693	12	12043675	12043702	12043919	12043940	AGCATTTTACCACTTTT TAGACAGTCA	TGGGACTCTAGGT GCTCCAGA
137694	12	25378493	25378522	25378720	25378749	ATAACAGTTATGATTTT GCAGAAAACAGA	GTGTTACTAATGA CTGTGCTATAACT TTT
130162	12	25380073	25380106	25380285	25380309	ATTAAATATTATATGCA TGGCATTAGCAAAGAC	CCTGTCTCTTGGA TATTCTCGACA
112650	12	25398047	25398077	25398285	25398313	TAATGGTTACATATAA CTTGAAACCCAAGG	TGAATATAAACTT GTGGTAGTTGGA GCT
124474	12	112887982	112888011	112888221	112888243	TTCTTATAGGGAATAG GTAAATTCGTTCC	GCCCGTGATGTTC CATGTAATA
124475	12	112890836	112890859	112891085	112891105	GGATGCAGATTTTCTG TCTCAGG	AATCTCCAGGGTG GCTCTGG
130166	12	112910605	112910632	112910844	112910874	CAGATGAACATTCTTG TAGCTATCGCA	ACTATTACTGAGC GGAATATTGATAC TTAC
124477	12	112915286	112915307	112915533	112915555	AGGCAGTGTTCACGTT ACTGT	ACTGTGAAAAGCA AAGCTTACC
55057	12	112924141	112924163	112924377	112924398	CTCACGAAGAGGACCT TTCAGT	GCTCTCCTGCTTAT GGTGCAC
121466	12	112926033	112926054	112926270	112926302	ACGTTGTACTGGAGAA GCTGA	TGATGTCAATAAG AATATCAATCACA ATGAAC
137700	12	112926670	112926694	112926917	112926939	TTGGTCAGAAGTTCAA CACTGTAG	TGCTGGACCGCCA TATAGATAA
198211	13	28589567	28589590	28589814	28589836	GTGAGGAAGACAGGC TAAAGGAC	CCGTCTGCCTGTA AAATGGATG
130172	13	28592473	28592495	28592704	28592724	GCAGACTGCTGTGAGG GTTTTT	TGTTCACAGAGAC CTGGCCG
124483	13	28597315	28597339	28597562	28597584	ATGCTACTACAATTAG CCAAGAGC	TGGAAGAAGAGG AGGACTTGAA
124484	13	28601016	28601042	28601257	28601285	CTCAGCTATAGTACCT GTACTGAAGG	AGATTTTCAAGGA ACACAATTTCAGT TT
113272	13	28602165	28602189	28602411	28602434	TGTCTAATTCCACTTGG GTTTGAG	CTTTGACAGAAAA AGCAGACAGC
223032	13	28608064	28608089	28608292	28608313	TTTGCTAATTCCATAAG CTGTTGCG	GACCGGCTCCTCA GATAATGA
223033	13	28608314	28608335	28608500	28608525	ACCTGTACCATCTGTA GCTGG	AGACAACATCTCA TTCTATGCAACA
121472	13	28609516	28609539	28609760	28609785	CTCACACACTGACCCTA TACTCT	GTCTGGAATAGAA AGGCTAACAGAA
124489	13	28609911	28609941	28610160	28610180	GTTATCAGAAACAGTC TATGACTATTGAGA	GGAAACCTCAAGT GCTCGCA
124743	15	90631705	90631725	90631935	90631960	CCGGTCTGCCACAAAG TCTG	GTGGAAAAGTCCC AATGGAACTATC
137750	17	7572743	7572764	7572991	7573012	ACAACAAAACACCAGT GCAGG	ACAGCCACCTGAA GTCCAAAA
113790	17	7573800	7573820	7574037	7574057	GGGCTGGGACCCAATG AGAT	TCTCCCCCTCCTCT GTTGCT

