Aus dem Adolf-Butenandt-Institut
Lehrstuhl: Molekularbiologie
Im Biomedizinischen Zentrum der
Ludwig-Maximilians-Universität München
Direktor: Prof. Dr. Peter B. Becker

Analysis of chromatin accessibility to characterize transcription factor activities in malignancies

Dissertation zum Erwerb des Doktorgrades der
Naturwissenschaften (Dr. rer. nat.) an der
Medizinischen Fakultät der
Ludwig-Maximilians-Universität München
vorgelegt von

Helena Domínguez Moreno

aus Alájar (Huelva), Spanien

2021

Mit Genehmigung der Medizinischen Fakultät der Ludwig-Maximilians-Universität München

Betreuer: Prof. Dr. rer. nat. Gunnar Schotta

Zweitgutachterin: Prof. Dr. María Colomé-Tatché

Dekan: Prof. Dr. med. Thomas Gudermann

Tag der mündlichen Prüfung: 22.03.2022

Eidestattliche Erklärung

Ich, Helena DOMÍNGUEZ MORENO, erkläre hiermit an Eides statt, dass ich die vorliegende Dissertation mit dem Titel

"Analysis of chromatin accessibility to characterize transcription factor activities in malignancies"

selbständig verfasst, mich außer der angegebenen keiner weitern Hilfsmittel bedient und alle Erkenntnisse, die aus dem Schrifttum ganz oder annähernd übernommen sind, als soche kenntlich gemacht und nach ihrer Herkunft unter Bezeichnung der Fundstelle einzeln nachgewiessen habe.

Ich erkläre des Weiteren, dass die hier vorgelegte Dissertation nicht in gleicher oder in ähnlicher Form bei einer anderen Stelle zur Erlangung eines akademischen Grades eingereicht wurde.

München, 31.03.2022	Helena Domínguez Moreno
Ort, Datum	Unterschrift

In the course of my doctoral research, I contributed to other colleagues' projects, some of which led to following publications or are undergoing a peer-revision process in order to be soon published:

Stief, S. M., Hanneforth, A.-L., Weser, S., Mattes, R., Carlet, M., Liu, W.-H., Bartoschek, M. D., **Domínguez Moreno, H.**, Oettle, M., Kempf, J., Vick, B., Ksienzyk, B., Tizazu, B., Rothenberg-Thurley, M., Quentmeier, H., Hiddemann, W., Bosberg, S., Greif, P. A., Metzeler, K. H., Schotta, G., Bultmann, S., Jeremias, I., Leonhardt, H., Spiekermann, K.; *Loss of KDM6A confers drug resistance in acute myeloid leukemia*; Leukemia, 14 June 2019;
DOI: 10.1038/s41375-019-0497-6

Garg, G., Muschaweckh, A., **Domínguez Moreno, H.**, Vasanthakumar, A., Floess, S., Lepennetier, G., Oellinger, R., Zhan, Y., Regen, T., Hiltensperger, M., Peter, C., Aly, L., Knier, B., Reddy Palam, L., Kapur, R., Kaplan, M. K., Waisman, A., Rad R., Schotta, G., Huehn, J., Kallies, A., Korn, T.; *Blimp1 prevents methylation of Foxp3 and loss of regulatory T cell identity at sites of inflammation*; Cell Reports, 12 February 2019;
DOI: 10.1016/j.celrep.2019.01.070

Eichelberger, L., Saini, M., **Domínguez Moreno, H.**, Klein, C., Bartsch, J. M., Falcone, M., Reitberger, M., Espinet, E., Vogel, V., Graf, E., Schwarzmayr, T., Strom, T.-M., Lehmann, M., Königshoff, M., Pfarr, N., Würth, R., Donato, E., Haas, S., Spaich, S., Sütterlin, M., Schneeweiss, A., Weichert, W., Schotta, G., Trumpp, A., Sprick, M. R., Scheel, C. H.; *Maintenance of epithelial traits and resistance to mesenchymal reprogramming promote proliferation in metastatic breast cancer*; Biorxiv preprint, 02 April 2020;
DOI: 10.1101/2020.03.19.998823

Lago, S., Cernilogar, F. M., Kazerani, M., **Domínguez Moreno, H.**, Nadai, M., Schotta, G., Richter, S. N.; *DNA G-guadruplexes: epigenetic regulators that shape the cell type-specific transcriptome*; already accepted by Nature Communications (pending publication)
Biorxiv preprint, 27 August 2020; DOI: 10.1101/2020.08.27.236778

Hiltensperger, M., Beltrán, E., Kant, R., Tyystjärvi, S., Lepennetier, G., **Domínguez Moreno, H.**, Bauer, I., Grassmann, S., Jarosch, S., Schober, K., Buchholz, V. R., Kenet, S., Gasperi, C., Öllinger, R., Rad, R., Muschaweckh, A., Sie, C., Aly, L., Knier, B., Garg, G., Afzali, A. M., Gerdes, L. A., Kümfel, T., Franzenburg, S., Kawakami, N., Hemmer, B., Busch, D. H., Misgeld, T., Dornmair, K., Korn, T.; *Skin and gut imprinted T helper cell subsets exert distinct effector functions in CNS autoimmunity*; Nature Immunology, 06 June 2021
DOI: 10.1038/s41590-021-00948-8

Danksagung/Acknowledgements

An dieser Stelle möchte ich mich herzlich bei allen bedanken, die mich bei der Anfertigung dieser Arbeit in irgendeiner Form unterstürzt haben.

Als erstes möchte ich mich bei **Prof. Gunnar Schotta** bedanken, für die Möglichkeit in seiner Arbeitsgruppe zu promovieren, für die wertvollen wissenschaftlichen Diskussionen, Räte und Unterstützung, sowie für die Korrektur und Begutachtung dieser Arbeit. Herzlichen dank.

Die Zusammenarbeit in der Arbeitsgruppe *Molekularbiologie* (BMC, LMU) war von großer Hilfsbereitschaft geprägt. Hier gilt mein besonderer Dank **Andrea Terrasi**, **Dr. Maryam Kazerani** und **Dr. Filippo Cernilogar** für die Zusammenarbeit, Unterstützung und stets interessante und produktive wissenschaftliche Diskussionen. Ein großes Dankeschön an **Christine Klement**: als ihre Betreuerin habe ich auch viel gelernt und außerdem hat die Zusammenarbeit an ihrer Seite viel Spaß gemacht. Danke an die gesamte Arbeitsgruppe für sowohl fachliche als auch persönliche Unterstützung.

Besonders möchte ich mich bei **Prof. Thomas Korn** bedanken, für die unschätzbare Unterstützung während des schwierigen letzten Jahres und dafür, dass er die Türen für mich offen gelassen hat. Ebenfalls danke an der ganze Arbeitsgruppe *Experimentelle Neuroimmunologie* (Klinikum rechts der Isar, TUM) für die tolle Zusammenarbeit während des letzten Jahres. Vielen Dank an euch alle.

Danke auch den Kollegen, mit denen ich das Glück hatte während meiner Promotion eine enge wissenschaftliche Kollaboration durchzuführen und die Möglichkeit mich weiter zu entwickeln: **Dr. Garima Garg** und **Dr. Andreas Muschawekh** (*Experimentelle Neuroimmunologie*, TUM), **Dr. Laura Eichelberger** (*Institut für Stammzellforschung*, Helmholtz Zentrum München) und **Dr. Sophie Janich** (*Experimentelle Leukämie- und Lymphomforschung*, LMU).

Ebenfalls möchte ich mich bei den Arbeitsgruppen bedanken, die uns die Proben für dieses Projekt zu Verfügung gestellt haben: **AG Jeremias** (*Research Unit Apoptosis in Hematopoietic Stem Cells*, Helmholtz Zentrum München), **AG Götze** (Klinikum rechts der Isar, III; medizinische Klinik, TUM) und **AG Spiekermann**, **AG Subklewe** und **AG Metzeler** (Medizinische Klinik und Poliklinik III, LMU). Danke auch an Lisa Richter (core facility Flow Cytometry, BMC, LMU), Stefan Krebs (LAFUGA, AG Blum, Gene Center, LMU) und Lucas Wange (*Anthropologie und Humanbiologie*, Fakultät für Biologie, LMU) für ihre essentielle Beiträge zu dieser Arbeit.

Außerdem möchte ich **Dr. Elizabeth Schroeder-Reiter** und **Elke Hammerbacher**, Koordinatorinnen der SFB 1243 *Cancer Evolution*, für deren Hilfbereitschaft und Unterschtützung danken.

Zu guter Letzt bedanke ich mich bei meiner Familie, die mich während meines gesamten Studiums unterstützt und ermutigt hat, und bei meinem **Illia**, der immer für mich da war. Muchísimas gracias. Красно дякую.

Abstract

In the context of the study of human disorders, exploring the chromatin accessibility applying the *Assay for Transposase Accessible Chromatin* followed by sequencing (ATAC-seq) in combination with the study of certain histone markers via *chromatin-immunoprecipita-tion* followed by sequencing (ChIP-seq) has proven to be a robust approach. It allows to unravel essential transcriptional regulatory circuitry, as well as to identify genes or genetic pathways, not necessarily presenting mutations, but relevant for any aspect of the disorder, from disease progression to relapse, or even the appearance of therapy resistance.

In the course of this work, three different human disorders were investigated. Our main objective was the study of Acute Myeloid Leukemia (AML), for which both experiments, ATAC-seq and ChIP-seq of the acetylation of lysine 27 of histone H3 (H3K27ac - a marker for active enhancers and super-enhancers) were performed. AML is a hematopoietic malignancy characterized by a broad genetic heterogeneity and the frequent occurrence of relapses, which compromises patient survival. In collaboration with other research groups, we additionally explored on the one hand, the mechanisms involved in the emergence of human breast cancer metastasis, and on the other hand, the response of the immune system in the context of severe autoimmune diseases using a mouse model commonly employed in the investigation of human multiple sclerosis.

To deeply cover all relevant aspects of the AML disease concerning our main goal, I studied AML cell lines as *in vitro* model, an AML patient-derived xenograft (PDX) mouse model as *in vivo* model and, most importantly, paired patient-derived samples at the stages of primary diagnosis and relapse. The analysis of patient samples was of great interest to capture the nature of the disorder in its whole complexity, free of the limitations intrinsic to the two aforementioned models.

The epigenetic landscape in terms of chromatin accessibility and enhancers was characterized based on 18 leukemic cell lines. They exhibited, as expected, a vast heterogeneity, setting the basis for an *in vitro* model to work with and validate hypothesis derived from the study of human samples. To explore the effects caused by the standard therapy over time *in vivo* we used PDX samples. Xenografted mice were subjected to several cycles of therapy. The derived samples were processed at three different time points to analyze the development of chromatin accessibility. Our experiments showed that at the beginning of the treatment chromatin changes barely occurred but as the therapy progressed, the effects accumulated, causing chromatin to significantly open up. The exploration and analysis performed on paired patient-derived samples clearly revealed a larger similarity with regard to the epigenetic landscape for the relapse stage among all patients analyzed. Especially apoptotic-related pathways, although primarily nonexistent in diagnosis, appeared to be commonly active in relapse. Nevertheless, due to the heterogeneity characterizing this disorder, no common pathways or genes could be identified for the whole patient cohort. Notably, for a patient subgroup, we could determine a bunch of relapse-associated relevant genes, some of which have been recently documented in the literature as playing an important role in the disease progression and even in therapy resistance, thus validating our approach.

Zusammenfassung

Im Rahmen der Erforschung menschlicher Erkrankungen, die Untersuchung der Zugänglichkeit des Chromatins mittels der *Assay for Transposase Accessible Chromatin* gefolgt von Sequenzierung (ATAC-seq) kombiniert mit der Untersuchung bestimmte Histonmodifikationen durch *Chromatin Immunoprecipitation* gefolgt von Sequenzierung (ChIP-seq) hat sich als robuster Ansatz erwiesen. Diese Strategie ermöglicht sowohl wesentliche transkriptionelle Regulationsschaltkreise als auch Gene oder genetische Pfade zu identifizieren. Diese weisen nicht unbedingt Mutationen auf, sind aber für manche Aspekte der Erkrankung relevant: von der Krankheitsprogression bis zum Rezidiv oder sogar dem Auftreten von Therapieresistenz.

In dieser Arbeit wurden drei verschiedene Erkrankungen untersucht. Unser Hauptziel war dabei die Untersuchung der menschlichen Akuten Myeloischen Leukämie (AML) und dafür führte ich sowohl ATAC-seq als auch ChIP-seq für Acetylierung des Lysins 27 am Histon H3 (H3K27ac - ein Marker für aktive Enhancers und Super-enhancers) durch. AML ist eine hämatopoetische Malignität, die durch eine breite genetische Heterogenität und das häufige Auftreten von Rezidiven gekennzeichnet ist, wodurch das Überleben der Patienten beeinträchtigt wird. In Zusammenarbeit mit anderen Forschungsgruppen haben wir zwei zusätzliche menschliche Erkrankungen untersucht. Zum einen waren es die Mechanismen, die an der Entstehung von Brustkrebsmetastasen beteiligt sind. Zum anderen, die Reaktion des Immunsystems im Kontext schwerer Autoimmunerkrankungen. Dafür wurde ein Mausmodell angewendet, das häufig für die Erforschung menschlichen Multiplen Sklerose benutzt wird.

Um alle relevanten Aspekte der AML umfassend abzudecken, untersuchte ich drei verschidene Modelle: erstens, AML-Zelllinien als *in-vitro*-Modell; zweitens, ein von AML-Patienten abgeleitetes Xenograft (PDX)-Mausmodell als *in vivo*-Modell; und drittens, gepaarte Patientenproben in den Stadien der Erstdiagnose und des Rezidives. Die Analyse der Patientenproben war vom großen Interesse um die Natur der Störung in ihrer ganzen Komplexität zu erfassen, frei von den Einschränkungen, die in den zuvor genannten Modellen implizit enthalten sind.

Die epigenetische Landschaft bezüglich der Chromatinzugänglichkeit und der Enhancer wurden an 18 Leukemie-Zelllinien charakterisiert. Diese Zelllinien wiesen erwartungsgemäß eine große Heterogenität auf. Das bildete die Grundlage für das *in vitro*-Modell, mit dem Hypothesen validiert werden können, die aus der Untersuchung menschlicher Proben abgeleitet worden sind. Mithilfe von PDX-Proben wurden die Auswirkungen der Standardtherapie über die Zeit *in vivo* untersucht. Dafür wurden xenotransplantierte Mäuse mehreren Therapiezyklen unterzogen. Zu bestimmten Zeitpunkten wurden die von Mäusen abgeleiteten Proben verarbeitet um die Entwicklung der Chromatinzugänglichkeit zu analysieren. Die Ergebnisse der Experimente haben gezeigt, dass zu Beginn der Behandlung kaum Chromatinveränderungen auftraten, sich jedoch im Verlauf der Therapie die Wirkungen akkumulierten, wodurch sich das Chromatin deutlich öffnete. Die Analyse von gepaarten Patientenproben zeigte eine deutliche Ähnlichkeit in Bezug auf die epigenetische Landschaft für das Rezidivstadium bei allen analysierten Patienten. Insbesondere schienen apoptotisch bedingte Signalwege beim Rezidiv häufig aktiv zu sein, obwohl sie in der Diagnose nicht vorhanden waren. Jedoch aufgrund der Heterogenität, die diese Erkrankung charakterisiert,

konnten für die gesamte Patientenkohorte keine gemeinsamen Signalwege oder Gene bestimmt werden. Nichtsdestotrotz konnten wir für eine Patientenuntergruppe eine Reihe von rezidivassoziierten relevanten Genen identifizieren, wobei einige von denen vor kurzem in der Literatur als wichtig für die Krankheitsprogression und auch für die Therapieresistenz dokumentiert wurden. Dadurch ist unser Ansatz grundsätzlich validiert.

Contents

List of Figures

List of Tables

List of Abbreviations

ab	antibody
AG	*Arbeitsgruppe*, working group (translation from original in German)
ALL	acute lymphoblastic leukemia
AML	acute myeloid leukemia
ATAC-seq	assay for transposase-accessible chromatin followed by high throughput sequencing
BLI	bioluminescence image
BM	bone marrow
BMC	Biomedical Center (LMU)
bp	base pair
CDH1	E-cadherin
ChIP	chromatin immunoprecipitation
ChIP-seq	chromatin immunoprecipitation followed by high throughput sequencing
CML	chronic myeloid leukemia
CNS	central nervous system
CNS2	conserved non-coding sequence 2 (in the first intron of the *Foxp3* locus)
CR	complete remission
CRE	cis-regulatory element
D	primary diagnosis
DNA	deoxyribonucleid acid
dLN	draining lymph node
EAE	experimental autoimmune encephalomyelitis
EDTA	ethylenediaminetetraacetic acid
EMT	epithelial-to-mesenchymal transition
EPCAM	epithelial cell adhesion molecule
ER	estrogen receptor
FAB	French-American-British classification of AML
FACS	fluorescence activated cell sorting
FC	fold change
FCS	fetal calf serum
FH	femoral head
GO	gene ontology
GRHL1/2	grainyhead like transcription factor 1/2
GSEA	gene set enrichment analysis
H3K27ac	acetylation at the 27th lysine residue of the histone H3 protein
HAT	histone acetyltransferase
HD	healthy donor
HDAC	histone deacetylase
Hep	heparin

h.i.	heat inactivated
HMLE	human mammary epithelial cell line
HOMER	Hypergeometric Optimization of Motif EnRichment
HPC	high performance computational (cluster)
HSC	hematopoietic stem cell
I	input
IFN	interferon
IGV	Integrative Genomics Viewer
IL	interleukin
IP	immunoprecipitation
K	lysine
KO	knock-out
LN	lymph node
MACS	magnetic cell sorting
MBC	metastatic breast cancer
MDS	myelodysplastic syndrome
Mio	million
MIR	microRNA
N	amino
NA	not available
NES	normalized enrichment score
NGS	next generation sequencing
OE	overexpression
OVOL	ovo-like zinc finger
PB	peripheral blood
PBS	phosphate buffered saline
PCA	principal component analysis
PCR	polimerase chain reaction
PDX	patient derived xenograft
PI	propidium iodide
PTM	post-translational modification
pval	p value
qPCR	quantitative polimerase chain reaction
R	relapse
rcf	relative centrifugal force
RNA	ribonucleid acid
RNA-seq	RNA high throughput sequencing
ROSE	Rank ordering of super-enhancers
rpm	revolutions per minute
RPKM	reads per kilo base per million mapped reads
RT	room temperature
SCC	single-cell clone
SE	super-enhancer
SEER	Surveillance, Epidemiology and End Results Program
SPL	spleen
SSC-A	side scatter pulse area
TAM	tamoxifen

Tconv	conventional T (cell)
TD	tagmentation DNA (buffer)
TF	transcription factor
TFBS	transcription factor binding site
TIC	tumor-initiating capacity
Tn5	hyperactive Tn5 transposase (used for ATAC-seq)
Treg	regulatory T (cell)
TSDR	Treg cell-specific demethylated region (in the first intron of the *Foxp3* locus)
t-SNE	t-distributed stochastic neighbor embedding
TSS	transcription start site
WHO	World Health Organization
WT	wild-type
YFP	yellow fluorescent protein
ZEB1	zinc finger E-box binding homeobox 1
2D	two-dimensional
3D	three-dimensional
3C	Chromosome Conformation Capture

Chapter 1

Introduction

1.1 Epigenetics and chromatin structure

The term *epigenetics* (*"epi-"* from the ancient Greek "επι-", meaning *over, beyond, around, outside of*) has evolved since it was first coined by C. Waddington in 1942, being nowadays generally understood as the research field studying heritable changes affecting gene expression when no alterations in the DNA sequence are involved (Waddington, 1942; Berger et al., 2009; historical review by Allis et al., 2016), although discussion about it still continues (Deans et al., 2015).