113791	17	7576458	7576480	7576628	7576654	AAAAAGCAGGCTAGGCTAAGCT	CAGACCAGCTTTCAAAAAGAAAATTG
223040	17	7576666	7576699	7576897	7576917	TATATATTATGGTATAAGTTGGTGTTCTGAAGT	ACAACACCAGCTCCTCTCCC
113018	17	7576923	7576944	7577153	7577179	AGTGCTAGGAAAGAGGCAAGG	CTTCTCTTTTCCTATCCTGAGTAGTG
4933	17	7577369	7577389	7577609	7577631	CCGGGGATGTGATGAGAGGT	CTTGGGCCTGTGTTATCTCCTA
124310	17	7578045	7578066	7578291	7578314	ACTTTGCACATCTCATGGGGT	CTGATTCCTCACTGATTGCTCTT
55074	17	7578333	7578353	7578556	7578579	CAACCAGCCCTGTCGTCTCT	CTCTGTCTCCTTCCTCTTCCTAC
121481	17	7579336	7579355	7579585	7579605	CCCAGAATGCAAGAAGCCC	TTCACCCATCTACAGTCCCC
591	17	7579680	7579700	7579929	7579949	AGCCCAACCCTTGTCCTTAC	CTCTTGCAGCAGCCAGACTG
137761	17	29482905	29482932	29483148	29483173	TTCCTAAAACGTCATGATTTTCAATGG	CCCCAAAACACAGTAACCCAAATAC
120208	17	29508599	29508627	29508842	29508868	GATAGTTTCACATTCATTTTCAGGAAGA	GGTTTTTATGTCACAAGTAGGCATTT
223049	17	29509359	29509384	29509596	29509625	ATTGCTTGTCTACTTACCAGAATGC	CACAAGATAAGGAGAATGATTTGTAGTGG
137764	17	29527911	29527932	29528148	29528169	CCACAGTATGGGTGCTTTGTG	GTTGGTTGTTGTGAGGGCTTA
121487	17	29528331	29528364	29528569	29528596	TTTTTAAACTTTCTATTTGCTGTTCTTTTTGGC	TGAAAACCAAGAGTGCATTTCTTAAAA
113320	17	29533143	29533165	29533385	29533412	TGGTCTTAGAAAGTTCCCGACA	TCAAAATTCGTGATTTATCTTACCGGT
121489	17	29551982	29552005	29552231	29552251	CAAGTTGGGGCATAGAGATTGAG	TTCCCTTCCGGAGAGAGGCT
223054	17	29553481	29553499	29553723	29553747	CCCCGATTTGCCGACAAG	TGCTTTGAGGCAGACTGAGTAAAA
137769	17	29554035	29554065	29554284	29554304	AATAGCTATTTTATTCTGTGGACATTGTAG	TGTTTCCTGCAGTGGGATGC
223056	17	29554351	29554374	29554593	29554620	CCTCAATTTGGAAGCCTCTTGTT	CCATCTTCCATTTTGGCTTTTGGATAG
121493	17	29556004	29556029	29556248	29556273	AGGTTTAATTCATGCTTTGCACAAA	CACTGAAATCATAGAACCCTTACGT
121494	17	29556640	29556661	29556879	29556909	TTGCCTCTTAGGAACTCTGGT	CTTCATTATAGCTATGGTTTGTTCTACAAA
223059	17	29557644	29557667	29557890	29557913	CCATAACTAAGGGCCATGATGGA	GCTTGGTTTGATGTTCCCATAAC
121496	17	29559829	29559848	29560076	29560098	GCACTGTACGGTCCTTGCA	TCTGTGCCTTGTTGAAGGATTT
113324	17	29562498	29562525	29562747	29562767	ATCAGTCATCATTTGCCTTAATTTAGC	TACTGGCCAAGCTGTTGCCT
121498	17	29562771	29562800	29563008	29563033	TAATGACATTCTGTTTCAAGGTTTGTATC	GGATCCACTTCAAAGCTAACATGTT
183214	17	29575873	29575902	29576118	29576142	TTTAAAGATTCCAATGAAGTCTACACGTT	CATACCTGGTATAAACAGTGGCAC
121500	17	29588506	29588528	29588751	29588775	AAGCTAAGGGTAAGCAATTGGG	GTGTTGCCATCTTATCAAAAGGTC
137779	17	29652697	29652726	29652912	29652936	AAAAATGAATCCAGACTTTGAAGAATTGT	AGGTCCACTACAATTTCATATGGC
223066	17	29654424	29654449	29654673	29654693	TTTGTTTGGTTGGTTGGTTTCTGGA	CTGGTGCATGAAGGTGAGCG
121503	17	29657135	29657164	29657375	29657404	GATTAAGGGGTATTTTGGTTTTACTGTAG	GATACATAAACCTGATGTCTCTAGTAACT
121504	17	29663498	29663526	29663748	29663767	ACTTTTCTTTTGCCTTCTGTACTATAGC	GCATCAGCATGTAGCGTGC
121505	17	29667283	29667308	29667528	29667549	AGAAATCAAAGAACAGGGAGAAGTC	AAAGAGGGCTTTGTGCAGAGG
121506	17	29683873	29683905	29684113	29684135	AGTATCTACTAAAGAAAGCTGTTGAATTTTAG	GGGGACTCAAAAAGGGGAAATC