Early studies previous to 1930 unraveled the structure allowing the dense packaging of DNA inside the eukaryotic cellular nucleus, the chromatin (Kornberg, 1974; Kornberg et al., 1999). In it, 145 to 147 base pairs (bp) of double-stranded DNA are wrapped around an octamer of histone proteins into nucleosomes separated by linker DNA (Figure 1.1). Nucleosomes constitute the basic structural units in the chromatin fiber and the nucleosome core consists of dimers of the histones H2A, H2B, H3 and H4, two copies of each. There is another histone stabilizing the octamer structure, the linker histone H1, which also may intervene in other relevant processes (McGhee et al., 1980; McGinty et al., 2015; Fyodorov et al., 2018; Zhou et al., 2018). Chromatin can present two states depending on its folding grade: *euchromatin* consists of decondensed, transcriptionally active chromatin, whereas the term *heterochromatin* refers to a highly compact state and, therefore, inaccessible and transcriptionally inactive chromatin. The chromatin structure is actually very flexible and can react dynamically through nucleosome remodeling to environmental, developmental or other stimuli changing the structure of some regions and modulating the spatial and temporal gene regulation but also influencing other fundamental DNA processes, as repair, replication and recombination (Bao et al., 2007; Altaf et al., 2007; Becker et al., 2013). In this context, histone modifications, histone tail cleavage or binding of certain proteins play a key role in the regulation of the nucleosome positioning and the higher-order chromatin structure and, therefore, in the way DNA is packaged (Henikoff, 2008; Boyle et al., 2011; Yi et al., 2018). Histone modifications consist in covalent post-translational modifications (PTMs), first described by V. G. Allfrey in the early 1960s (Allfrey et al., 1964). Solving the nucleosome structure in 1997 (Luger et al., 1997; Fletcher et al., 1996) allowed the observation of the overhanging histone amino (N)-terminal tails, which could interact with adjacent nucleosomes and recruit non-histone proteins and other complexes with specific enzymatic activities (Luger et al., 1998).

Histones can be modified in several different ways but the most common and best studied histone modifications consist of acetylation, methylation, phosphorylation and ubiquination (Kouzarides, 2007; Bannister et al., 2011). Already in early studies, it was observed the link between variable levels of histone acetylation and methylation and alterations in transcription rates (Allfrey et al., 1964; Brownell et al., 1996). Indeed, histone acetylation influences fundamental changes in the chromatin structure, being its presence associated to a state of accessible chromatin, whereas repressive chromatin is characterized by lack of acetylation (Eberharter et al., 2002; Verdone et al., 2006). Acetylation is one of the most studied histone modifications in the context of gene regulation (Grunstein, 1997; Struhl, 1998; Kurdistani et al., 2003). It takes place at lysine residues on the amino-terminal histone tails and neutralizes the positive charge of the histone tails, therefore reducing their affinity to DNA (Hong et al., 1993).

The opposing processes of acetylation and deacetylation are mediated by two main enzyme families: histone acetyltransferases (HATs) and histone deacetylases (HDACs). HATs catalyze the transfer of an acetyl group to the lysine ε-N groups on the amino-terminal histone tails using acetyl CoA as cofactor (Roth et al., 2001). Through this process, HATs neutralize the positive charge of lysine, causing a decrease in the capacity of histones to interact with DNA and, as a consequence, increasing chromatin accessibility. HDACs present antagonistic effects, reversing lysine acetylation and restoring in this way the lysine positive charge, which results in a reduction of chromatin accessibility (Kuo et al., 1998; Turner, 2000; Hyndman et al., 2017).

1.2 Enhancers and super-enhancers (SEs) as key regulatory elements

Spatial and temporal gene regulation is fundamental for the correct development and differential function of cells and tissues. Enhancers are *cis*-regulatory elements (CREs) consisting in short DNA sequences, typically between 100-1000 bp, capable of modulating the transcription of their target genes, regardless of their orientation and relative distance to the transcription start site (TSS), in contrast to promoter regions, which are located in close vicinity to TSSs.

Enhancers are key elements in the regulation of cell-type specific genes (Heinz et al., 2010b) and, although it has been estimated that mammalian genomes contain millions of enhancers, only a small portion of them is active at a specific time in a given cell type. The long-range communication between enhancers and the promoters of their target genes requires complex three-dimensional chromatin folding to achieve physical proximity (Smallwood et al., 2013; Robson et al., 2020) and it is mediated by sequence-specific DNA binding proteins known as transcription factors, TFs (Spitz, 2012), whose binding primes enhancers and contribute to the recruitment of other coactivator proteins and chromatin regulators required for the regulation of gene expression in a temporal- and cell-type-specific manner (Roeder, 2005; Weake et al., 2010). Thus, enhancers are central regulatory elements during normal development (Shlyueva et al., 2014; Schoenfelder et al., 2019) but also play an essential role in pathogenesis (Miguel-Escalada et al., 2015; Luo et al., 2016) and, of course, in cancer (Aran et al., 2013; Herz et al., 2014; Sur et al., 2016).

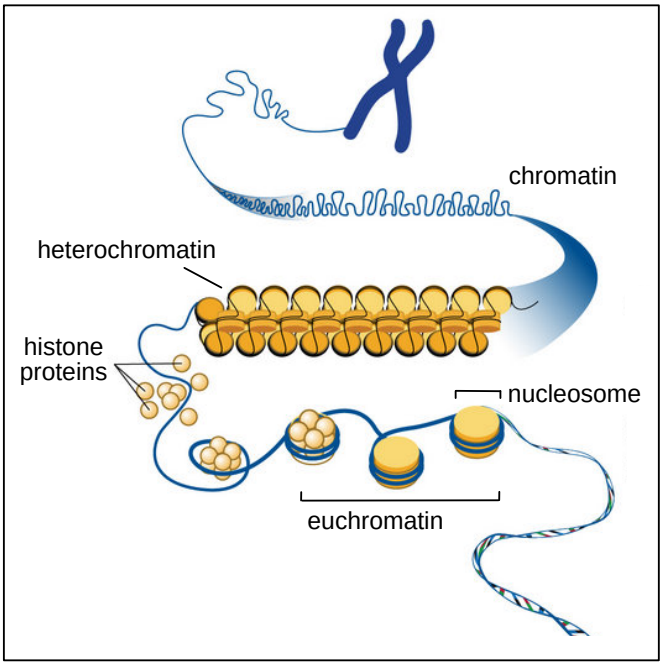

FIGURE 1.1: Schematic representation of the chromatin: the DNA double helix is wrapped around histone octamers into nucleosomes and these are packed into a more compact structure, the chromatin. Chromatin can exist in an active state and be accessible, *euchromatin*, or on the contrary, it can be in a repressed or silenced state, *heterochromatin*. Figure modified from *National Human Genome Research Institute*.

Large clusters of proximal enhancers are known as super-enhancers (SEs). These enhancer clusters characteristically present high level of TFs binding and have been demonstrated to be particularly relevant for the expression of genes defining cell identity (Whyte et al., 2013; Pott et al., 2015). As such, super-enhancers have also been observed to be implicated in human tumors and other developmental disorders (Niederriter et al., 2015; Quang et al., 2015; Gartlgruber et al., 2020). Specifically, Hnisz et al., 2013, besides creating a SE catalogue comprising several human cells across different tissues, observed that tumor cells are able to generate SEs at oncogenes and other genes involved in tumor development and relevant for the acquisition of favorable features for cancer cells. These SE features make them a valuable tool to explore regulatory characteristics in the context of disease and potentially exploit them for therapeutic purposes (McKeown et al., 2017).

Enhancers can be classified attending to their state as inactive, primed, poised or active (Ernst et al., 2010). In brief, inactive enhancers are characterized by a compact chromatin conformation, preventing the binding of TFs and other histone modifications. Primed enhancers, although placed in nucleosome-depleted, accessible chromatin regions for TFs to bind, may require the concurrence of other TFs and co-activators for fully enhancer activation. Poised enhancers, first described in the context of human embryonic development (Rada-Iglesias et al., 2011), are basically primed enhancers presenting some repressing chromatin modifications preventing them to be active. Enhancers can change their state dynamically, activating or increasing the transcription of the genes they regulate or, on the contrary, silencing those genes. Histone modifications are important features for the identification of enhancers and the characterization of their state. In this regard, it has been observed that enhancers exhibit enrichment of histone H3 lysine (K) 4 monomethylation, H3K4me1 (Heintzman et al., 2009) or H3K4me2 (He et al., 2010) and deficiency of H3K4me3 compared to gene promoters (Heintzman et al., 2007). Nevertheless, these histone characteristics are not sufficient to define functional enhancers. Primed enhancers display H3K4me1 and H3K4me2 and lack acetylation, whereas poised enhancers present additionally the repressive mark H3K27me3 (Rada-Iglesias et al., 2011; Zentner et al., 2011). Indeed, acetylation of lysine 27

on histone H3, i.e., H3K27ac, has been reported to characterize active enhancers (Creyghton et al., 2010; Calo et al., 2013 for review).

1.3 Methods to study chromatin structure and chromatin accessibility

Active enhancers and super-enhancers are usually located in chromatin regions presenting certain characteristics. They tend to overlap to chromatin areas scarce in nucleosomes and these particular regions can be identified through specific epigenetic hallmarks, such as chromatin accessibility and specific histone marks, as H3K27ac (Visel et al., 2009; Ong et al., 2011; Spicuglia et al., 2012). The genome accessibility has been estimated to be about 2-3% of total DNA, accumulating these accessible regions more than 90% of TFs bindings (Thurman et al., 2012). This means, in other words, that the vast majority of TFs bind to already open chromatin regions and only a minor part of TFs, the so-called pioneer transcription factors, are actually able to bind condensed chromatin regions and initiate structural changes in its conformation (Cirillo et al., 2002; Zaret et al., 2011; Cernilogar et al., 2019) to promote local access to DNA (Felsenfeld, 1996).

Several methods have been developed to assess chromatin accessibility: DNase-seq, FAIRE-seq, and ATAC-seq and its variants are all methods based on the direct isolation of the accessible chromatin regions, in contrast to the MNase-seq assay, which is an indirect method to evaluate chromatin accessibility. All these approaches measure in a direct fashion the effects caused by modifications in the chromatin structure on gene transcription. In contrast, approaches based on the detection and characterization of chromatin-bound proteins, comprising methods such as ChIP-seq or the more recent CUT&RUN, deduce those effects on gene transcription from the detection of certain overlapping histone modifications or bound TFs. Klemm et al., 2019, Klein et al., 2020 and Minnoye et al., 2021 recently reviewed and compared the aforementioned assays, among others.

Since enhancers and clusters of enhancers are distal regulatory elements, one of the main challenges consists in establishing the link between a particular enhancer or SE and the gene or genes it regulates. For this purpose, several methods have been developed in the past years, both experimental, mostly based on Chromosome Conformation Capture (3C) technologies (Dekker et al., 2002; Dostie et al., 2006; Belton et al., 2012), but also on CRISPRi-mediated epigenome editing techniques (Fulco et al., 2016; Kim et al., 2019) and computational methods (Hariprakash et al., 2019; Moore et al., 2020), relying on the information compiled in databases generated over time.

The two main technologies applied in the course of this work are H3K27ac (as histone mark associated with active enhancers) ChIP-seq and one of the ATAC-seq protocol variants, specifically the Omni-ATAC-seq. The principles on which these techniques are based are briefly described below and the corresponding protocols are comprehensively documented in Chapter 4.

1.3.1 ChIP-seq

Chromatin immunoprecipitation followed by high throughput sequencing (ChIP-seq) has been an essential technique for the study and exploration of the epigenome. This method

relies on antibodies capturing specific DNA-binding proteins or histone modifications to determine genome-wide enriched regions at base-pair resolution (Rhee et al., 2012; Furey, 2012). To do so, the DNA is first cross-linked with the associated proteins on chromatin and, in a later step, DNA-protein complexes are sheared into small DNA fragments applying sonication (Figure 1.2).

There are multiple variants of the ChIP-seq protocol, optimized for the cell type to be processed, the amount of available input cells or the target proteins or histone modifications. The specific ChIP-seq protocols used for the completion of this work are described in detail in Chapter 4, Sections 4.3.5 and 4.3.6.

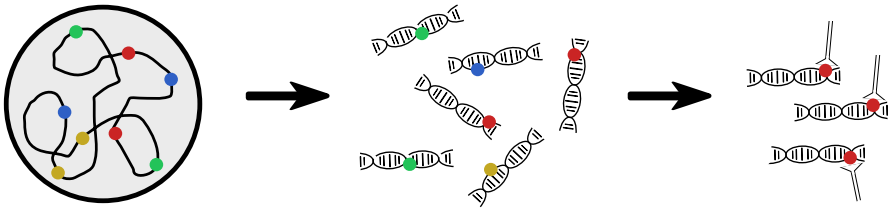

FIGURE 1.2: Schematic representation of the ChIP-seq process: chromatin in cell nucleus is cross-linked with bound proteins (represented by colored circles). Then, DNA-protein complexes are sheared in small fragments and specific DNA fragments bound by the protein of interest are co-immunoprecipitated. In a later step (not shown) DNA fragments are decross-linked for ulterior library preparation and sequencing, to identify genome-wide sites associated with the studied protein.

1.3.2 ATAC-seq

The Assay for Transposase-Accessible Chromatin followed by sequencing (ATAC-seq) was first described by Buenrostro et al., 2013 as an alternative technique to evaluate accessible chromatin profiles. It is based on the hyperactivity of a genetically engineered DNA transposase, Tn5, preloaded with sequencing adapters (Goryshin et al., 1998). This hyperactive enzyme is able to simultaneously cleave and tag accessible chromatin regions, i.e., nucleosome-depleted regions, genome-wide (Figure 1.3). This method has several advantages over other approaches used up to that date: higher sensitivity, lower input cells requirement, relatively low sequence depth needed, significantly shorter preparation time and simpler protocol due to the combination of chromatin fragmentation and tagmentation in one step.

The first version of the protocol was improved after some years by Corces et al., 2016 to perform better for hematopoietic human cells (being known this new version as Fast-ATAC-seq) and again one year later some modifications were added by Corces et al., 2017, improving the signal-to-noise ratio as well as reducing the mitochondrial DNA interference (Omni-ATAC-seq version).

Detailed protocols used for processing the samples in the context of this work are included in Chapter 4, Sections 4.3.3 and 4.3.4.

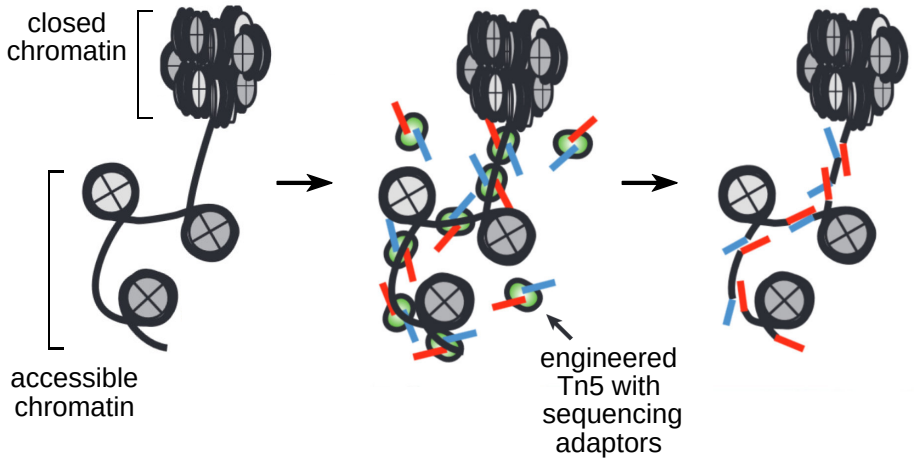

FIGURE 1.3: Schematic representation of the ATAC-seq process: the engineered hyperactive Tn5 transposase, preloaded with sequencing adapters, fragments and tags chromatin accessible regions. Figure modified from Buenrostro et al., 2013.

1.4 Epigenetics and human diseases

1.4.1 Cancer

For long time, genetic factors were thought to be the main trigger behind certain human health disorders. This vision started to change half a century ago, when C. Waddington introduced the term *epigenetics* in the middle of the 20th century (Waddington, 1942). The initial definition has evolved and nowadays it is commonly accepted that the term defines the study of heritable features associated with no changes in the nucleotide sequence but to chemical DNA changes or other modifications regarding structural or regulatory proteins binding it.

As previously described, a particular chromatin status, which ultimately conditions gene expression, is the result of the interplay of all epigenetic processes (post-translational histone modifications, including acetylation, methylation or phosphorylation, as well as binding of DNA-specific proteins). All these processes constitute what Waddington referred to as "epigenetic landscape", today named "epigenome", or the epigenetic status determining how the genome reveals itself in a time- and cell type-dependent manner. When this intricate regulatory machinery does not work properly, cancer and other disorders may arise. Moreover, tumor development and other human diseases have been documented to show abnormal epigenetic regulation (Esteller, 2007; Lee et al., 2013; Kuznetsova et al., 2016). Particularly in the case of cancer, the epigenetic landscape displays aberrant histone modification patterns as well as altered DNA methylation (comprising global hypomethylation but also local methylation gains). In this sense, multiple studies have reported patterns of hypermethylation at cytosine bases within CpG island in promoters of genes with tumor-suppressor function, repressing their expression (Galm et al., 2006; Boultwood et al., 2007), as well as other histone marks alterations (Fraga et al., 2005). Additionally, altered function of enzymes in charge of adding, removing or recognizing particular histone modifications has been described in many types of human cancers (Chi et al., 2010). In hematological malignancies, for example, aberrant function of the family of lysine acetyltransferases due to chromosomal translocations has been documented (Yang, 2004).

Given the advances occurred in the field of epigenetics in the past decades, including increasingly refined and precise techniques to explore the epigenome, there is a growing interest in investigating and developing new therapeutic strategies targeting altered regulatory mechanisms (Cheng et al., 2019; Mohammad et al., 2019; Lu et al., 2020). Different to gene mutations, epigenetic alterations are highly dynamic and can be by its nature, therefore, potentially reversed. This is the reason, why multiple studies have been recently analyzing potential epigenetic targets for cancer therapy (Kelly et al., 2010; Arrowsmith et al., 2012; Furtado et al., 2019; Montalvo-Casimiro et al., 2020), although human disorders frequently exhibit a complex mixture of both genetic and epigenetic lesions or dysregulation (Li et al., 2016), usually accumulating over time.

Taking all the previous into consideration, the main goals pursued by our research were to investigate the role of epigenetic factors in three different disorders, namely: acute myeloid leukemia, metastatic breast cancer and inflammation associated to immune diseases.

1.4.1.1 Acute Myeloid Leukemia and the challenge of relapse

Hematopoiesis is a fundamental process driven by hematopoietic stem cells (HSCs), in charge of generating all lineages of blood and immune cells (Lemischka et al., 1986; Naik et al., 2013; Busch et al., 2015). When the homeostasis in this delicate system is disrupted, leukemogenesis can arise (Bonnet et al., 1997). Leukemia is a hematopoietic disorder initiated in the bone marrow where blood precursor cells, known as *blasts*, are not able to normally develop and differentiate into fully functional and mature blood cells and, instead, begin to build up, causing an accumulation of immature blood cells and simultaneously suppressing the development of normal, functional blood cells (Figure 1.4). Leukemia is the most common form of cancer in children, accounting for 30% of all cancers diagnosed in children under 15 years according to the World Health Organization, *WHO*, and in adults leukemia is included in the top ten of the most common forms of cancer (Surveillance, Epidemiology and End Results Program *SEER*). Pathologically, leukemia can be divided into several subgroups, based on the type of blood precursor cells blasts originate from (myeloid or lymphoblastic), and the rate at which the leukemia progresses (acute or chronic). The increase of immature blood cells or blasts leads to a reduction of the functional blood cells and can cause symptoms such as weakness or fatigue, associated with a loss of red blood cells (anemia), increased number or repetitive infections, when neutrophils are reduced (neutropenia) or abnormal bleeding, linked to the loss of platelets (thrombocytopenia).

Acute myeloid leukemias (AMLs) are neoplastic proliferations of myeloid progenitor cells, i.e., immature granulocytes, monocytes, erythrocytes and megakaryocytes. AML was first classified by the FAB criteria (French-America-British AML classification), established in the 1970s and based on cytomorphologic and cytochemical features, with lineage assigned as myeloid by the presence of Auer rods or ≥ 3% myeloperoxidase positive blasts (Bennet et al., 1976; Bennet et al., 1985). More recently, WHO developed a new classification based on a combination of morphological, immunophenotypic, genetic and clinical features (Brunning et al., 2001; Vardiman et al., 2002; Swerdlow et al., 2016). A blast count below 2% in bone marrow aspirates or peripheral blood is considered to be normal. AML is usually diagnosed when these immature blood cells represent more than 20% of the total cell count. Blast count in between is termed as myelodysplastic syndrome (MDS), although this classification has changed towards the last decades and is still subject of debate (DiNardo et al., 2016). Even if

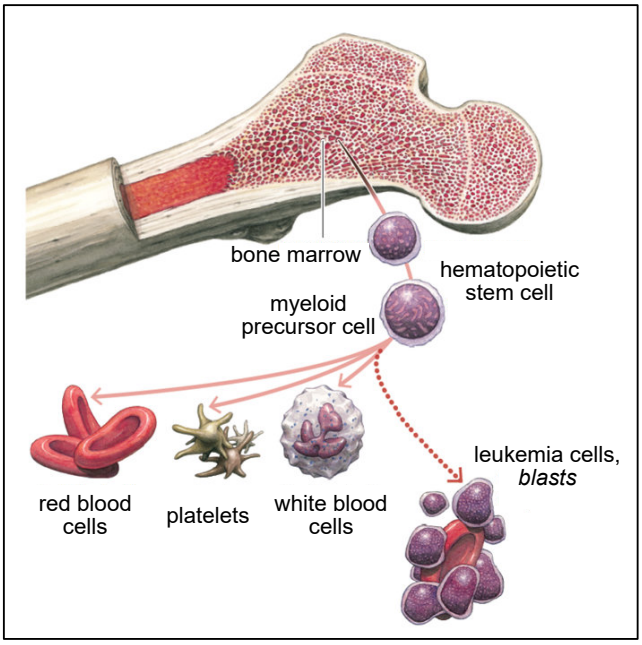

FIGURE 1.4: In AML, leukemic cells or *blast* originate from a myeloid progenitor and accumulate in the bone marrow interfering with the normal production of mature, fully functional blood cells. Figure modified from OSHU Knight Cancer Center *OHSU Knight Cancer Institute*.

most patients diagnosed with any type of leukemia or MDS usually present a higher percentage of blasts in the bone marrow than in the peripheral blood, there are evidences suggesting that the opposite can also occur, in which case the disease appears to progress more aggressively (Amin et al., 2005).

AML is indeed the most frequent type of acute leukemia diagnosed in adults (Kouchkovsky et al., 2016). According to data collected by *SEER*, the median age at diagnosis is 68 years. Moreover, adults over the age of 65 represent more than 50% of the total new cases diagnosed per year. The outcome is worse, the older the patients are at age of diagnosis, with 70% of patients older than 65 years dying within one year from diagnosis (Meyers et al., 2013).

The standard treatment administered to AML patients remains the same since it was first established in 1973. It consists in first place in applying induction chemotherapy to achieve remission eliminating blasts from blood, followed by a consolidation phase with cyclical chemotherapy targeting potential blasts leftovers. The standard induction chemotherapy, known as *"7+3"*, consists in a combination of cytarabine applied for 7 days, followed by an anthracycline drug, such as daunorubicin, for 3 days (Lichtman, 2013). This treatment is associated with toxic side effects, making elderly patients more vulnerable, especially in presence of comorbidities, such as diabetes or heart disease (Kimby et al., 2001). Many patients, thus, do not tolerate the aggressiveness of the standard chemotherapy and only palliative care can be applied (Bryan et al., 2015; Palmieri et al., 2020). Even when the survival within five years after diagnosis has improved over the past decades, this fact is predominantly due to the major advances and refinement in supportive care (Derolf et al., 2009, Ferrara et al., 2013) rather than to improvements in the therapy itself (Gbadamosi et al., 2017; Kolitz et al., 2018; Zeidner et al., 2018). Therefore, there is the urgent need of developing new alternative therapeutic strategies to treat this disease in a more efficient way.

AML usually develops in a stepwise manner, through accumulation of genetic and epigenetic lesions (Shlush et al., 2014; Xu et al., 2019; Chopra et al., 2019). Furthermore, the bone marrow microenvironment maintaining and regulating hematopoiesis is increasingly

thought to play a crucial role in hematological disorders and therapy resistance (Wang et al., 2018; Kogan, 2019; Méndez-Ferrer et al., 2020). Generally, AML patients present highly heterogeneous mutational profiles, with several AML sub-clones potentially co-existing in parallel in the same patient. Moreover, one of the main reasons of treatment failure is the high occurrence of relapse after complete remission (CR). Even when patients respond well to the induction chemotherapy, most patients will eventually relapse (Döhner et al., 2015; Oliva et al., 2018), with tumor cells becoming frequently treatment-resistant, what compromises the long-term patient survival (Hackl et al., 2017).

Relapse is associated with clonal evolution at the cytogenetic level (Testa et al., 1979; Garson et al., 1989; Ding et al., 2012; Vosberg et al., 2019). Chemotherapeutic treatment represents a new environment in which AML cell populations may be able to adapt through the accumulation of random and heritable mutations with the subsequent survival and regrowth of some clones. Nevertheless, the mechanisms for which AML patients become treatment-resistant, even if they were treatment-sensitive in first term, remain poorly understood (Brown et al., 2014; Döhner et al., 2017; Zhang et al., 2019).

Previously, many studies focused on particular cytogenetic abnormalities in AML patients to predict outcome and other prognostic implications (Grimwade et al., 1998; Grimwade et al., 2001; Byrd et al., 2002). More recently several studies have taken advantage of next-generation sequencing (NGS) to try to identify relapse-specific gene mutations in certain AML-subgroups, which could explain the course of the disease. However, no consistent gene mutations were found to be acquired in relapse across different studies, being the genetic AML profile very heterogeneous among different patients (Jan et al., 2012; Papaemmanuil et al., 2016; Metzeler et al., 2016), even when focusing in specific AML subgroups (Gröschel et al., 2015; Wang et al., 2016; Ahn et al., 2017). Moreover, AML was shown to have a low rate in genetic lesions compared to most solid tumors in general (Kandoth et al., 2013; Lawrence et al., 2013b; Lawrence et al., 2014), evidencing the importance of changes at the epigenetic level in the course of the disease and the need of considering the intrinsic and vast disease heterogeneity in order to explore alternative drug development strategies (Vicente-Dueñas et al., 2018; Schwenger et al., 2021; Levin et al., 2021).

In the course of our work, we confirmed the broad regulatory heterogeneity intrinsic to the AML disorder among the patient-derived samples, which was consequently reflected in the *in vitro* model based on AML cell lines. Despite this noticeable diversity among patients regarding the regulatory landscape, we were able to find out common features differentiating the stages of primary diagnosis and relapse. Indeed, we detected the activation of regulatory pathways involved in apoptotic processes only during relapse but not in the primary diagnosis. Moreover, we were able to identify some promising genes in the relapse stage shared by a subgroup of patients, which potentially confer the tumor cells resistance to therapy and, therefore, could be interesting targets for developing more effective anti-cancer therapies in the context of this disease.

1.4.1.2 Breast cancer and the challenge of metastases

When focusing on solid tumors, one of the main challenges when treating patients suffering from cancer, even when reacting initially well to the treatment, is the emergence of distal metastases. Distal metastases compromise enormously the patient survival. Particularly, for breast cancer patients, approximately 6-10% of new diagnosed cases are initially

metastatic and 20-30% of all breast cancer cases are estimated to become metastatic, while less than 30% of patients presenting distal metastases survive more than 5 years from diagnosis (*SEER*; O'Shaughnessy, 2005). Metastasis constitutes a complex process, still not fully understood, by which tumor cells develop the capacity to colonize and proliferate in distant tissues different from the tissue of origin, being a threat for the normal function of vital organs and causing ultimately organ failure and even death.

In the context of breast cancer, it has been previously reported that circulating tumor cells (CTCs) present an increased metastatic tumor initiation capacity (TIC) when the Epithelial Cell Adhesion Molecule (EPCAM) is highly expressed, compared to those CTCs with low EPCAM expression (Baccelli et al., 2013). Besides, the epithelial-to-mesenchymal transition (EMT), a complex molecular and cellular mechanism, critical for tissue and organ development, has been extensively investigated and many preceding studies have related it with tumor initiating capacity, metastatic outgrowth and chemoresistance in many different types of cancer (Tsai et al., 2012; Krebs et al., 2017; Legras et al., 2017), although the results from different studies differ substantially and there is still debate regarding the role of the EMT or which are the key factors involved in the metastatic process (Ocaña et al., 2012; Pastushenko et al., 2018; Kröger et al., 2019). Nevertheless, there are indications suggesting that the transient and not the permanent expression of EMT-related transcription factors (TFs) is in first term responsible for metastatic outgrowth (Schmidt et al., 2015), which is in line with the observation that most macroscopic metastases show an epithelial phenotype (Kowalski et al., 2003). We confirmed that high expressing EPCAM cells ($EPCAM^{high}$) are more prone to initiate tumors and ultimately metastases in immunocompromised mice compared to cells with low EPCAM expression ($EPCAM^{low}$). Our results suggest that $EPCAM^{high}$ cells are able to epigenetically prevent the repression of epithelial genes upon EMT and, thus, preserve their epithelial identity. In contrast, $EPCAM^{low}$ cells, with lower tumor-initiating capacity, transition to a stable mesenchymal phenotype upon EMT. In the light of our results, this capacity of retaining epithelial status while displaying epithelial-mesenchymal flexibility is key for initiating metastatic proliferation.

1.4.2 Inflammation and regulatory T cells

Regulatory T (Treg) cells are crucial for restoring the homeostasis upon emergence of immune diseases. Nevertheless, T cell identity and functionality are compromised in case of severe inflammation (Zhou et al., 2009). In this context, rodent experimental autoimmune encephalitis (EAE), an autoimmune disease used commonly as model for human multiple sclerosis (Constantinescu et al., 2012) was used to investigate Treg cell behavior at the peak of the disease in the central nervous system (CNS).

Forkhead box P3 (Foxp3) is a key regulatory gene encoding a crucial transcription factor in the development and function of Treg cells (Hori et al., 2003, Yagi et al., 2004). When inflammation occurs, there is a reduction in the Foxp3 expression in Treg cells caused by Dnmt3a mediated methylation of the Foxp3 intronic element known as conserved non-coding sequence 2 (CNS2) or Treg cell-specific demethylated region (TSDR), whose demethylation is required for the normal expression of Foxp3 (Floess et al., 2007). This loss of Foxp3 expression contributes to destabilize Treg cells. We observed that $Foxp3^{+}$ Treg cells expressed Blimp1 in the inflamed tissue, preventing the methylation of the *Foxp3* locus through inhibition of the methyltransferase Dnmt3a. Blimp1, encoded by *Prdm1*, is a zinc finger protein

serving as transcriptional regulator promoting the differentiation of fully functional $CD8^+$ T cells (Kallies et al., 2009, Rutishauser et al., 2009, Shapiro-Shelef et al., 2003) and it has also been found as a part of a transcriptional module mediating the presence of $CD8^+$ T cells in non-lymphoid tissues (Mackay et al., 2016). The results and conclusions of our investigations, describing a novel function of the transcriptional regulator Blimp1, were published (Garg et al., 2019): Blimp1 is expressed in Treg cells in an inflamed environment to protect themselves from destabilization, since it counteracts the loss of expression of the key factor Foxp3.

1.5 Objectives

For the completion of this work, we have studied three different disorders from a regulatory point of view. The first section presented in Chapter 2 consists of my main doctoral project, focused on the study of AML. The second and third sections are devoted to two collaborations we were involved in. On the one hand, we contributed to the study of the emergence of metastases in breast cancer and on the other hand, the preservation of regulatory T (Treg) cell identity in the context of inflammation.

The main methods this work relies on are H3K27ac ChIP-seq, for the identification of enhancers and super-enhancers characterizing the epigenetic landscape, and ATAC-seq, for the exploration of chromatin accessibility and the prediction of key transcription factors (TFs) playing a role in the regulation of the differential enhancers and super-enhancers (SEs).

1.5.1 AML: emergence of therapy resistance in relapse

With this project we aimed in first place to explore the regulatory landscape of samples derived from AML patients at both stages, primary diagnosis and relapse, to be able to identify epigenetic changes between both disease phases. The generated data were subsequently analyzed to determine resistance mechanisms, which could ultimately serve as input for developing new therapeutic strategies.

We divided this project into three main sections, namely:

1. Exploratory analysis of the enhancer and super-enhancer landscape, as well as chromatin accessibility patterns in primary diagnosis and relapsed AML patient-derived samples, aiming to determine significant differences between both conditions.
2. Exploratory analysis of the regulatory changes induced in tumor cells during the application of the standard AML therapy using the patient-derived-xenograft (PDX) mouse model.
3. Characterization of AML cell lines regarding their epigenetic landscape as *in vitro* model, serving as future resource for the validation of candidate genes found in the previously described steps to make a difference between primary diagnosis and relapse

1.5.2 Breast cancer and emergence of metastases

For this project, I collaborated with investigators from the Institute of Stem Cell Research (Helmholtz Center Munich). Our main goal was to understand the mechanisms behind the

tumor outgrowth process in breast cancer. With this purpose, we worked with cells from patient biopsies xenotransplanted into immunocompromised mice but also with single-cell clones in culture.

In concrete, the main goal of my collaboration was to identify fundamental differences in terms of the chromatin opening patterns between subpopulations of single-cell clones sensitive or resistant to the epithelial-to-mesenchymal transition (EMT) upon stimulation.

1.5.3 Maintenance of Treg cell identity during inflammation

In our second collaborative project, carried out in the Experimental Neuroimmunology department in Klinikum rechts der Isar (Technical University of Munich) our main goal was to unravel the mechanisms preventing the loss of Treg cell identity in the context of inflammation.

Particularly, my main contribution consisted of a series of ATAC-seq experiments aiming to understand the transcriptional mechanisms by which the factor Blimp1 controlled CNS Treg cells in an inflamed environment.

Chapter 2

Results

2.1 AML: epigenetic mechanisms involved in relapse and therapy resistance

As already stated, AML is a disorder presenting a vast heterogeneity regarding genetic lesions and affecting mostly relatively elderly people. The older the patients are at diagnosis, the worse the outcome is and the higher the probability for the patient to relapse. Relapses represent a major challenge in the treatment of AML, since tumor cells at this stage frequently present resistance to the therapy. This, together with the fact that the standard therapy cannot always be applied to elderly patients due to its strong side effects, reflects the current need for developing new and more efficient treatments.

To identify and characterize relevant regulatory differences between the AML stages of primary diagnosis and relapse, AML cell lines as well as AML patient-derived samples were processed. In addition, the effects of the standard therapy on the chromatin status *in vivo* were explored using the PDX mouse model. The experiments performed as well as the derived results are presented below.

2.1.1 Characterization of AML cell lines as *in vitro* model

In first term, we aimed to assess the chromatin status and to characterize the enhancer and super-enhancer landscape of several AML cell lines (Table 4.1). To the 18 available leukemic cell lines, provided by AG Spiekermann [1], I applied both assay for transposase-accessible chromatin followed by sequencing (ATAC-seq) and chromatin immunoprecipitation of the histone modification H3K27ac followed by sequencing (H3K27ac ChIP-seq). Data derived from this first round of experiments would serve as guideline for later selection of cell line models reflecting specific regulatory landscapes to validate potential key genes found to be involved in relapse-development and/or therapeutic resistance mechanisms in patient-derived samples.

ATAC-seq and H3K27ac ChIP-seq derived peaks were first identified for each cell line and used as input for a Principal Component Analysis (PCA). For both experiments the coverage was restricted to enhancer regions, and promoter regions were filtered out. No clear clusters could be identified for the analyzed cell lines based on their FAB subtype (Figures 2.1.b and 2.1.e), evidencing a high regulatory heterogeneity among them. Nevertheless, attending to

[1] Medizinische Klinik und Poliklinik III, Ludwig-Maximilians-Universität, Munich

the risk group each processed AML cell line is associated with, according to their cytogenetic and molecular profile (Estey, 2013; Kansal, 2016), cell lines could be slightly stratified in two major clusters corresponding to intermediate- and adverse-risk groups (Figures 2.1.c and 2.1.f). These two clusters, although not strongly separated, were more clearly differentiated along the PC1 axis in the case of the H3K27ac ChIP-seq-derived data (Figure 2.1.f).

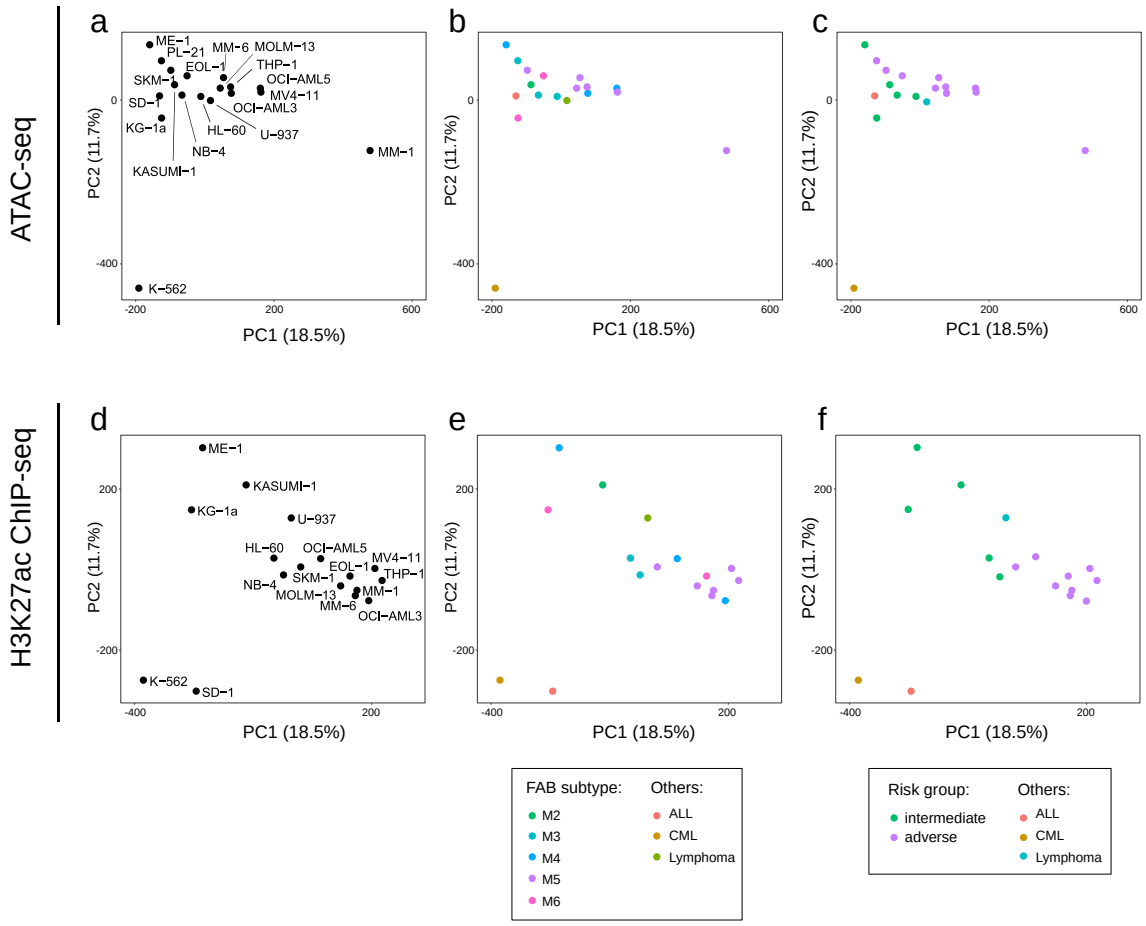

FIGURE 2.1: PCA of ATAC coverage (a, b, c) and H3K27ac coverage in enhancer regions (c, d, e) of all processed AML cell lines. Color scale in b and e indicates the AML subtype each cell line corresponds to, according to the FAB classification (Bennet et al., 1985). Color scale in c and f indicates the prognosis and outcome risk, intermediate or adverse, each cell line is associated with, according to their cytogenetic and molecular profile. Cell lines K-562, SD-1 and U-937, which are chronic myeloid leukemia (CML), the first one, and other types of lymphoid leukemias, the second and third ones, were also included. Each dot represents one processed cell line, labeled with the corresponding name in a and d.