223071	17	29684139	29684170	29684304	29684330	CCTCAAATTTTTATTCCAGTCTACTTTTAGG	ATTCGTTGATCAAACTCATCTGTGGT
121508	17	29685392	29685417	29685632	29685661	TAATTTTGGCACATTATTCTGGGGA	TCTCCAGATACATTTATTTACTTTGCAGG
137788	17	58740204	58740232	58740440	58740465	TAGTTCAGTAAATAGAAGGTGAACGAAT	GCTATCTCAGCTGAAACCTCTAAAA
121511	17	58740505	58740530	58740750	58740774	TCCAGAACCACTTGAAGAAAATTGC	CCACTACTTCGACTTAAGCCATTT
55076	17	74732172	74732194	74732410	74732430	AAAAGACCTACCCCAAATCCCA	CCTCCAAGTCCAGATCCGCA
121514	18	42530222	42530241	42530436	42530456	CTGAGTGCAACGGGCTTCA	CTGGCTGGTGATGCCTCTTG
121515	18	42531778	42531798	42532022	42532040	GGAACCTGGAAGCTGTCTCC	TTTGTGCTGGTGTCGGAC
110777	18	42532445	42532467	42532693	42532714	ACTATGCACCCTATGGAATGCC	CCTTGGGAGGGTTCAGAGAAG
137793	18	42532955	42532975	42533165	42533187	CCGGCCACAAAGCTTCTAAG	GTCACTGTCCCACTTGTCATTC
137794	18	42533188	42533210	42533355	42533377	GTGAGTGGGAGTAAAAGGAGGA	AGATGTGTCTGAGGTGCAAAGC
223076	19	4043156	4043176	4043396	4043425	GGTGTGGACGAGGTGCTCAA	GGGAGGGAGAAGCTTTAAAAGTAATAAAA
223077	19	4044580	4044600	4044767	4044788	GACCTCACTCCCGCTCTCTT	TGAAAGGACACAGCTCAAAGC
223078	19	4047174	4047196	4047420	4047438	GGAGACCCCTCCCTTAGTAAGT	AATCCGAGACCCCGCACA
232697	19	4048076	4048095	4048324	4048344	GGACGAAGGTCTTGCAGCA	AAGGGTGATGATCCAGGGCC
223079	19	4053967	4053986	4054163	4054189	ACCTGGTGAAGCGGACCTT	GTCATGGACTACTACCTGAAGTACTT
223080	19	4054910	4054929	4055157	4055177	GATGTCACCCACGTTGGCT	CGACATCCTGAGTGGGCTGA
223081	19	4059561	4059582	4059780	4059800	CTGCAGTCCTGGCCTAAGGAT	CCTTTCCTTAGGGGAGGCCT
223082	19	4063019	4063038	4063270	4063288	ACCCAAGCTCGTTCCTCTT	TTGGATGAGCCCCGTGAC
137795	19	13054383	13054405	13054628	13054652	CATCACCAACGATGAGGCATAC	CTCATCTTTGTCCTCATCATCCTC
223084	19	33790535	33790555	33790755	33790784	CCGGGGCAGATAAGAAGCCT	AGGTCACATTTGTAAATAATACAGCATTT
223085	19	33792251	33792270	33792437	33792455	CAGTTGCCCATGGCCTTGA	GCGAGCGCAACAACATCG
223086	19	33795002	33795029	33795245	33795269	TCAACATAGTCCCAGTGATTAATAGCC	CCATTAGTTTTAGCACCGCAGAAA
223087	19	33797186	33797207	33797436	33797455	CAAAGCCAGTCCTAAGAGCAG	TCTCGTGGGCCTTGCATTA
130201	20	30956629	30956659	30956873	30956898	TCTGAGTGTGATTTATGTGAATTTCATTTG	GGCAGTTTATAAAACAACCCCTCTC
124506	20	31015701	31015729	31015946	31015966	GAGGCACCCTATAAATGTTTAATGAATG	TGTAGCTGGATGGCGAGACC
124507	20	31021228	31021247	31021477	31021497	GAAACGCTCTCGGCCAGAT	CGTGCCAAGTTGTCTGGCTC
124508	20	31021498	31021518	31021671	31021690	TGCCTCTGCATCTCCAGACA	AGTGGGCTGTGGCTTTTCG
223092	20	31022029	31022051	31022277	31022298	GTGTGGCATATAACAGCCCTTG	GGCATATCTGGTAAGTGGGCT
121525	20	31022347	31022367	31022556	31022577	CCAGGACCCTCGCAGACATT	TACACTTTCCAGGGGTGCTCG
121526	20	31022689	31022710	31022937	31022958	GAAAGGAGGAAAGCTGCCTAC	GAATGGGACCATTGTCTGCAG
121527	20	31022959	31022984	31023132	31023158	TGTCTCTAGTGGGAGATGATACATT	TCTAGTATCACTTTCCCTCATAGGAG
121528	20	31023167	31023188	31023411	31023431	CTTGAAAACCAAGGCTCTCGT	GCCATTGCTGTCACTGCCTC
55833	20	31023499	31023518	31023732	31023753	ACGGTGAGTCCACGGATAC	GGGAATCTGGGATCTTTTGGC