The regulatory heterogeneity among AML cell lines can also be observed in particular loci important for the AML condition, as is the case for both *RARA* and *RXRA* loci (Figure 2.2). Both retinoic acid receptors (RARs) and retinoic X receptors (RXRs) have been documented

to be relevant in certain AML subtypes as druggable therapeutic targets (McKeown et al., 2017; Martino et al., 2019; Daramola et al., 2019).

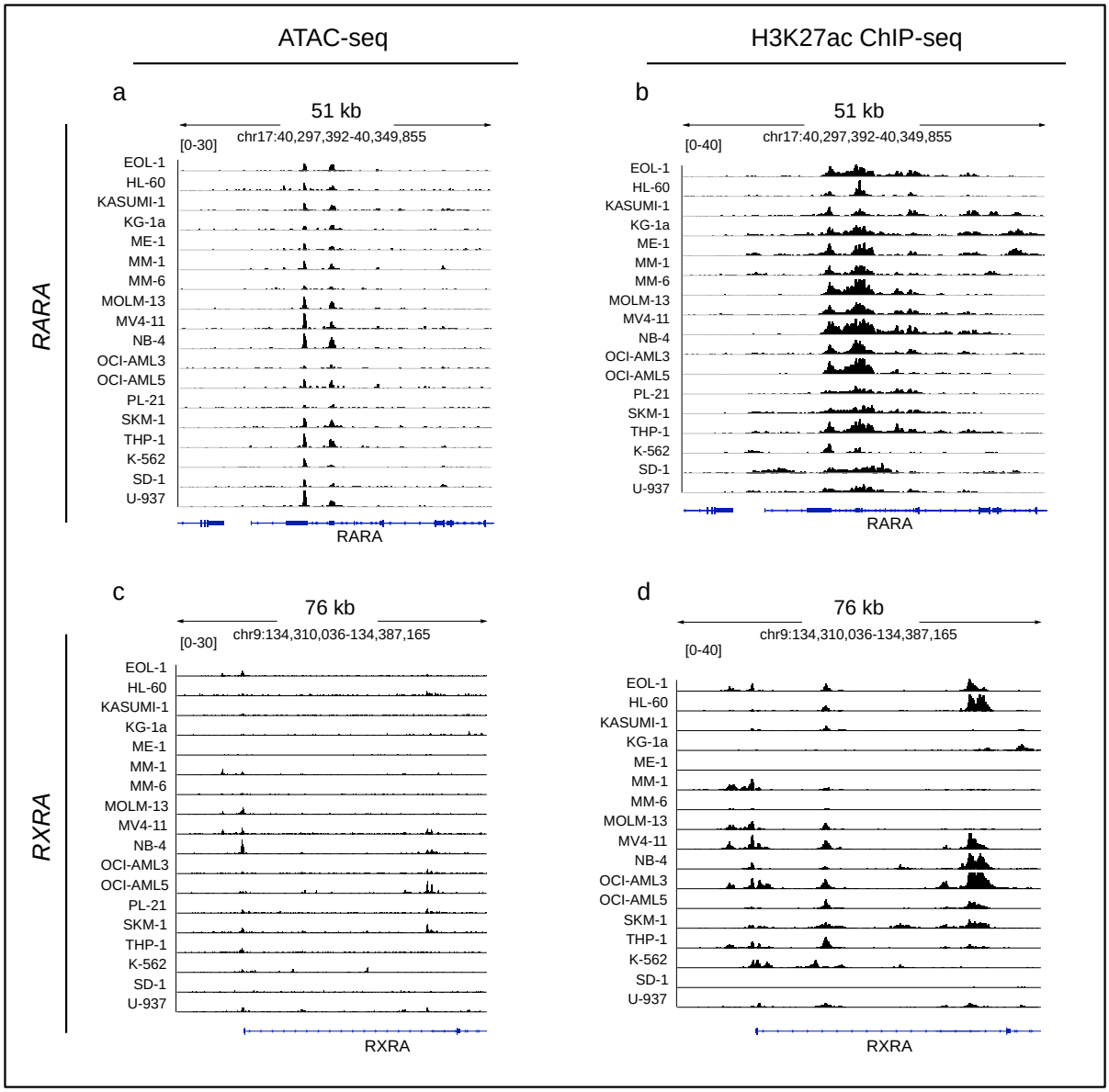

FIGURE 2.2: Example IGV genomic regions around the *RARA* locus (a, b) and the *RXRA* locus (c, d) for all ATAC-seq (left) and H3K27ac ChIP-seq (right) processed AML cell lines. Each track corresponds to one cell line.

In this section, 18 leukemic cell lines (AML principally, were characterized attending to their chromatin status and enhancer landscape. These cell lines display a broad regulatory heterogeneity, as it is distinctive for AML, serving as well-established *in vitro* model for future investigations regarding the validation of candidate genes found to be relevant for therapy resistance or emergence of relapses in AML patient-derived samples.

2.1.2 PDX samples as model for studying the effect of the standard therapy *in vivo*

Patient-derived xenograft (PDX) models have been shown over the past decades to be an invaluable help to better understand the biology and therapy of human diseases *in vivo* (Schultz et al., 2007). Particularly, xenograft mouse models represent essential tools in cancer research (Sánchez et al., 2009; Malaise et al., 2011) and are extremely important to assess the effects of drugs when developing new therapeutic strategies (Ehrhardt et al., 2012; Alves et al., 2012), since primary tumor cells far outperform the capacity of established cell lines for modeling the complexity of human disorders.

To study potential chromatin dynamic changes upon therapy in AML samples *in vivo*, the PDX mouse model was used. Mice xenografted with AML patient-derived cells were subjected to several cycles of the standard treatment and tumor samples gained at three different time points were processed for ATAC-seq. With this strategy, we aimed to investigate regulatory shifts in terms of chromatin accessibility caused by the therapy pressure.

All PDX samples were generated and provided by AG Jeremias [2]. Severe immunocompromised mice were injected with patients' bone marrow-derived AML cells, genetically engineered to allow luciferase expression for *in vivo* monitoring through bioluminescence imaging (Terziyska et al., 2012; Vick et al., 2015; Figure 2.3.b). Concretely, all samples generated in the context of this study originated from AML tumor cells isolated from a patient at the time of first relapse, eleven months after initial diagnosis. The treatment design consisted in exposing xenotransplanted mice to several rounds of the equivalent standard "*7+3*" therapy, conveniently adjusted. Leukemic burden, monitored by bioluminescence imaging (BLI), begins to descend with treatment start. After some days of remission, a minimal amount of residual tumor cells of $\sim 5 \times 10^4$ were allowed to regrow up to $\sim 5 \times 10^7$ cells before the new cycle of therapy was applied. This strategy was repeated for a total of three treatment cycles. With each successive therapy cycle, tumor cells were losing their growing potential, as the less and less steep slope of the curve indicates (Figure 2.3.a). At key time points, tumor cells were isolated for analysis. Particularly for ATAC-seq processing, samples were collected at following time points: before treatment, as baseline control, after one cycle of treatment (C1) and after three consecutive cycles of treatment (C3).

In total, 23 484 ATAC-seq peaks were identified for all samples processed. ATAC peaks corresponding to biological replicates belonging to the same treatment time point were merged. The vast majority of ATAC peaks (over 98%) were shared by all three treatment time points, suggesting that only minor changes occurred at the chromatin level driven by the applied therapy (Figure 2.4.a). Moreover, when applying a dimensionality reduction analysis to the ATAC-seq data and considering all samples independently, no clear clustering of samples attending to the treatment time point could be observed in a 2D representation (Figure 2.4.b, left). However, the 3D plot of the data suggests that samples corresponding to the latest treatment stage (C3) may cluster separately from both, untreated samples (control) and samples at the early treatment stage (C1) (Figure 2.4.b, right). Indeed, control samples and samples after the first round of therapy (C1) could not be spatially distinguished from each other in this representation.

[2] Department of Apoptosis in Hematopoietic Stem Cells (AHS), Helmholtz Zentrum München, Munich, Germany

When focusing on differential ATAC peaks showing logFC ≥ 2 between groups, we could identify 1444 ATAC peaks (~ 6% of the total number of peaks identified) distributed in four main peak clusters representing different chromatin dynamics (Figure 2.5):

cluster 1 29 ATAC peaks; continuous chromatin closing down upon treatment.

cluster 2 86 ATAC peaks; progressive opening up of chromatin as treatment advanced.

cluster 3 7 ATAC peaks; chromatin opening up right after the first round of therapy, remaining then open.

cluster 4 1322 ATAC peaks; chromatin opening up after three cycles of treatment, i.e., at the latest therapy phase.

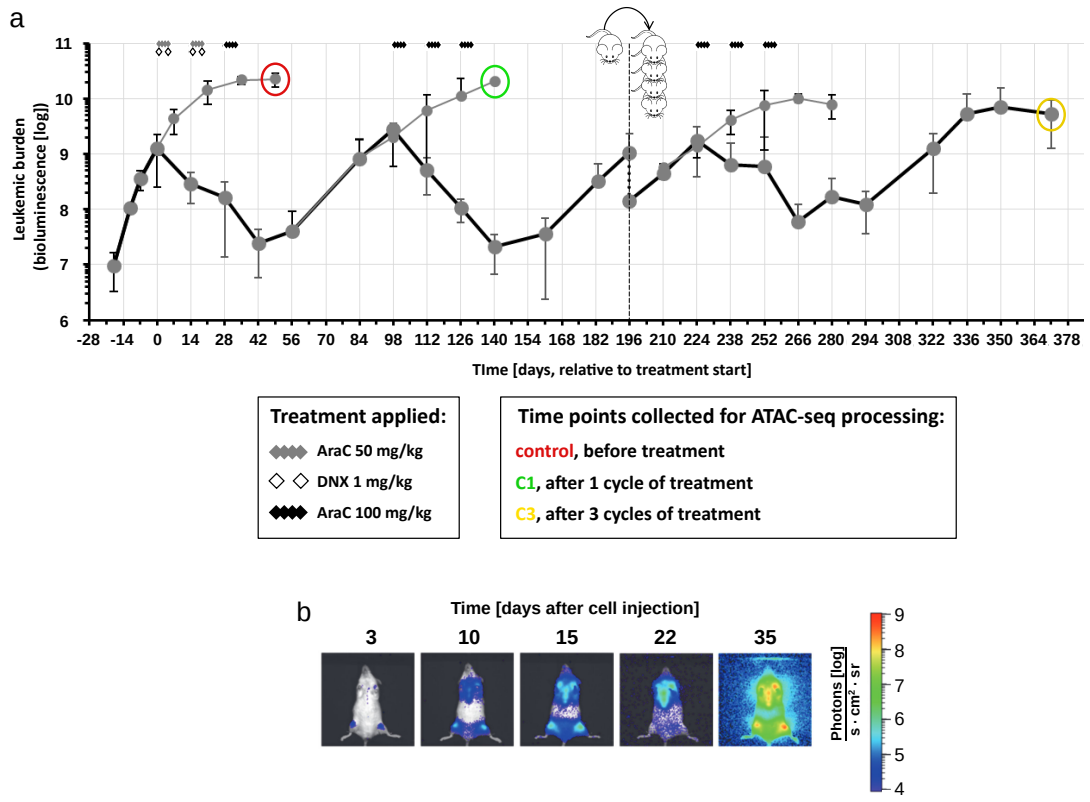

FIGURE 2.3: a, Experimental scheme describing the treatment applied to PDX mouse model. Mice were xenotransplanted with AML cells originally derived from a first relapsed patient. Mice were subjected to three cycles of therapy equivalent to the standard "*7+3*" treatment applied to AML patients. Leukemic burden of mice was monitored by bioluminescence imaging every 14 days. Cells corresponding to following time points: untreated (control), after one cycle (C1) and after three cycles (C3) of treatment, were isolated and processed for ATAC-seq. At day 196 after treatment mice died and cells had to be re-transplanted in new immunocompromised mice. Figure modified and used with permission from B. Vick (Department of Apoptosis in Hematopoietic Stem Cells (AHS), Helmholtz Zentrum München, Munich, Germany). b, Example of *in vivo* mouse monitoring by BLI after 3, 10, 15, 22 and 35 days of injection of 10^5 engineered patient-derived AML cells. Figure modified from Vick et al., 2015.

Among the four ATAC peak clusters identified, clusters 2 and 4, corresponding to genomic regions gaining accessibility at the latest treatment stage, included 1329 peaks, representing roughly 92% of the total dynamically differential ATAC peaks detected. This is in consonance with our previous observation, indicating that the most prominent effects caused by the therapy at the chromatin level appear only at later treatment stages rather than at the beginning of the therapy.

We then aimed to further investigate the proximal genes to the found differential ATAC peaks opening-up as the treatment progresses (peaks belonging to clusters 2, 3 and 4). For doing so, we set $pval < 10^{-7}$ to focus on the most relevant genes. The generated list of genes included *HCN1*, *OCA2*, *TENM2*, *IQCM*, *FAT1*, as well as the long intergenic non-protein coding RNA *LINC02541* and *LINC02063*. Among these genes, both *HCN1* and *FAT1* have been previously reported to be overexpressed in several types of cancer, including leukemia (Phan et al., 2017, Gawdat et al., 2017), and *OCA2* appears to be relevant in the clonal evolution of myelodysplastic syndrome, MDS (Silva-Coelho et al., 2017).

All in all, the ATAC-seq processing and subsequent analysis of PDX samples subjected to therapy showed in general minor changes in terms of the chromatin status. Moreover, after the first cycle of therapy, no significant changes could be detected. These minor alterations of the chromatin structure happened rather at the latest treatment stage, after completion of three cycles of therapy, and they consistently consisted in increased chromatin accessibility in certain regions. The regions presenting the most prominent changes were proximal to the loci of genes which could play a role in some leukemia subtypes or even in the pre-leukemic phase according to the literature.

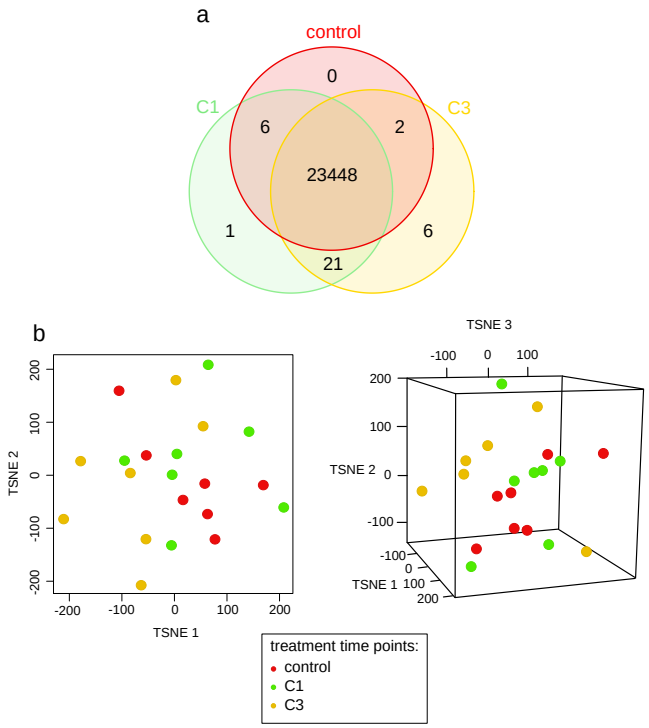

FIGURE 2.4: a, Triple Venn diagram showing the ATAC peaks detected for all three conditions explored: before treatment start (control, red), after one cycle of therapy (C1, green), and after three cycles of therapy (C3, yellow). For each condition seven biological replicates were processed. b, 2D and 3D t-SNE plots representing ATAC-processed samples. Each dot corresponds to one sample.

Although the use of PDX mice models represents an invaluable tool in the study of human diseases and drug efficacy, it needs to be considered that there are still some limitations

mainly related to the model capacity for recapitulating the original human tumor microenvironment (Ben-David et al., 2017; Murayama et al., 2019). For this reason, we aimed to study the original tumor in its native microenvironment by processing directly samples derived from AML patients, both at the stage of primary diagnosis and at relapse.

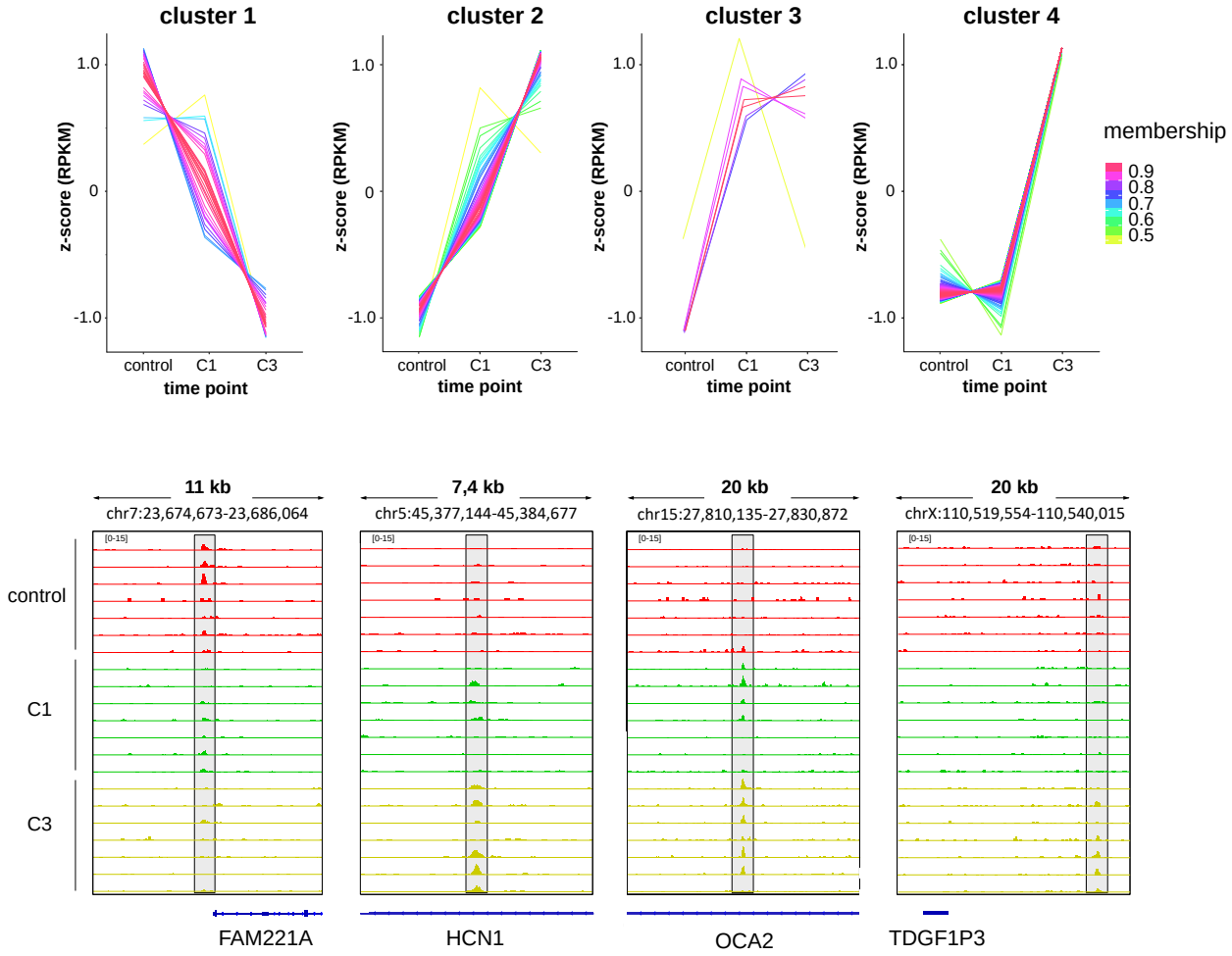

FIGURE 2.5: Upper panel, diagrams representing the normalized ATAC coverage at the different time points and corresponding to four different chromatin dynamics upon treatment: cluster 1 (29 peaks), chromatin closing down upon treatment; cluster 2 (86 peaks), chromatin opening up gradually upon treatment; cluster 3 (7 peaks), chromatin opening already after the first cycle of therapy and staying open; and cluster 4 (934 peaks), late opening chromatin. Bottom panel, IGV genomic regions around the *FAM221A*, *HCN1*, *OCA2* and *TDGF1P3* loci, exemplifying the distinct chromatin dynamics correspondent to the four ATAC peak clusters. Each track corresponds to one PDX sample: untreated (control, red), after one cycle of therapy (C1, green) and after three cycles of therapy (C3, yellow). RPKM normalization, reads per kilo base per million mapped reads. Membership color code within a particular cluster indicates to which extent a concrete ATAC peak fits into the pattern represented in that cluster.

2.1.3 Regulatory differences between primary diagnosis and relapse in AML patient-derived samples

The ultimate goal of this project is to elucidate whether regulatory alterations in AML cells contributed to therapy resistance, frequently observed in relapsed patients. To address this question and complement the previously obtained results from the study of the standard treatment over several therapy cycles in the PDX model, we intended to collect AML samples extracted from patients who experienced relapse and to process them for ATAC-seq, in order to assess chromatin accessibility, H3K27ac ChIP-seq, to explore the enhancer and super-enhancer landscape, and RNA-seq. This strategy would allow us to compare directly the regulatory landscape of paired samples (initial diagnosis and the corresponding relapse) derived from the same patient.

Indeed, the studies on paired AML patient-derived samples are relatively scarce, due to the challenges and difficulties derived from the limited availability of these precious samples.

We collected 24 AML paired samples in total, derived from 12 patients. Samples were isolated over a period of 8 years, between 2009 and 2017, with the corresponding patient informed consent for scientific use of sample material in all cases [3]. The tissue of origin was either bone marrow or peripheral blood and to separate serum and plasma from blood after extraction either heparin or EDTA were used. Patients' age was between 33 and 75 years old at age of initial diagnosis and in the cohort were included five females, four males and three cases where gender was not specified. Elapsed time between initial diagnosis and relapse was for most patients 1-2 years, although two individuals relapsed within the first year after being diagnosed and other two individuals relapsed after more than 5 years from diagnosis. The blast count, even when not known in all cases, was very sample-dependent, ranging between 8% and 99%. This fact made extremely important to develop a strategy for blast enrichment, discarding the usually co-existing normal hematopoietic cells. Samples were provided by AG Götze [4], AG Spiekermann [5], AG Subklewe [5] and AG Metzeler [5] in frozen vials stored at -80 °C since extraction (supplementary Table A.1).

We also collected 10 samples from healthy donors as control, extracted either from the bone marrow or the femoral head of donors aged between 30 and 71. Among the donors cohort we had five females, one male and three samples for which donor gender was not specified. All healthy donor samples were processed immediately after isolation, without being frozen, and were provided by AG Götze and AG Subklewe (supplementary Table A.2).

As previously mentioned, the high heterogeneity between AML patient-derived samples regarding the blast count, made essential to enrich for blasts before further processing them. To do so, we decided to target the surface marker CD33, a myeloid-associated marker highly expressed in AML cells (Bernstein, 2000). For healthy cells we targeted the surface marker CD34, characteristic of the hematopoietic progenitor cells (Civin et al., 1984). With this in mind, I carried out several tests to decide which strategy performed best for thawing samples, removing the dead cell fraction and enriching for blasts, while minimizing cell loss.

[3] As established by the *WMA* Declaration of Helsinki concerning the ethical principles for medical research involving human subjects

[4] Technical University of Munich, Klinikum rechts der Isar, Clinic and Polyclinic for Internal Medicine III, Munich, Germany

[5] Department of Medicine III, University Hospital, LMU Munich, Munich, Germany

One of the main difficulties encountered while thawing cells was the frequent presence of huge cell clumps (Figure 2.6.a). Facing the impossibility of dissociating them, these cell clumps had to be eventually removed, which limited the final number of recovered viable cells. The ultimate processing design, described in detail in Section 4.3, Chapter 4, consisted in following steps: first, samples were thawed and, when detected, cell clumps were removed; second, viability was assessed and samples were incubated with the appropriate amount of antibody (CD33 for blasts and CD34 for control cells); and finally, living cell population were sorted by either FACS or AutoMACS, discarding dead cell fraction (Figure 2.6).

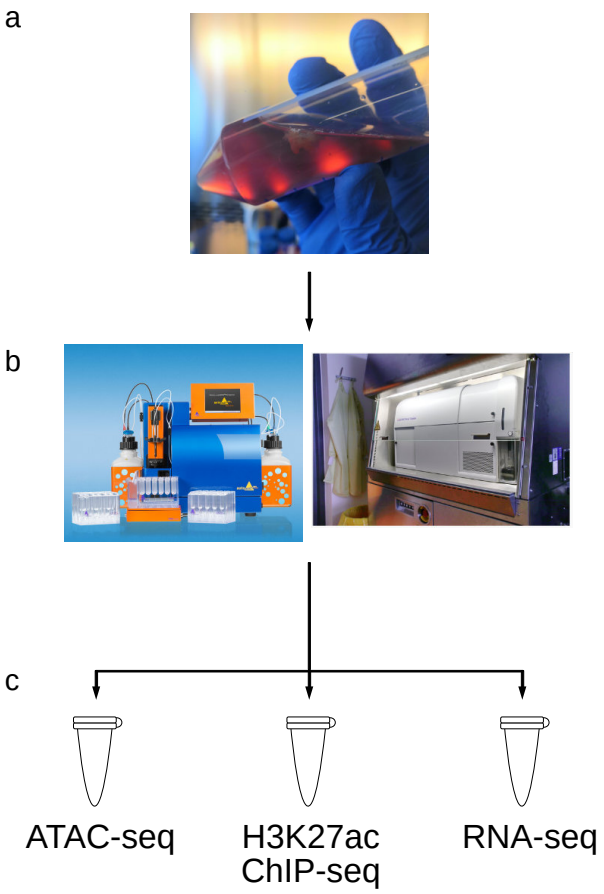

FIGURE 2.6: Scheme illustrating the AML patient-derived samples pre-processing steps: a, thawing cells, removing cell clumps when present, filtering out dead or dying cells; b, staining cells with CD33 marker (blasts) or CD34 marker (control cells) and sorting them with AutoMACS or FACS; and c, further processing cells by ATAC-, H3K27ac ChIP- and RNA-seq. RNA-seq performed in collaboration of AG Enard (Anthropology and Human Genetics, Faculty of Biology, LMU Munich). Image b, left, from *AutoMACS Pro Separator* and image b, right, from *Core Facility Flow Cytometry*, BMC, LMU.

Once AML samples were enriched for blasts, 50×10^3 cells were set aside for being immediately processed for ATAC-seq, 10^4 cells were sorted directly in the corresponding buffer and stored for subsequent RNA-seq processing and the rest of the cells were fixed as the first step for later H3K27ac ChIP-seq processing.

The final number of viable $CD33^+$ cells was different for each sample and we could not always sort enough cells for all three planned assays, since at least 300×10^3 cells were needed for performing H3K27ac ChIP-seq.

Only for six paired samples, primary diagnosis and relapse derived from the same patient, we could obtain enough tumor cells to perform all three planned assays, namely, ATAC-seq, H3K27ac ChIP-seq and RNA-seq. For the rest of the samples, due to the cell number constraints, we had only partial information gained from ATAC- and RNA-seq assays.

2.1.3.1 First AML patient paired samples-derived data analysis as proof of principle

AML patient-derived samples HS08 and HS09 (supplementary Table A.1) were processed and analyzed (not systematically, this was done at a later step, together with the rest of the

patient-derived samples) in first place as a proof of principle to validate the experimental strategy and confirm that, as expected, we could find significant regulatory differences between primary diagnosis and its corresponding relapse. We received those samples already thawed and performed CD33-based blasts enrichment with AutoMACS. Indeed, the ulterior tests carried out to optimize the pre-processing outcome demonstrated that FACS improved the number of recovered viable cells, sped up the process (essential for maximizing cell recovery rate) and reduced hands-on work. For this reason, the rest of the AML patient-derived samples, as well as the healthy donors-derived samples, were FACS sorted, instead. However, the results obtained from the processing of the aforementioned samples were promising, showing fundamental regulatory differences between the two disease stages. This is exemplified in the genomic region around the *NECTIN2* locus. In AML, NECTIN2 has been documented to be involved in the interaction between natural killer cells and leukemic cells (Pende et al., 2005). Indeed, we observed that this genomic region presents a significant increase in both the H3K27ac ChIP- and the ATAC-seq signals in the relapse sample compared to the diagnosis one (Figure 2.7).

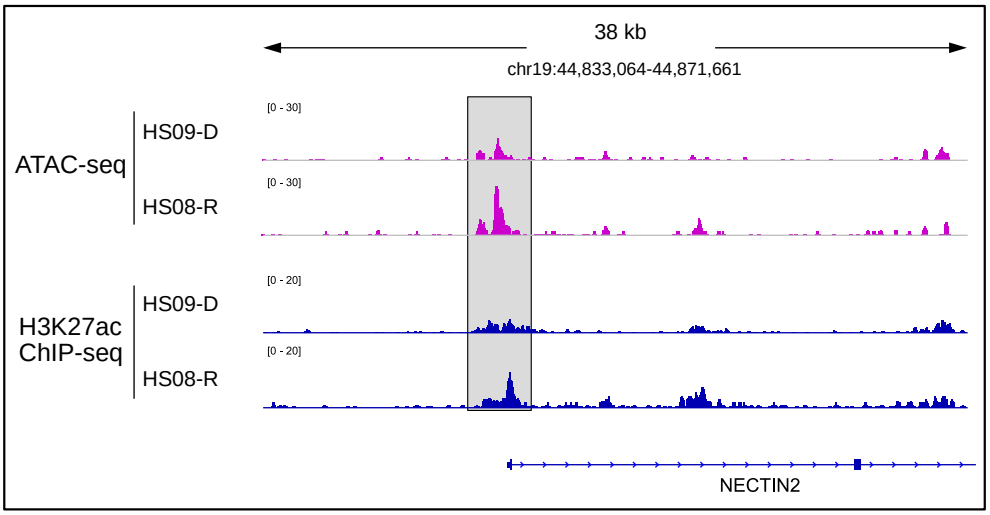

FIGURE 2.7: IGV genomic region around the *NECTIN2* locus showing the ATAC-seq tracks (magenta) and the H3K27ac ChIP-seq tracks (blue) corresponding to the first paired AML patient-derived samples, primary diagnosis (HS09) and relapse (HS08).

2.1.3.2 Protocols test: Fast-ATAC- vs. Omni-ATAC-seq

When we started processing the first AML samples (cell lines and PDX samples) and even the very first AML patient samples pair, the most suitable protocol for interrogating chromatin accessibility in human blood cells was the Fast-ATAC-seq protocol (Corces et al., 2016), an improved ATAC-seq protocol based on the previously described protocol by Buenrostro et al., 2013. The Fast-ATAC protocol was optimized for hematopoietic cells, generating improved quality data by reducing the proportion of mitochondrial reads up to 5-fold, while improving the signal-to-noise ratio.

One year later, Corces et al., 2017 published further improvements in the classical ATAC-seq protocol. The new so-called Omni-ATAC protocol allowed to generate even higher quality data using less material, as well as it made possible to process frozen material.

In order to be able to decide to which extend it was convenient to switch protocols, I compared both versions of the ATAC-seq protocol, Fast- and Omni-ATAC, side by side. For doing so, I processed 50 x 10^3 MM-1, MV4-11 and K-562 cells following both versions of the protocol (Sections 4.3.3 and 4.3.4).

For all three cell lines tested, the number of ATAC peaks detected upon analysis of data generated with the Fast-ATAC protocol was significantly lower compared to the Omni-ATAC protocol: on average, only 64% of the peaks detected by Omni-ATAC were also detected by Fast-ATAC (Figure 2.8). Moreover, the data quality, based on signal-to-noise ratio, increased when samples were processed with the Omni-ATAC protocol (Figure 2.9), meaning not only that we were able to detect more peaks, but also that the detected signal was more robust. The amount of mitochondrial reads also improved with the new protocol, with an average reduction of 58% of the lost reads for the three cell lines interrogated.

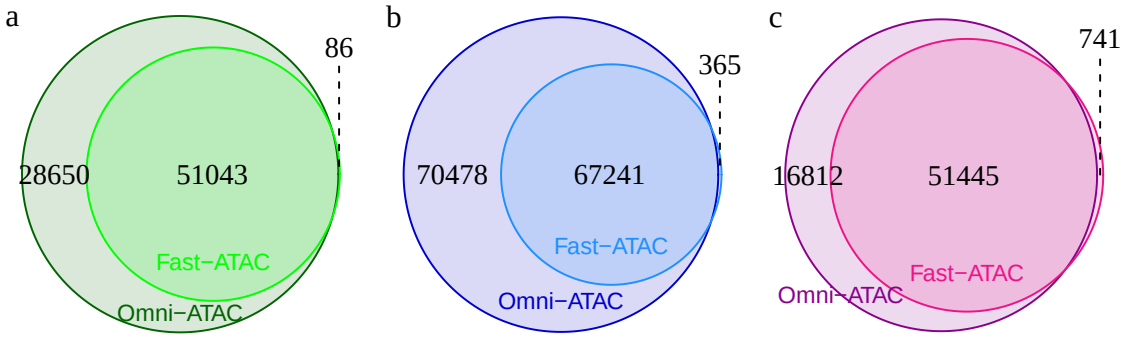

FIGURE 2.8: Venn diagrams showing the number of peaks called by HOMER for the AML cell lines a) MM-1, b) MV4-11 and c) K-562 processed following both Fast- and Omni-ATAC protocols.

Based on the previous results, we decided to definitively switch to the Omni-ATAC protocol and to discontinue the use of the previous Fast-ATAC protocol version. Additionally, we checked the impact of some other parameters on the data quality to optimize the new protocol in our experimental context. With this purpose, we tested the effect in the outcome of varying three different parameters, namely: amount of Tn5 enzyme used for transposition reaction, cell number used as input and commercial TD buffer vs. in-house made TD buffer (Appendix B).

2.1.3.3 AML patient-derived samples: no clear clustering patterns based on chromatin accessibility

Since ATAC-seq data could be generated for most of the patient- and healthy donor-derived samples, our first goal was to interrogate these data regarding potential common chromatin accessibility patterns among subsets of samples.

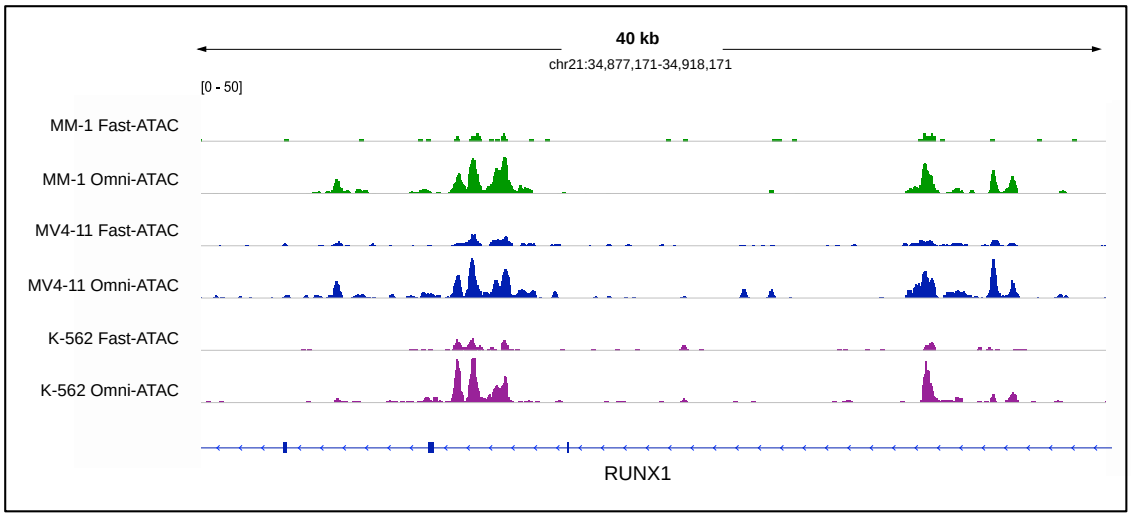

FIGURE 2.9: IGV region around the *RUNX1* locus showing the differences in the ATAC signal for the three cells lines processed for both protocols; from the top to the bottom: MM-1 Fast-ATAC, MM-1 Omni-ATAC, MV4-11 Fast-ATAC, MV4-11 Omni-ATAC, K-562 Fast-ATAC and K-562 Omni-ATAC

Using the identified ATAC peaks for all considered samples, we performed a PCA and generated the corresponding scree plot of the first fifteen PCs. The scree plot showed that already the first principal component, PC1, could explain 46.93% of the data variance (Figure 2.10.a). Nevertheless, this principal component was mainly associated with the number of ATAC peaks detected in samples (Figure 2.10.b, upper right). For this reason, we considered the six first principal components, PC1-PC6, which accounted for over 75% of the data variance. However, none of the PCs considered showed a clear clustering of samples attending to the disease status (Figure 2.10.b, lower left). Healthy donor samples did cluster together in some of the PCs but, since they all corresponded to samples with the highest number of ATAC peaks detected, a potential bias could not be discarded.

When repeating the PCA just for patient-derived samples, i.e., not considering healthy donor samples, the results did not differ substantially: the first PC explained roughly 35% of the data variance and it was mainly linked to the number of detected ATAC peaks (Figure 2.11). However, based on disease status, no well-defined clusters of samples were distinguishable along PC1-PC6. When displaying the patient affiliation, though, it could be recognized that in some cases, samples isolated from the same patient tended to group close to each other (Figure 2.12). This can be best seen along principal components PC2-PC3, where the masking effect of the number of ATAC peaks detected associated with PC1 is not present (Figure 2.13). These results strongly suggest that the origin of the samples (which patient samples were derived from) may be more defining than the disease status itself (primary diagnosis or relapse), highlighting the heterogeneity of the AML patient-derived samples in terms of the chromatin status. Moreover, in that case, single-patient-based analyses could potentially be more meaningful than the analysis of all samples considered altogether.

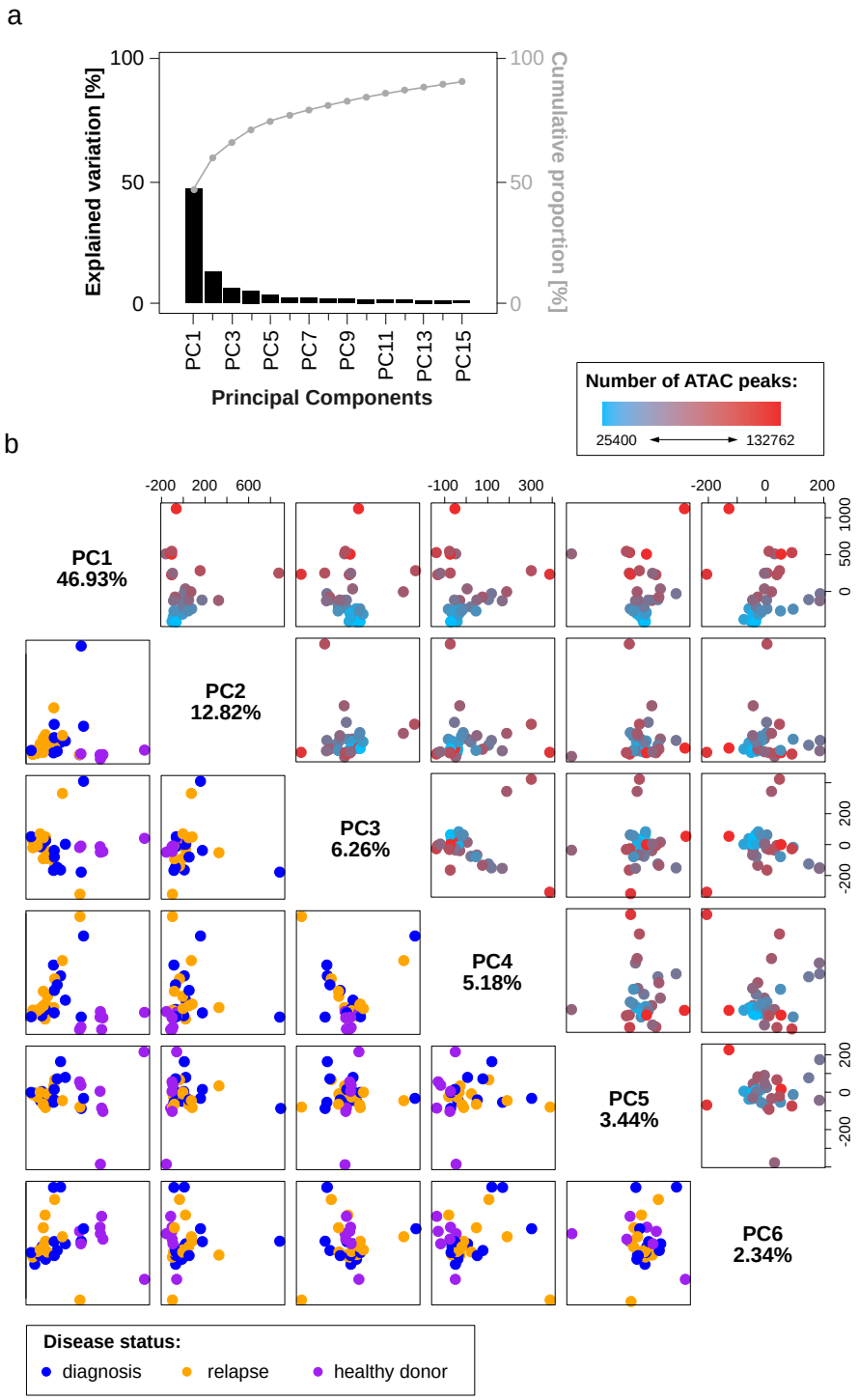

FIGURE 2.10: a, Scree plot showing the variance by each PC (black bars) and the cumulative proportion of total variance (gray line) of the first fifteen principal components based on ATAC data of patient-derived samples and healthy donor-derived samples. The first six PCs explain already >75% of data variance. b, PCA plots with the first six PCs. Color code in the top right plots is based in the number of ATAC peaks detected per sample, ranging from 25400 to 132762 peaks (light blue to red), and in the bottom left plots in the disease status: diagnosis (blue), relapse (yellow) and healthy donor (purple). Each dot corresponds to one sample.