121530	20	31023774	31023793	31024021	31024042	TACCAGCCAAGAGCCGTGT	GGCTCCTTCAAAGCCCTTAAA
183271	20	31024113	31024134	31024344	31024367	CCCCAAAAGAATTGCAAGGCA	TGGACCAAAGGAGATCACATTTG
121532	20	31024537	31024558	31024784	31024806	GCACAAACTCCATGTCTGGTG	GGAGAGAAGAAGGCTCCAGAGG
121533	20	31024807	31024828	31025035	31025055	CCTCCCAACTCAGCATCAAGC	TCATGGCTTTCAGGCTGCAC
137806	21	36164620	36164642	36164826	36164844	CCATGGAGAACTGGTAGGAGCC	ACCCCCGCATGCACTATC
124514	21	36171492	36171519	36171728	36171748	ACCATGTTTTACTCAATAATGTTCTGC	ATCCAACCATCCCCACCGTG
121537	21	36193742	36193767	36193990	36194010	TTAATCCTCACAATTGTCCCATGTG	ACACATGGCTCCCTCAGAGG
55086	21	36206682	36206702	36206892	36206913	GCAGTGGGCTCCATCTGGTA	TGCTCCCCACAATAGGACATC
124516	21	36231626	36231648	36231874	36231894	ATGCAACTTTTTGGCTTTACGG	CTCTATCGTGTCCCCACAGG
189368	21	36252782	36252810	36252996	36253017	TCACTAGAATTTTGAAATGTGGGTTTGT	TGTTTAGGTGGTGGCCCTAGG
121541	21	36259133	36259154	36259324	36259342	CCAGTACCTTGAAAGCGATGG	GCAAGATGAGCGAGGCGT
137813	21	36420930	36420959	36421175	36421198	TTTATTAAAACATTTCTGAAGAGCTTCCA	CGATGGCTTCAGACAGCATATTT
121544	21	44514609	44514628	44514845	44514870	GCCGCAGCTCTCTGGAAAT	CTGTGATTGACTTGAATAACCGTTG
223111	21	44524253	44524274	44524487	44524516	CTGGAAGTCGATCACCTGCCT	ACAGAGTCAACTGTTCATTTTATTTCAAA
121548	X	15808876	15808896	15809123	15809145	CTGTCTGGGAGTGGGAGGAT	GTCCATCACCTGAGTCTCTCAG
113960	X	15817775	15817808	15818015	15818040	TGATTAATTTTTAAATGTAATGTGCTATCTGCC	TGTTCTTCAATAAAAGTGTCCTCCT
137867	X	15821651	15821674	15821898	15821920	TGCTCTTCTCGTTCTCTTACACC	CCTCTTGTTCTTCTTGCCGTTT
223115	X	15822080	15822109	15822319	15822341	AAAAAGAGTTGCCTATCTTGAACTTAAAA	ACATCCCATGACAGGAATCACC
137869	X	15826169	15826199	15826402	15826431	AAAAGAGAAAAGTCTCAGCTATATATCCTT	TTGGTCTCAAATTATCAATGTAAGGTTTT
137870	X	15827250	15827283	15827433	15827465	TTAATATACTTTGAAACATTTCGTCTTTCATGG	ACATGATAAAACAAAATTAGTTACCTATCTCC
137871	X	15833718	15833740	15833945	15833971	CCATGCCTGGTCTAAAGCAGTT	CACATCCTCATAGAAGTCTAGGAACT
121555	X	15838171	15838191	15838419	15838440	GCTTTGGGAGCTTGACCATT	CCACAAATCGCCATTTTTCCAC
137873	X	15840703	15840730	15840948	15840972	ACATCAATTTATGTAAGCCCCTTTTAC	CCGATCTGGAGACAAGTAGATGTC
121557	X	15841060	15841081	15841309	15841329	AACCCTAGTCCAGACCACTCC	ATTACCCGACCTCCTCCTGC
223122	X	39911364	39911386	39911605	39911633	ACCAGTAGTTGTCTGAGGCCAG	TTTCGGATGTCCTTAAGAAATTGAAAAT
137876	X	39912910	39912933	39913153	39913179	AGTGATGAGGGCAGGTTTATACC	CTCTTACCGTGTTATAACATCCAAGT
223124	X	39913263	39913292	39913510	39913532	CTAAAACATCATAGCCACTTTCATCATCT	ACTTCTATGGCAGCTCTGTTTG
121561	X	39914509	39914534	39914756	39914777	AAGCTTTACAGTTTCAGCCTTTGTG	TCTCTTCCTAGGCCTCTGCAC
121562	X	39916294	39916315	39916535	39916563	GTCACCCTCTGTCTTTCCTCC	GTACTGCTTAGAGAACAAGATTTGTGAT
137880	X	39921256	39921274	39921504	39921524	ATCCTGTTGGAGCACCGG	ACAGCTGCCATGCTCAAGTT
121564	X	39922003	39922022	39922245	39922264	CCATCGGCATTCTCCACGT	GCAAAAAGCAGGCTCAGCC
121565	X	39922851	39922872	39923088	39923113	CCCCGCATACCTTGTTCATTG	GAATTAACAGGGCTCCATCCTAAAA