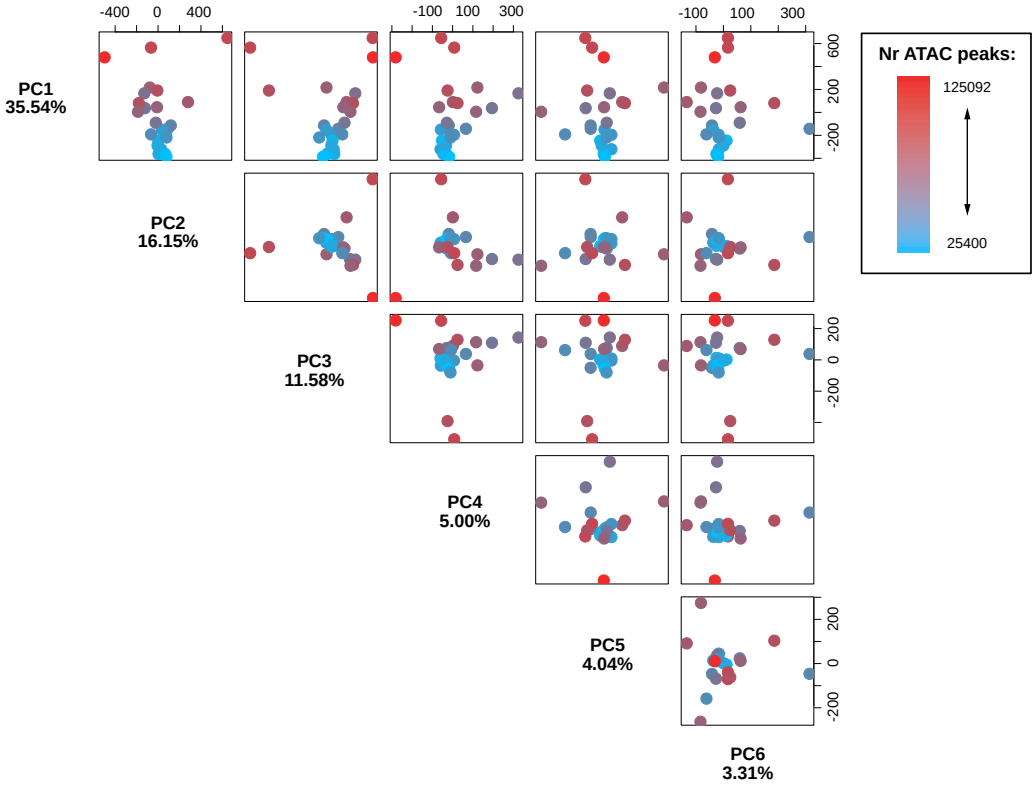

FIGURE 2.11: PCA based on ATAC data from patient-derived samples only. PCA plots considering the first six principal components are shown. Color code indicates the number of ATAC peaks detected for each sample. Each dot represents one sample.

2.1.3.4 Patient-based integrative analysis

In the light of the previous results, we decided to change the initial analysis strategy and, instead of interrogating samples altogether, we re-analyzed samples in a patient-by-patient basis. The advantage of such an approach lies in the fact that the regulatory heterogeneity derived from the origin of the samples would be ignored and the differences due to the disease status (primary diagnosis/relapse) would be prioritized. Common disease status-specific features among patients could be evaluated afterwards.

To start with, we aimed to identify regulatory differences between both disease stages for each individual patient based on the combination of ATAC-seq data and RNA-seq data (available for the major part of the patient cohort). Our goal was to predict transcription factor binding for each individual patient based on both, differential ATAC peaks found in diagnosis which are down or lost in relapse and *vice versa*, i.e., differential ATAC peaks going up in relapse compared to the diagnosis stage. Moreover, we performed gene set enrichment analysis on proximal genes to those differential regions for diagnosis or relapse. The detailed analysis description can be consulted in Section 4.4.3, Chapter 4.

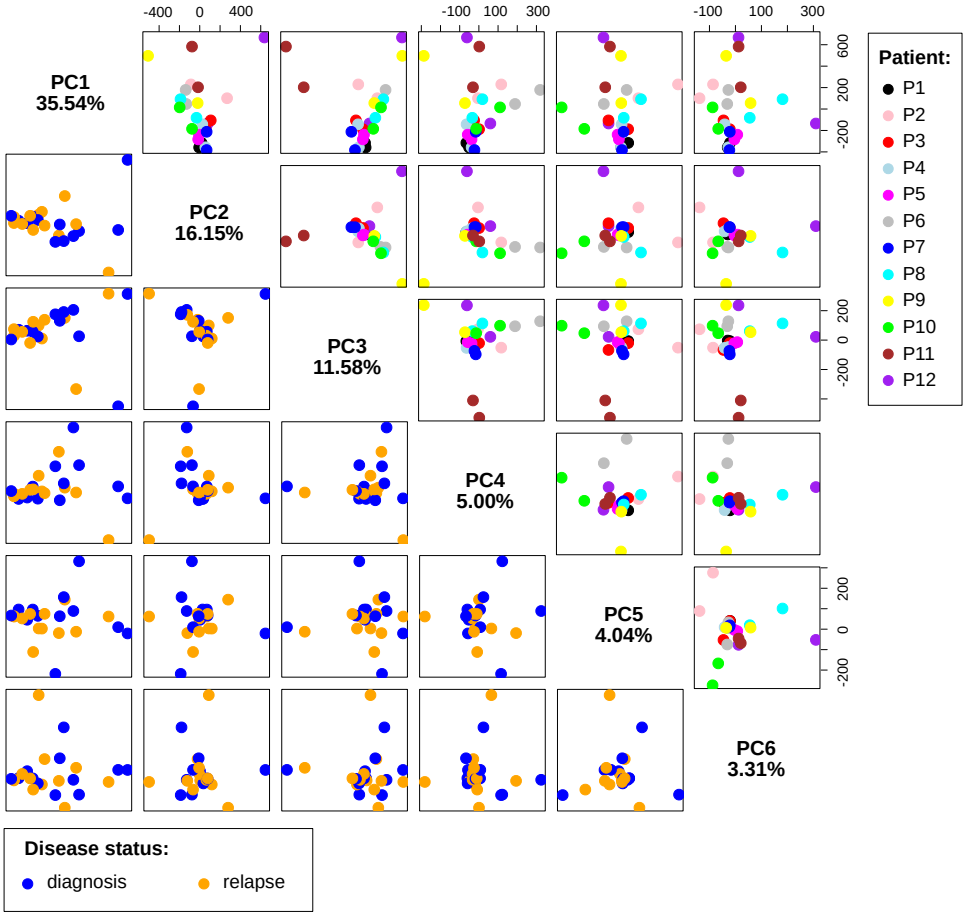

FIGURE 2.12: PCA based on ATAC data of patient-derived samples only. PCA plots corresponding to the first six PCs are shown. Color code in the top right plots depicts in the same color primary diagnosis and relapse samples originated from the same patient (each color corresponds to one patient); in the bottom left plots color code corresponds to the disease status: diagnosis (blue), relapse (yellow). Each dot represents one sample.

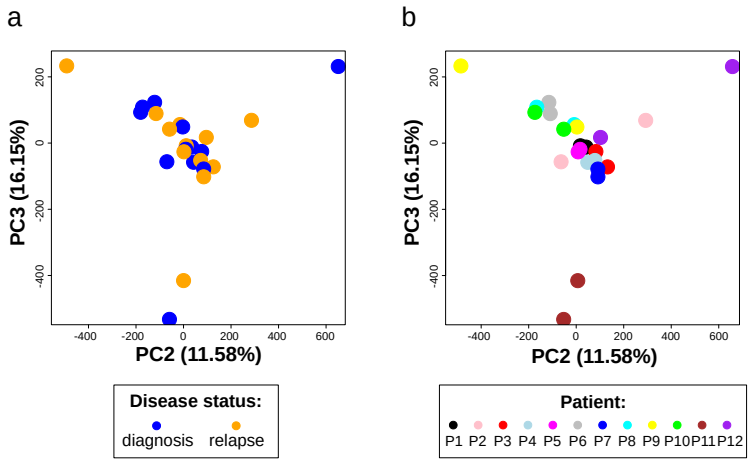

FIGURE 2.13: PCA based on ATAC-seq data showing the second and third PCs of patient samples only. a, Color code based on disease status: primary diagnosis (blue) and relapse (orange). b, Color code shows primary diagnosis and relapse samples originated from the same patient in the same color (each color corresponds to one patient). Each dot represents one sample.

Given the regulatory heterogeneity already observed among individuals, the results obtained highly varied from patient to patient. As example, for patient 2 (P2), pathways related to the immune response were found overrepresented at primary diagnosis (Figure 2.14.a). During relapse, in contrast, other pathways became more relevant: among others, activation of JUN kinase appeared within the top differential pathways for this disease stage and also within the top differential motifs predicted (Figure 2.14.b), having been this factor previously associated with therapy resistance in AML (Cripe et al., 2002). Among the differential ATAC regions associated with differentially expressed genes in diagnosis and relapse for this specific patient, the *STAT4* and the *RARA* loci, respectively, are highlighted (Figure 2.14.c). *RARA* has been shown multiple times to play a role in AML outcome (Daramola et al., 2019; Martino et al., 2019), being even a *druggable* target (McKeown et al., 2017) and the STAT transcription factor family has also been documented to be activated during leukemogenesis (Spiekermann et al., 2001; Dorritie et al., 2013). Regarding TF activity, the GATA family, which has been reported to be related to hematological disorders (Gao et al., 2015; Crispino et al., 2017), was actually observed to be significantly more active at the relapse stage (Figures 2.14.d and e). Nevertheless, even if the results showed fundamental regulatory differences between both disease stages, we found anew a remarkable heterogeneity among patients, which made it difficult to draw global conclusions (Appendix A, Section A.4 includes additional information corresponding to patients P3-P12).

A subset of patient-derived samples including patients P7 and P9-P12 could additionally be processed for H3K27ac ChIP-seq. For this subset of patients, we aimed to integrate all datasets, namely, ATAC-, H3K27ac ChIP- and RNA-seq, to obtain a global regulatory picture including not only the chromatin status information and the expression data, but also information regarding the enhancer and SE landscape. First, enhancers and SEs were identified and ranked using the algorithm ROSE2 (Lovén et al., 2013; Whyte et al., 2013; Lin et al., 2016), based on H3K27ac ChIP-seq signal, as described in Section 4.4.2.2, Chapter 4. This resulted in an average of 1392 SEs per sample for the considered subgroup of patients. Next, for integrating the ATAC-seq and the H3K27ac ChIP-seq generated data, the coltron python package (Federation et al., 2018), conceived to build transcriptional regulatory networks, was applied with default parameters. As input, the previously generated list of ranked enhancers, as well as the aligned H3K27ac ChIP- and ATAC-seq reads, were used. The idea was to determine the network of active TFs based on SEs-associated genes and ATAC peaks within those regions, and identify common patterns for the same disease status among different patients. This part of the analysis included also samples corresponding to patient P1.

All identified TFs were then merged and, for each patient, the differential factors (only present at primary diagnosis but not at relapse, or *vice versa*) were selected and ranked (Section 4.4.3.2, Chapter 4). Among the top 6 differential TFs found for each patient, two factors were shared by at least two patients at diagnosis, GFI1 and ETS2, and three factors were common to at least two patients at relapse, namely, FOXO3, RXRB and KBTBD6. Interestingly, both TFs FOXO3 and ETS2 have been previously documented to be linked to a poorer prognosis in AML (Ge et al., 2007; Santamaría et al., 2009; Fu et al., 2017) and GFI1 has been shown to influence the progression of AML in a dose-dependent manner (L. Botezatu et al., 2016).

Nevertheless, no TFs were found in common either at diagnosis or relapse for the entire subgroup of patients considered. At this point, RNA-seq data were used to check the expression of the identified differential TFs, being this analysis extended to the complete patient

cohort. TFs found to be differentially active for diagnosis in the subgroup of patients analyzed, ETS2 and GFI1, were also overexpressed in average in diagnosis compared to relapse, when considering the entire patient cohort (Figure 2.15). Nevertheless, no significant overexpression in relapse compared to diagnosis could be observed for the TFs previously identified to be differentially linked to relapse, not even for all the patients in the subgroup for which the transcriptional regulatory networks analysis was conducted. Post-translational mechanisms involving these TFs may explain the differences in TFs activity found to be associated with diagnosis or relapse, even when these TFs were not found to be differentially expressed.

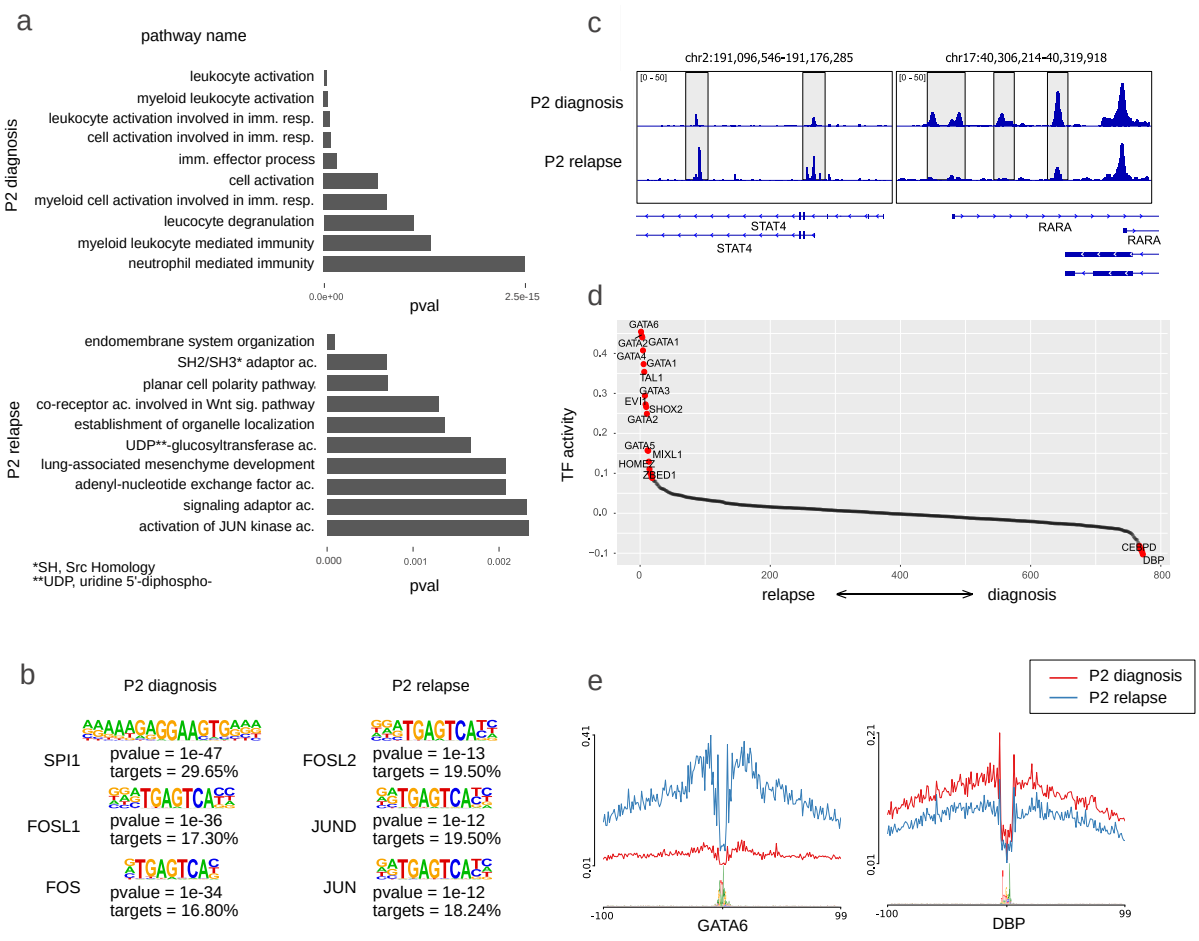

FIGURE 2.14: Integrative analysis of ATAC-seq and RNA-seq data for patient 2 (P2). a, *l2p* analysis showing the top 10 pathways associated with differential genomic features in combination with differentially expressed genes for diagnosis (top) and relapse (bottom); abbreviations: *imm. resp.*, immune response; *ac*, activity; *sig.*, signaling. b, Top three HOMER *known* predicted TF motifs for assessed differential genomic features. c, IGV tracks showing differential ATAC peaks around *STAT4* and *RARA* loci for diagnosis and relapse. d, Waterfall plot based on HINT TF differential activity. Each dot represents one TF; TFs showing significant differences in activity score (pval < 0.05) are labeled and highlighted in red. e, Detail of ATAC-seq profiles of diagnosis and relapse for motifs GATA6 and DBP.

Alternatively, we conducted a pathway analysis based on the top differentially active TFs identified for the considered subgroup of patients. This analysis strategy is detailed in Section 4.4.3.2, Chapter 4, and it basically consisted in annotating the genomic regions containing the motifs of interest, identifying the TSSs of genes within an arbitrarily defined 50 kb window and selecting only the unique genes for each disease status, diagnosis or relapse. Subsequently, these generated gene lists were filtered based on expression data and finally the relevant pathways, as well as the genes involved in those pathways, were identified.