121566	X	39923114	39923134	39923293	39923313	AATGCACACCTTCAGGTTGG	AGAGACCAGTGACCACACCC
121567	X	39923467	39923487	39923709	39923727	CTTCCCCTCACCGACTCTGT	TCGGACATGCCTCACAGC
223132	X	39930701	39930730	39930940	39930968	TGAAGGTAAATAAAACAAAACGAAAACAC	TTGAGTGCCCATTTTAATTACATAGATG
137886	X	39931428	39931448	39931678	39931697	TCAGTGCTACCACTCACGTT	AACTCAGCGGGTTACGTGG
223134	X	39931698	39931718	39931909	39931929	GGCGATTCTCTTTGCCAGCT	GAACCTAGGGTTGCCAGTCT
121571	X	39932264	39932287	39932513	39932533	CATCTGGATGTAACTTGGTGCTG	GCAGTCTGTTTCCTGGGCAC
121572	X	39932548	39932568	39932798	39932817	ACTGGGTGAGGGTAGACAGG	TATTCAGCACGTGGGCCAG
223137	X	39932818	39932839	39933037	39933060	ACGGATGGTGTGGTTTCTACA	GCCAAGTCCTAACGAAGAGAACA
113976	X	39933093	39933123	39933339	39933361	GATGATTTCAGATCTATAGATAGCACAACC	CAGCGGTTCAAGACAGAAAAGA
121575	X	39933652	39933672	39933872	39933891	GGAACCCTGGGCTGCTTACT	AGTCTGCACCAATGGGGAG
121576	X	39934179	39934200	39934426	39934448	TTTTCCAGAGACGGCAGAAGC	TGTTCTTCTGCAAAGGTGGATG
121577	X	44820414	44820443	44820650	44820683	TTTTCAATATTGACTTCTTAGGTGATCGA	CCAACACATGTATAAATCAAACAATAATGAACA
223142	X	44833736	44833763	44833974	44833998	ACCATGGTTTTACGTTTCTGAACTGAG	AACACAAACATGAAATCAGCATGC
121579	X	44879826	44879858	44880035	44880068	TAAGATATCTTATGTCTATATTCTTTCAGGGC	CCAACATTTCTCAATGCTCTAATATTTAAGAAA
121580	X	44896671	44896698	44896915	44896940	ATAATGACCAACAATGAATACATGTGG	ACTTACCTGGGTTTCATATAAGTGG
138905	X	44910801	44910832	44911017	44911043	TGTTAAAGAAACATTCAATAATGGAATCAGC	CTGTTGTAAGACAGTTGCTTTTACTT
223146	X	44912917	44912947	44913148	44913170	GAATTGGCAGTTCTAAATTTTGGAAATATG	TTAGGATCTGCTTCCAAGGACT
138907	X	44918149	44918178	44918338	44918365	AAAAATTTGCTGTGATATATAGTCCATCC	TTATACAAAGGCTTACCCTATTGAACA
121584	X	44919133	44919160	44919382	44919402	GTCTTCTAAAGCCCCAAATTTAGAATC	CCTGCTGTGCTGTGTTCATG
138909	X	44921717	44921745	44921960	44921986	TTCTCAAGCTTATTAAAGGACTTCATCT	TGTTGCATTAAGACAAACTGACTTTC
121586	X	44922797	44922817	44923036	44923063	TCTCTCAGCCTGGAGTCCGT	CCTGAGTGGAGTTAGATAGTTGGTTTT
138911	X	44928929	44928952	44929171	44929190	CAGAATCAGAACGGACATCCCAC	GGCTAGGACTGCAAACAGC
121588	X	44929240	44929261	44929487	44929509	CTCTGCCTTGTTGATGGGAAA	AGATCTGTTTTCATGGGGCTCT
121589	X	44938199	44938220	44938447	44938468	GATTCAGGGCTTCAGCGTATG	GCAACAACTGTGTCCTCACTT
121590	X	44942614	44942638	44942844	44942872	TCTGATTGGAACACAAGGGTTTTG	GTTAGTTCATACAAATTACCTGGTGTTC
121591	X	44948904	44948932	44949124	44949143	AAAAATTTGTGACATTTTCTTCCAGTCT	TTGCACCAGCCAATAGCCT
138916	X	44949848	44949877	44950068	44950093	TTATAGTGTTTTGAGGTTTTCAGAATTGA	GCTTTGGATCTGAGACCTTGATATT
138917	X	44966473	44966492	44966717	44966741	AGGCCTGCTGAGCATTGTC	CGCCCATGCCATATAATCTCTTTT
121594	X	44969244	44969275	44969480	44969513	TAGAATTCCATGATTATACTAAAGCAGTACA	ACATGTTGATTTGACTTACTAATGTAAATTGGT
121595	X	53423306	53423325	53423544	53423564	AAAGGTCCAGGGGCTAGGT	TTGTCCTGTTCCCACCTGCTG
121596	X	53426368	53426389	53426618	53426637	TTTTCCCGTTTAGGTCTTGGT	TCAGAAGACTCGCTTGGGC