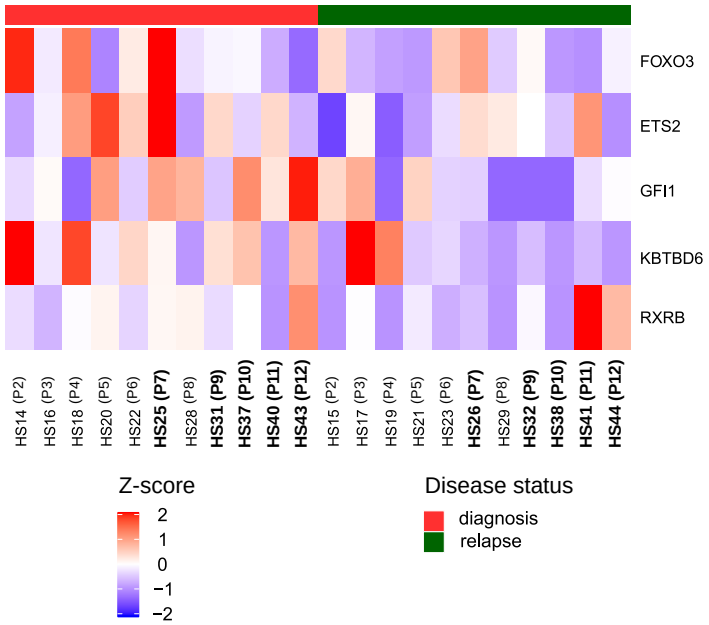

FIGURE 2.15: Heatmap showing the expression of the top differential transcription factors between primary diagnosis (*ETS2*, *GFI1*) and relapse (*KBTBD6*, *FOXO3*, *RXRB*) extended to the complete patient cohort. Differential TF activity was previously assessed by coltron based on ATAC- and H3K27ac ChIP-seq data of a patient subgroup (highlighted samples). Disease status: red, diagnosis; green, relapse. Color scale: Z-score. Note: no RNA-seq data available for patient P1 (samples HS08 and HS09).

All pathways related to apoptosis or apoptotic processes were of especial interest. These kind of pathways appeared actually overrepresented in relapse compared to primary diagnosis. Moreover, the subgroup of patients examined, considering only pathways involving more than two genes, did not show any apoptosis-related pathways at the stage of diagnosis. Nevertheless, for each considered patient-derived relapsed sample, the genes found to be involved in apoptotic processes diverged, with no genes common to the entire patient subgroup. The list of genes related to apoptotic processes and shared by at least two patients consisted of: *MUL1*, *MSH2*, *MOAP1*, *IKBKG*, *GADD45G*, *RGCC*, *CSF2RB* and *FCER1A*. Among them, the only gene shared by at least three patients was *MUL1*, whose expression has been confirmed to be associated with a poorer prognosis and a decreased disease-free survival in AML (Bester et al., 2018). The rest of the found genes appeared in at least two of the patients in the subset considered and from them, the tumor suppressor gene *MSH2* and the apoptotic modulator *MOAP1*, have also been documented to play a role in AML progression and therapy resistance (Mao et al., 2008; Law et al., 2020; Liu et al., 2021). Moreover, the gene

GADD45G (*Growth Arrest and DNA Damage Inducible Gamma*), involved in several different processes including tumor suppression (Tamura et al., 2012) and cellular stress response (Liebermann et al., 2007), has very recently been demonstrated to act as a novel tumor suppressor with important implications in AML treatment (Guo et al., 2021). Interestingly, one more of the found genes, *CSF2RB*, was newly identified in a bioinformatic analysis to correlate with worse prognosis in the overall survival in AML (Cai et al., 2021). All in all, these findings confirm our research approach as successful to determine relevant genes which could serve as potential targets for new therapeutic strategies. In this sense, the rest of the genes identified, *FCER1A*, *IKBKG* and *RGCC* (whose link with AML is, to the best of my knowledge at the time of writing the present work, still not well known), could serve as promising candidates to be further studied as possibly relevant for AML progression or resistance to therapy, using for it AML cell lines as *in vitro* model.

Remarkably, in terms of shared pathways in general, relapsed samples from different patients appeared to be significantly more similar to each other than diagnosis samples, since they presented more common pathways (Table 2.1). Taken any two samples belonging to the same disease stage, 5.9 pathways in average were found to be common in diagnosis, whereas 27.2 pathways were shared in relapse. When considering any three samples together, 0.3 pathways were in average common in diagnosis and 3.6 in relapse. No common pathways could be found in diagnosis for any four considered samples, although 29 pathways in total were still found to be shared by four out of the six relapsed samples (Table 2.3). This finding highlighted the fact that relapse samples presented more similarities than diagnosis samples did, in terms of regulatory processes. However, no pathways were found shared by the complete subset of patients, not even among relapsed samples.

The *JNK cascade* pathway was the only apoptosis-related pathway common to four out of six analyzed relapsed samples although, as already mentioned, other apoptotic pathways appeared overrepresented among relapsed samples and barely any was present in the diagnosis sample subgroup. Noteworthy, the list of all apoptosis-related pathways found in at least one of the relapsed samples included very broad terms, making not possible to infer to which extent these active pathways may contribute to therapy resistance (Table 2.2).

Nr. samples taken (out of six)	**Average number of pathways**	
	Diagnosis	Relapse
2	5.9	27.2
3	0.3	3.6
4	0	1.9
5	0	0

TABLE 2.1: Relapsed samples show more similarities to each other in terms of shared pathways than diagnosis samples do. Number of shared pathways in average for any 2, 3, 4 or 5 taken samples (diagnosis or relapse). No common pathways were found for the entire subset of samples analyzed and not even when considering any 5 samples at any disease status.

We then checked the expression of the identified most common genes (shared by at least two patients) involved in apoptotic processes in the subgroup of patient-derived relapsed samples, extending the analysis to the complete patient cohort. Nevertheless, no clear correlation could be identified between the genes of interest and the disease status (Figure 2.16.a). Next, we analyzed the expression of the top genes involved not only in apoptosis-related pathways but in the topmost shared pathways

within the subgroup of relapsed samples (Table 2.3). In fact, for this particular pathways-related gene list, we could observe an overexpression correlating with the relapse status for most patients of the cohort (Figure 2.16.b). These observations suggest the existence of certain features which may be common to all patient-derived samples at the same disease status. However, the prevailing trend deduced from the collected data evidences a general underlying (regulatory) heterogeneity among patients, regardless to the stage of the disease.

Pathway name
dimerization of procaspase-8
regulation of cysteine-type endopeptidase activity involved in apoptotic signaling pathway
regulation of extrinsic apoptotic signaling pathway
apoptotic process
lymphocyte apoptotic process
negative regulation of apoptotic signaling pathway
negative regulation of extrinsic apoptotic signaling pathway
regulation of apoptotic signaling pathway
B cell apoptotic process
death receptor signaling
activation of cysteine-type endopeptidase activity involved in apoptotic signaling pathway
positive regulation of apoptotic signaling pathway
death-inducing signaling complex
T cell apoptotic process
positive regulation of endoplasmic reticulum stress-induced intrinsic apoptotic signaling pathway
apoptosis
regulation of cysteine-type endopeptidase activity involved in apoptotic process
positive regulation of endothelial cell apoptotic process
regulation of programmed cell death
regulation of apoptotic process
apoptotic process involved in embryonic digit morphogenesis
death receptor signaling
positive regulation of fibroblast apoptotic process
regulation of hydrogen peroxide-mediated programmed cell death
protein insertion into mitochondrial membrane involved in apoptotic signaling pathway
intrinsic apoptotic signaling pathway
extrinsic apoptotic signaling pathway in absence of ligand

TABLE 2.2: List of apoptosis-related pathways identified in at least one of the relapse samples within the patients' subgroup analyzed by coltron. The found pathways are based on differentially active TFs for relapse.

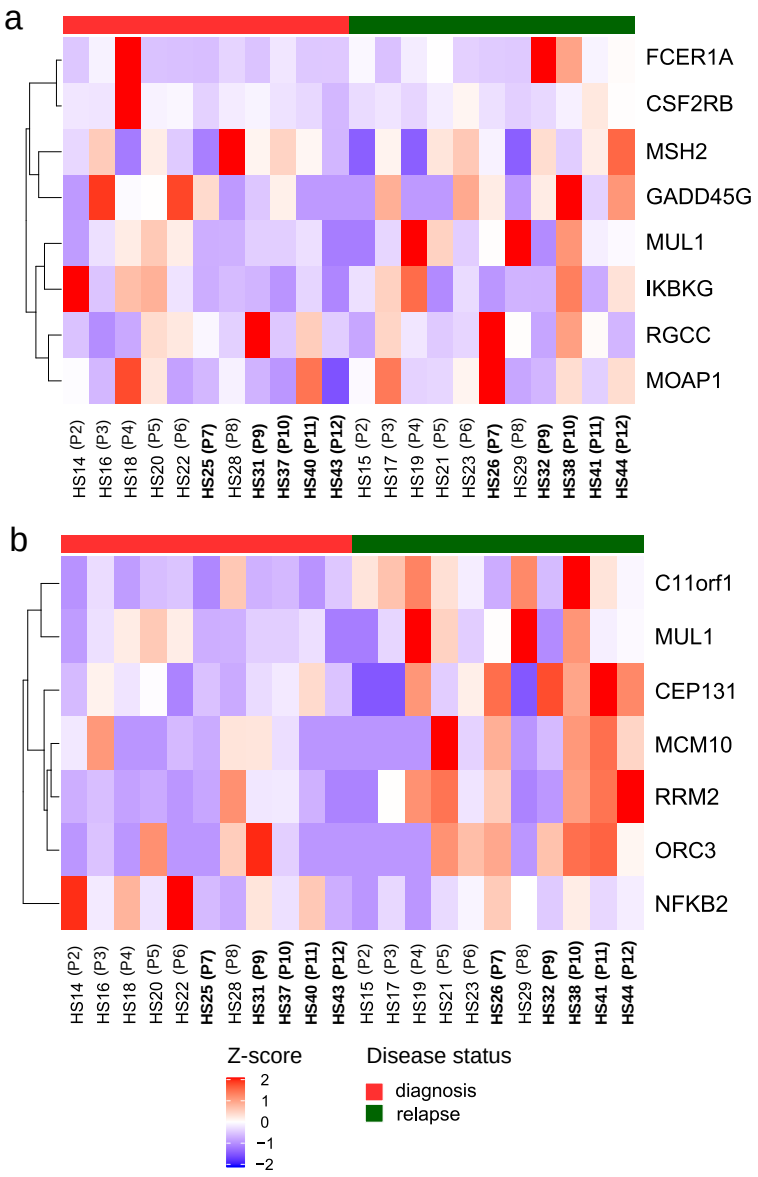

FIGURE 2.16: Heatmaps showing the expression of top genes for, a, apoptosis-related pathways and, b, most shared pathways (not only apoptosis-related) by the considered relapsed samples subset, extended to the entire patient cohort. Pathway analysis was conducted in a patient subset (highlighted samples) for which both ATAC- and H3K27ac ChIP-seq were available. Disease status: red, diagnosis; green, relapse. Color scale: Z-score. Note: no RNA-seq data available for patient P1 (samples HS08 and HS09).

We also analyzed the mean expression of all genes involved in the most shared pathways among the subset of relapsed samples, grouping pathways containing related terms together into one single entry and including the pathway *apoptosis* for comparison. For the considered patient subgroup there is a clear correlation between the disease status and the mean expression of the genes involved in the selected pathways (Figure 2.17, left). However, when extending this analysis to the entire patient cohort, the result differs, making evident that such a correlation cannot be extended to the rest of the patients (Figure 2.17, right). This highlighted once more the intrinsic heterogeneity among different AML patients, which is itself more defining than the disease status.

This observation is in agreement with certain studies appeared during the past years, suggesting the need of developing personalized therapy approaches, especially in case of drug resistant AML, given the enormous heterogeneity displayed by patients affected by this disorder (Collignon et al., 2020; Levin et al., 2021).

Pathway name	
JNK cascade	negative regulation of cell cycle
sensory perception	G1/S transition
nuclear lumen	negative regulation of cell cycle process
nucleoplasm	regulation of cell cycle phase transition
cellular response to stress	organelle envelope
neurological system process	envelope
cell-cell signaling	organelle biogenesis and maintenance
cell periphery	signal transducer activity
plasma membrane	molecular transducer activity
G-protein coupled receptor activity	system process
detection of stimulus	single-multicellular organism process
transmembrane receptor activity	mitochondrial large ribosomal subunit
membrane-enclosed lumen	mitochondrion
organelle lumen	organelle large ribosomal subunit
intracellular organelle lumen	

TABLE 2.3: List of 29 pathways found to be common to four out of the six relapsed patient samples in the patients' subgroup analyzed by coltron. The found pathways are based on differentially active TFs for relapse. No common pathways were identified for four or more diagnosis samples.

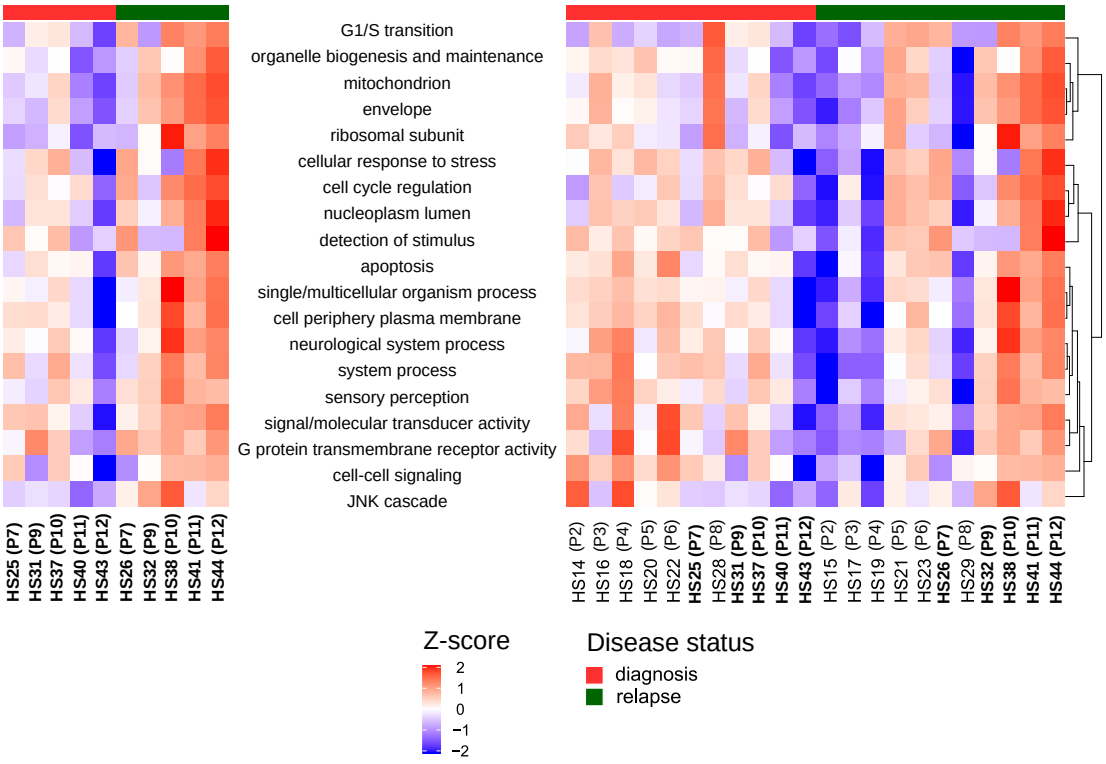

FIGURE 2.17: Heatmaps showing the mean expression of all genes involved in the previously identified relapse-relevant pathways; left, only the patient subset considered for conducting the analysis; right, analysis extended to the entire patient cohort. Disease status: red, diagnosis; green, relapse. Color scale: mean Z-score. Note: no RNA-seq data available for patient P1 (samples HS08 and HS09).

2.1.4 Discussion I

Intrinsic tumor heterogeneity of acute myeloid leukemia has already been reported multiple times in the past years (Swaninathan et al., 2018; Horibata et al., 2019; Karantanos et al., 2019). This phenomenon can actually not be reduced to inter-patient heterogeneity, but also clonal heterogeneity has been recently documented during disease progression and under therapy pressure (Morita et al., 2020). Moreover, Sandén et al., 2020 also reported the presence of multiple co-existing clones already at the diagnosis stage. Due to this vast diversity, AML have been sub-classified into several groups depending on particular features to be treated following specific therapeutic strategies (Borthakur et al., 2014; Abaza et al., 2017). Even more, novel studies have proposed the development of personalized chemotherapy as the best tool to treat AML, especially when tumor cells develop therapy resistance (Testa et al., 2021; Levin et al., 2021). A very recent review article (Kantarjian et al., 2021) compiles the major advances and novel drugs approved for the clinical practice during the past years for the treatment of different AML sub-types. Nonetheless, treatment based on cytarabine and anthracyclines, usually known as *"7+3"*, still remains the standard therapy, emphasizing the big interest and urgent need of developing new therapeutic strategies to treat AML efficiently in all its possible variants.

Our approach was based on the comparison of primary diagnosis and relapse coupled samples derived from AML patients. The patient cohort included 12 patients in total, for which we developed a processing strategy involving blasts enrichment based on the cellular surface marker CD33, characteristic of the leukemic cells. This was a crucial step, since patient-derived samples usually contain other normal hematopoietic cell populations, which would potentially distort the results. Patient-derived samples were extracted within a relatively long timespan (at the time of processing, the first isolated sample of the cohort was stored over 10 years and, in contrast, the last gained sample was stored for just 2 years). This could be one of the reasons explaining the different samples capacity to survive the thawing and sorting processes. Another aspect to consider as potentially influencing the samples quality was the human factor, since samples were provided by four different research groups, in two different hospital clinics, being presumably several people involved in the samples extractions over time. The number of frozen cells contained in received vials for each sample was also very different, ranging between 12-300 Mio frozen cells; this fact represented an additional determining factor linked to cells viability after thawing, being usually the vials with less cellular content the ones resulting in the proportionally lowest final recovery (additional information regarding the number of frozen cells per vial, as well as cell viability after thawing and final cellular recovery after sorting, can be consulted in supplementary Table A.3). As much as possible, due to the aforementioned limitations, samples were processed for ATAC-seq and H3K27ac ChIP-seq, as well as for RNA-seq, to characterize their chromatin status and determine their enhancer and super-enhancer landscape [6].

Through the course of this work, we were able to examine the characteristic AML heterogeneity reflected in broad regulatory differences among the studied patient cohort, as well as between both stages analyzed, primary diagnosis and relapse (Figures 2.10, 2.12). The high heterogeneity observed between patients induced us indeed to adopt a patient-by-patient analysis approach, comparing both disease stages for the same patient before trying to draw

[6] Not all samples could be processed for all three planed assays, since the cell number was very variable and insufficient in some cases.

more general conclusions. In this sense, we carried out a data integration resulting in the major regulatory differences between diagnosis and relapse for all patients individually (Figure 2.14 and Figures in Appendix A). Based on these results, we aimed to identify common features in terms of regulatory networks and shared pathways among samples at the same disease stage.

Our rationale was to identify in first term enhancers and super-enhancers from the H3K27ac ChIP-seq signal and then to integrate these data with the ATAC-seq data to build transcriptional regulatory networks. In this way, we managed to identify the network of active TFs based on SEs-associated genes and ATAC peaks within these regions and, thus, determine common patterns for the same disease status among different patients, as described in Section 2.1.3.4.

We detected significantly more shared pathways among the relapsed patient-derived samples than among the group of diagnosis samples (Table 2.1). Interestingly, no apoptosis-related pathways could be identified for any of the diagnosis samples, whereas those pathways were found in almost all relapse samples analyzed. This fact may reflect the selective pressure the tumors underwent during chemotherapy. Nevertheless, apoptosis-related pathways identified included mostly broad terms such as *apoptosis signaling pathway*, *apoptotic process*, *regulation of programmed cell death* or *death-inducing signaling complex*, making extremely difficult to hypothesize whether the apoptotic processes relapsed samples seem to be commonly involved in, could be exploited to target these usually therapy-resistant tumor cells. This question should be further explored in future research.

In any case, these observations were in consonance with the results obtained in the experiment conducted on the PDX mouse model, where xenotransplanted mice were subjected to three consecutive therapy cycles, mimicking the standard AML treatment applied to patients, as described in Section 2.1.2. In this experiment, data derived from the ATAC-seq processing of samples collected at three different treatment time points suggested that the sustained therapy pressure over several cycles caused samples to converge to a distinct regulatory status, when compared to untreated samples or samples right at the beginning of the treatment, after just one round of therapy (Figures 2.4 and 2.5).