183323	X	53432340	53432363	53432580	53432605	GAAGAGAGAAGAGGGGGAGAAGC	GGTGAAGCCTACAGATGAGAAACTC
121598	X	123164719	123164746	123164963	123164988	GTTATGTACCCTATCAGCTACTTCTAC	ACATAAATCTTACCTGCATAGCACT
137931	X	123171168	123171201	123171412	123171437	AGGTCATTTTGTGTTATCAGATAATTGTTTTAC	GGTCAAGAAGTGCTATATCTCGGTC
121600	X	123176266	123176299	123176495	123176527	TAGAGTATCATGGTAAAGATTATTAGATGATGG	AAGTTATCAAAACACTTAAGGTAGTAAGTTAC
137933	X	123178968	123178995	123179212	123179237	AGTCATGCATTCTAAATGAAATTGCTG	AATATTTCTAAATACCGACCTGCCA
223166	X	123181156	123181183	123181394	123181424	TTGTGTCTGTTAGATTAGTTTCACCAT	AGTTATATCCTAATTAATGGTGACCAAGTT
223167	X	123197697	123197719	123197912	123197941	GCATTTGGATGCCTTATTGCGA	GCTCAACTGACAAAATACTTGTAATAGTT
137936	X	123204988	123205016	123205194	123205225	TTTTTAAATAGGCCTTCACTATTCTGTG	GTTTGATTTTATGGTGGACACAGATTTAATA
121605	X	123217147	123217177	123217380	123217408	ATTTTCTATTCCTCAAATTAAAAGCTTGGC	AATACATACACTGTTCTTTTGTCTTGTC
137938	X	123220354	123220382	123220578	123220601	CAGATTTCTGTATCAAAGCTAACAGTTT	CTCAACCACTTTCCGTTTTCCTG
232790	X	129148850	129148871	129149098	129149119	CCTGGGTGAAGAACTCAACTG	CAATGCGGGCTGGAGTATATC
232791	X	129149333	129149352	129149581	129149602	CCACCAGCGTTGTTTCGGA	CCTCAGGCCCCTGATTCATTC
181123	X	129154859	129154882	129155108	129155128	CTCATGCCTCTAGGTCAGAAACG	TACCTTCCTCGTGGCTGTCG
137939	X	133511553	133511586	133511782	133511804	ACTGATTTTAAAATAAAATTAACATTGTCGCCC	TTGCTGCCGGTATACTTACCAT
137940	X	133527420	133527449	133527637	133527663	ACTTTTCAATAACCAATTTGTTTTCCTTG	TGTAAATTCCTTGTGAAGGTTTCTCT
223173	X	133527738	133527769	133527984	133528007	CCATTTCAAAGCATTTCATCATCATCATAAA	GCTGAAACGTGGCTAAATGATGT
223174	X	133547370	133547397	133547614	133547639	ATTGGGTGGCTTTATTGAACATACCAC	GGTATCTTCTGACAGCCCTTTAAAA
121636	X	133547756	133547788	133547992	133548017	GGTAAAGTTCATCATTTGAAATGTTAAGTAAG	GCTTAAAAGAACCATGCTTACCATG
125816	X	133548947	133548979	133549162	133549195	TCTTTCTTCGGAAATTAAATATAACACACTTG	GCAAATCAAGTGTAATGTTGCATAACAAATATA
121638	X	133551064	133551089	133551305	133551333	ATCAAAGTATGGTTTAAGTTGCAGC	CTTGTAAATTCCTCGTGACATATTTTCA
121639	X	133559114	133559134	133559350	133559383	AGTGTGGATGAGGAAACCCA	TCCCATATAATTAACTTTCATACCTAGTTTCCA

Acknowledgements

I want to express my gratitude to everyone who has helped me, in their own unique ways, along this PhD journey.

First of all, I would like to thank Prof. Dr. Karsten Rippe for giving me the opportunity to perform my doctoral studies in his lab. Thank you, Karsten, for making it possible for me to try out all the fancy single-cell methods I wanted to apply for my studies and for giving me the time to learn how to analyse the data on my own. Thanks for providing me the chance to learn new skills and for all the conversations and time you invested. I appreciate your ongoing encouragement and advise as well as your excitement for every aspect of the project.

Further, I want to thank my TAC members, Dr. Sevin Turcan and Dr. Carl Hermann, for the continuous support of my project, the constructive discussions, and the participation on my defense committee. I would additionally like to thank Prof. Dr. Ingrid Lohmann for chairing my PhD defense.

I thank the Genomics & Proteomics Core Facility, the Omics IT, the Data Management Core Facility and the Single-Cell Open Lab of the DKFZ for their support. Especially I would like to thank Dr. Jan-Philipp Mallm and Katharina Bauer for providing me a lab space to conduct my single-cell experiments, for being patient with all of my questions and sharing your knowledge how to lyse erythrocytes in patient samples or to recover messed up libraries.

I would like to thank my collaboration partners, Prof. Dr. Michael Lübbert, Prof. Dr. Hartmut Döhner and Prof. Dr. Konstanze Döhner, for enabling me to work with AML samples from the clinics.

A big thank you to all the past and present members of the Rippe lab: Afzal, Alexandra, Anastasiya, Anne, Armin, Caro, Delia, Emma, Fabian, Isabelle, Jorge, Lara, Lukas, Markus, Nick, Norbert, Robin, Sabrina, Sofie, Stephan and Verena. I really appreciate your support, whether it was just venting about unsuccessful experiments, ordering chemicals, feeding my cells, or helping me with experiments, and analyses. Thanks for all the chats in the coffee corner, the lunch breaks that could turn into little therapy sessions, the Friday beer hours, the Fail Army and Buzzfeed-quiz hangouts, hours of playing cards against humanity or exploding kittens and the Christmas parties. The environment in the lab was always very positive, cooperative, and fun. I'm glad to have worked with you. I would like to particularly thank Stephan for mentoring me in the first half of my PhD and

always picking up the phone when I needed help with data interpretation or analysis during the lockdown. Thank you to Sabrina for helping me out with all the single-cell experiments. You're the most well-organized and detail-oriented technician that one could ever ask for. Thanks to Caro, Lara and Markus for lending me a hand, when the experiments required even more than 2 helping hands. Thanks to Anastasiya, Isabelle and Lara for supporting me with all ATAC-related issues. Especially thank you Simon for installing and deinstalling R packages back and forth and for helping me with all kinds of problems that came up once I decided to do bioinformatics. Big thanks to Robin for all the emotional support during hard times. Additional thanks to my lab bestie Armin who proofread this thesis and provided insightful feedback.

A huge thank you to all the people who made my time in Heidelberg very special.
To my rock and best friend Agnes, who moved from Vienna to Heidelberg so we could be together; happy or miserable, but at least together. No matter what life may throw at us, I know that we will be there for each other and that makes me very blissful.
To all the Nachttänzer for the amazing raves and parties we went to and for turning nights into days. Ahmad, Agnes, Anita, Antonia, Armin, Dominik, Franzi, Janina, Jonny, Joschka and Julia – I will miss our garden chill-outs watching ducks, the weekly P11 appointments, spontaneous trips to the lovely Europapark in Frankfurt and joint dinners.
To Mara and Michel, for prosecco breakfasts that lasted till midnight, all the beer pong tournaments and sewing sessions.
To Sofie, for all the weekends well spent on the couch with pizza, wine, trash TV and cats.
To Jonathan for being the most stubborn person I know and encouraging me to do things I did not think I could and who never stopped believing in me, even when I did.
To all the people from the DJK Handschusheim volleyball team and from the Neckargang who I could spend hours playing with while forgetting all the stress around me.

Most importantly, I want to thank my family from the bottom of my heart. This would not have been possible without you. I will always be thankful for the constant support, helping me move everywhere I go and for giving me the freedom to do what I love. I appreciate how you helped pave my path and for always being there for me.

Last, but not least, thanks to my arts teacher Anna Riebe, who said I should mention her in my speech if I ever win the Nobel Prize or do anything else important. I guess this is as close as it gets.