Regarding the AML patient-derived samples, when focusing on the top genes involved in the common pathways found for the relapsed samples, we found some very interesting candidates, such as *MUL1*, *MOAP1*, *IKBKG*, *FCER1A* or *RGCC*, among others. Many of them have already been documented to play a role in the progression of the disease or even in the response to therapy (Mao et al., 2008; Bester et al., 2018; Law et al., 2020). Our analysis also identified the gene *GADD45G* as relevant for relapse. Interestingly, this gene has very recently been reported to act as a novel tumor suppressor with important implications in AML treatment (Guo et al., 2021). This fact positively reinforces the validity of our approach, based on the study of the regulatory landscape to identify interesting candidate genes involved in the relapse process and the response to treatment. Indeed, other genes identified in our screening, which are so far not extensively documented in the literature, such as *FCER1A*, *IKBKG* or *RGCC*, could serve as potential candidate genes to be tested as relevant for the progression of the disease, using for it the AML cell lines *in vitro* model, established for this purpose in the course of this work (Section 2.1).

We mainly focused on apoptotic-related pathways as well as on the most common pathways in general and, within them, on the top shared genes among the samples analyzed;

however, other relevant pathways, such as pathways related to cell cycle regulation or response to stress, to name just a few, could be explored following the same strategy. The integrative analysis of data derived from ATAC-seq and H3K27ac ChIP-seq, besides using expression data as filtering criterion, constitutes a powerful proceeding for unraveling particular features which could potentially be exploited for the development of new treatment strategies in AML, but could also be applied in the contexts of other diseases.

Nevertheless, it is necessary to take into consideration some constraints affecting our approach. On the one hand, and as described in the corresponding section, we encountered difficulties regarding the availability of patient-derived samples, especially since we aimed to compare primary diagnosis and relapse coupled samples derived from the same patient. This kind of samples are scarce and, thus, difficult to obtain. That is the reason why the patient cohort we had access to comprised just 12 patients, with the limitations that such a small study cohort means for generalizing obtained results. In the future, complementing the results obtained in this work with data derived from additional AML patients would be very valuable. Another challenge we faced was related to the different qualities exhibited by the processed samples resulting in a very sample-dependent recovered number of viable cells. However, the above mentioned factors are inevitably inherent to the handling of patient samples and, although it is necessary to have all these constraints in mind, this fact is not at odds with the interest of the obtained results. Another consideration I would like to comment on, is the fact that we seemed to have a bias affecting differently the complexity of ATAC libraries derived from AML patients on the one hand, and from healthy donors on the other hand. This derived in a very different number of ATAC peaks detected for both groups of samples, making any comparison between them inconsistent. Whether this is due to intrinsic differences between samples or other reasons, remains to be elucidated.

The present work, despite all the limitations associated with the investigation of small patient cohorts, constitutes a successful pilot study laying the foundation for future research including more paired AML patient-derived samples. Additionally, as result of the performed analyses, some genes were identified which could be potentially involved in AML progression. In future experiments, those genes could be tested in the already characterized AML cell lines, knocking them down or completely deleting their expression to evaluate the outcome and its implications as a possible therapeutic approach.

2.2 Mechanisms underlying the development of metastasis in breast cancer

To explore chromatin structure patterns in terms of accessibility in the context of other malignancies, we were involved in a collaborative project working with a different cancer model, specifically breast cancer. Applying ATAC-seq, we aimed to analyze the chromatin structure associated with the tumor outgrowth process.

For this research I was collaborating with investigators from the Institute of Stem Cell Research (Helmholtz Center Munich), where metastatic breast cancer (MBC) cells from biopsies of patients were xenotransplanted into immunocompromised mice to study the mechanisms controlling the tumor outgrowth process. The results and conclusions of our work

have been already pre-published in an open access repository (Eichelberger et al., 2020) and are undergoing a peer-review process in order to be soon published.

To assess the tumor-initiating capacity (TIC) of different tumor cells phenotypes, cancer cells derived from liquid biopsies of metastatic breast cancer patients were sorted attending to the expression level of the surface marker EPCAM (Figure 2.18.a) and xenotransplanted into highly immunodeficient recipient mice. From both EPCAM subtypes, EPCAM^{high} cells demonstrated in all cases to have a greater capacity to initiate tumors compared to EPCAM^{low} cells, regardless to the initial number of tumor cells injected: in total, EPCAM^{high} cells showed a 30 times higher TIC frequency than EPCAM^{low} cells. Moreover, only EPCAM^{high} cells were found to be able to generate distal metastases in the lung (Figure 2.18.b).

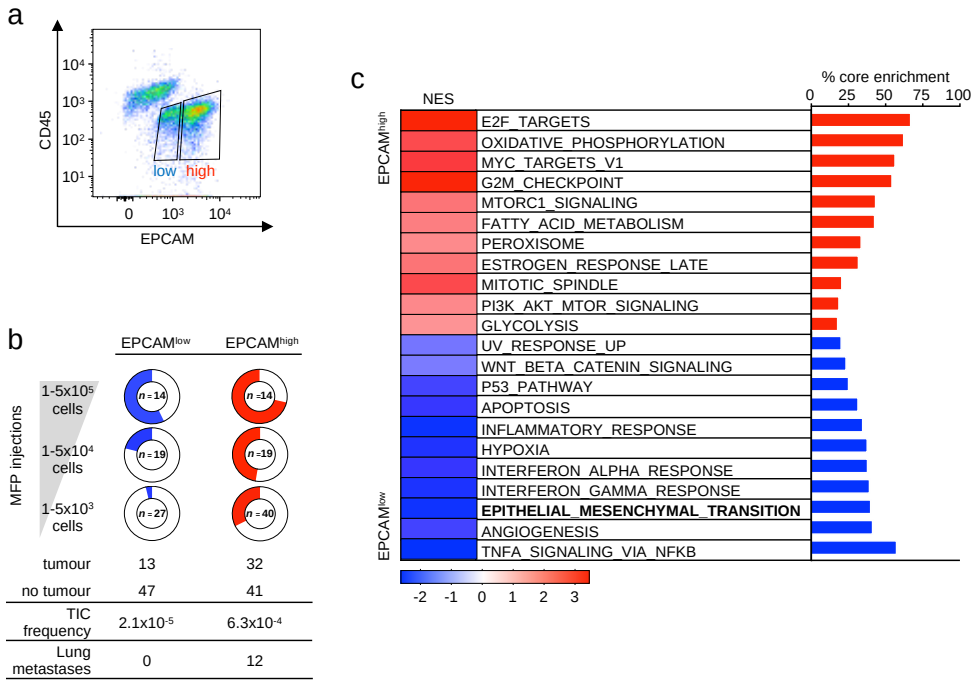

FIGURE 2.18: a, FACS plot showing two distinct cancer cell groups based on expression of the surface marker EPCAM: EPCAM^{high} and EPCAM^{low}; b, Tumor-initiating capacity and frequency of observed distal lung metastases for EPCAM^{high} and EPCAM^{low} cells when xenotransplanted into immunodeficient mice. Three different amount of initial EPCAM cells were injected into recipient mice. n, number of mice injected in each case; c, GSEA of RNA-seq data from EPCAM^{high} and EPCAM^{low} cells. Signatures were ranked based on % of core enrichment. Figures modified from Eichelberger et al., 2020.

Additionally, when checking the estrogen receptor (ER) status of the collected biopsies from patients, it was observed that, although EPCAM^{high} cells presented a higher TIC frequency for both ER^{+} and ER^{-} cell subpopulations compared to EPCAM^{low} cells, particularly for ER^{-}-derived biopsies, both EPCAM^{high} and EPCAM^{low} cells generated tumors at a relatively high rate, in consonance with the previous observation relating a greater phenotypic plasticity to ER^{-} biopsies (Yu et al., 2013).

The mutational profiles of EPCAM^{high} and EPCAM^{low} cell populations were also checked, although no significant mutational divergences were detected. This indicated that the observed differences in the capacity of the EPCAM subpopulations to generate tumors must be due to epigenetic rather than to genetic differences. RNA-seq was then performed in both EPCAM^{high} and EPCAM^{low} cells. Interestingly, *epithelial-to-mesenchymal transition* (EMT) appeared as one of the most enriched pathways for EPCAM^{low} cells in Gene Set Enrichment Analysis (GSEA). In contrast, EPCAM^{high} cells were found to be enriched for pathways related to cell growth and cell proliferation, as E2F targets, G2M targets, mTOR signaling or MYC targets (Figure 2.18.c). These results confirmed EPCAM expression level as a key indicator for the tumor-initiating capacity of breast cancer cells and identified the EMT as one of the mechanisms responsible for the reduced TIC observed in the EPCAM^{low} cell population.

To further examine the effects of EMT in breast cancer cells, immortalized human mammary epithelial cells (HMLE cell line) were engineered to express in a conditional manner the EMT-inducing TF Twist1 (termed HMLE-Twist-ER) upon tamoxifen (TAM) activation (Schmidt et al., 2015; Casas et al., 2011). Cells were treated with TAM for 21 days continuously (+TAM 21d) and then checked for surface EPCAM expression. Most cells lost EPCAM expression after EMT Twist1-induced activation, indicating that they have transitioned to a mesenchymal phenotype. Nevertheless, a small portion of cells remained EPCAM positive, implying a resistance to EMT-induction (Figure 2.19.a). Following the Twist1 activation, cells were exposed to TAM-free conditions for 10 more days (+TAM 21d/-TAM 10d), to mimic the transitory EMT stimulus observed frequently during cancer development (Guarino et al., 2007; Trimboli et al., 2008; Gavert et al., 2008). Interestingly, EPCAM negative cells remained stably mesenchymal when the EMT stimulus disappeared, as shown by the absence of typical epithelial markers like E-cadherin, *CDH1* (Figure 2.19.b). In contrast, EPCAM positive cells were further characterized by an epithelial phenotype, indicating a certain extent of resistance to EMT-stimulus.

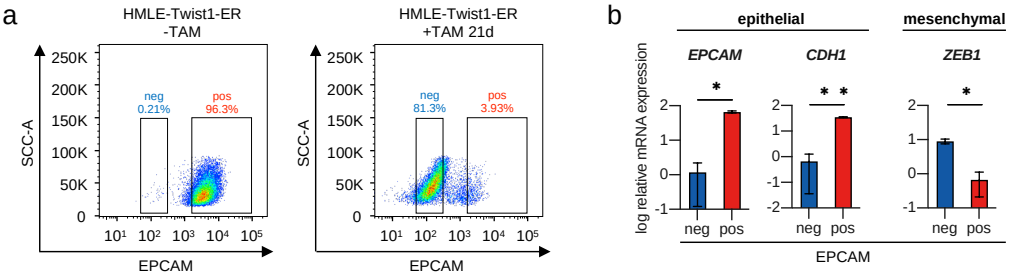

FIGURE 2.19: a, FACS plots showing EPCAM expression of HMLE-Twist1-ER cells before (-TAM) and after 21 days of continuous TAM treatment (+TAM 21d) for Twist1 activation; b, log relative mRNA expression of *EPCAM*, E-cadherin (*CDH1*) and *ZEB1* of EPCAM positive (pos) and negative (neg) cells after 21 days of TAM treatment immediately followed by 10 days without TAM. Figures modified from Eichelberger et al., 2020.

To better understand the mechanisms underlying the distinct susceptibility of HMLE-Twist1-ER cells to the EMT process, single-cell clones (SCCs) were generated and exposed independently to TAM treatment. From all TAM treated-SCCs, only a fraction was observed

to undergo EMT, developing mesenchymal characteristics and losing at the same time their previous epithelial features (M-SSCs, 72% of all SCCs). It was even observed that this subpopulation retained their mesenchymal features despite discontinuation of TAM treatment, indicating that these cells undergo EMT and acquire a stable and irreversible mesenchymal phenotype. The rest SCCs were resistant to EMT and retained their epithelial attributes (E-SSCs, 28% of all SCCs). Even more, E-SCCs were observed to acquire mesenchymal traits, as upregulation of Vimentin (*VIM*), to a similar extent than M-SCCs upon TAM-induced Twist1 activation, displaying a hybrid phenotype characterized by the co-expression of both epithelial and mesenchymal markers (Figure 2.20).

We then performed ATAC-seq to have a better insight into the epigenetic mechanisms underlying the observed resistance/susceptibility of both subpopulations of breast cancer cells to undergo EMT. This was my major contribution to Eichelberger et al., 2020. For this first ATAC experiment, we processed in total 24 samples derived from HMLE-Twist1-ER SCCs, 12 E-SCCs and 12 M-SCCs, three biological replicates at four different time points: untreated as baseline condition (-TAM), after 7 days of continuous TAM treatment (+TAM 7d), after 14 days of continuous TAM treatment (+TAM 14d) and finally after 14 days of TAM treatment followed by 7 days of TAM discontinuation (+TAM 14d/-TAM 7d) to mimic a transient EMT stimulus (Figure 2.21 and Table 2.4 at the end of this section). As HMLE cells are adherent, we introduced some changes in the harvesting protocol to maximize the viability of collected cells (Chapter 4, Section 4.3.4).

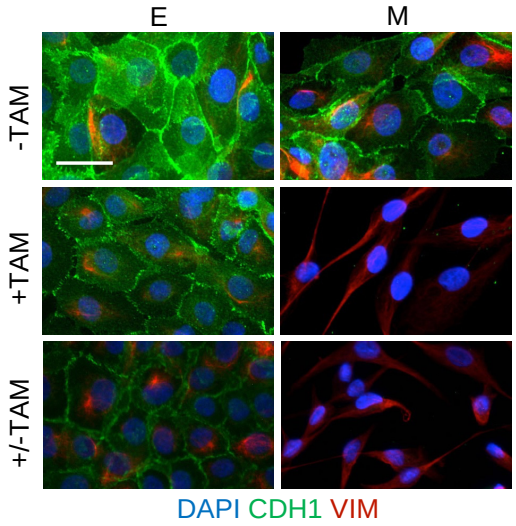

FIGURE 2.20: Immunofluorescence microscopy image of one example representative for E-SCC and M-SCC before TAM treatment (-TAM), after 15 days of continuous TAM exposure (+TAM) and after 15 days of TAM exposure followed by 9 days without TAM (+-TAM). DAPI (blue), CDH1 (green) and Vimentin (VIM, red). Scale bar equivalent to 20 µm. DAPI (blue), CDH1 (green) and Vimentin (VIM, red). Scale bar equivalent to 50 µm. Figure modified from Eichelberger et al., 2020.

The PCA of the generated ATAC data shows that, once the TAM-induced Twist1 activation was over, E-SCCs but not M-SCCs, returned to a chromatin state similar to the initial one, before the TAM treatment was initiated (Figure 2.22). Dynamic changes in the opening of the chromatin were observed around loci encoding key epithelial and mesenchymal genes. Interestingly, the epithelial genes *EPCAM*, *CDH1* and *OVOL2* display a reduction of the ATAC signal for M samples upon TAM-induced activation of Twist1, while E samples retain their chromatin status to a certain extent even after withdrawal of the TAM treatment (Figure 2.23). Moreover, E-SCCs show a temporary increased chromatin opening at the loci of mesenchymal markers *VIM* and *FN1*, which is reversed upon TAM discontinuation. For M samples, on the contrary, both *VIM* and *FN1* promoter regions show stable chromatin openness through all treatment stages (Figure 2.24).

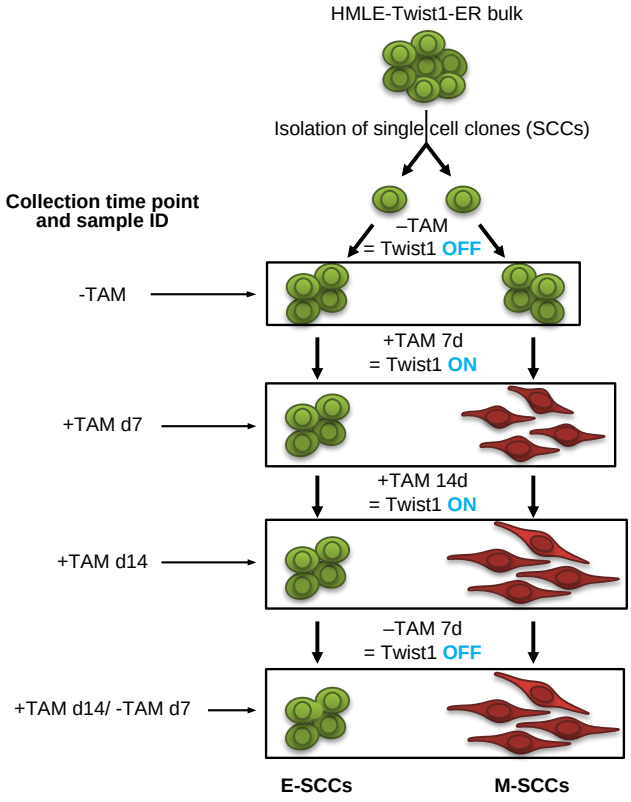

FIGURE 2.21: Overview of HLME-Twist1-ER-derived SCCs, treatments and time points used for the ATAC-seq experiment. Figure adapted from L. Eichelberger.

Together, these results confirm our previous observations, in which E-SSCs transiently reach a hybrid phenotype upon Twist1 activation, maintaining their epithelial features but also displaying mesenchymal traits, while M-SCCs remain in a stable and permanent mesenchymal state even after the EMT stimulus is over.

When further analyzing ATAC peaks, we could establish 12 peak clusters (c1-c12), grouped into three categories: A, chromatin regions not accessible previous to the Twist1 activation but becoming accessible upon TAM treatment for both E-SCCs and M-SCCs; B, chromatin regions with increased accessibility only for M-SCCs upon Twist1 activation and C, regions accessible only for E-SCCs but with reduced accessibility for M-SCCs upon TAM treatment (Figure 2.25.a). We then focused in all peak clusters included in group C (c1, c2, c5, c8, c10 and c12), i.e., genomic regions presenting a decreased chromatin accessibility during Twist1 activation only for M-SCCs but not for E-SCCs, and performed HOMER *de novo* motif analysis (Figure 2.25.b). Among the top motifs enriched in this subset of peaks, we found ZEB1, a TF known to be involved in the induction of EMT having a repressive function over epithelial genes as *EPCAM* or *CDH1* (Sánchez-Tilló et al., 2010; Vannier et al., 2013). We also found an enrichment of GRHL, a TF family described to interact with epithelial enhancers (Jacobs et al., 2018). Moreover, GRHL2 has been described to be a pioneer TF priming epithelial genes (Chen et al., 2018), as well as to play important roles in breast cancer, interacting bidirectionally with ZEB1 in a regulatory negative feedback loop (Werner et al., 2013). Therefore, both ZEB and GRHL TFs seem to be important for the process regulating the response to an EMT-inducing stimulus. We then performed the same analysis focusing this time on the clusters of ATAC peaks included in group A (c3, c4), namely, chromatin regions gaining

accessibility for both E-SCCs and M-SCCs upon TAM treatment (Figure 2.25.c). Among the top motifs enriched for this subset of peaks, we found the consensus binding sequence for Twist. These results, which are in consonance with the RNA-seq data collected from the same samples, confirmed that Twist1 is active to a similar extent in both E-SCCs and M-SCCs, suggesting that the observed differences regarding the resistance or susceptibility of E and M samples to undergo EMT must rather be due to Twist1-downstream regulation, where ZEB1 and GRHL1/2 come into play.

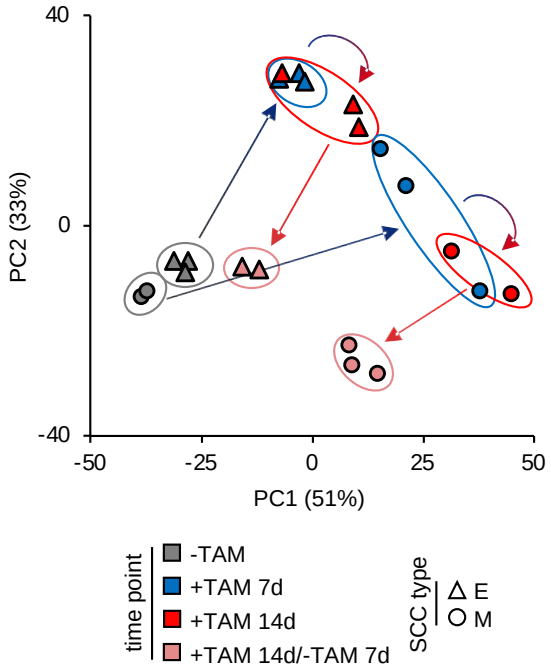

FIGURE 2.22: PCA of HMLE-derived E-SCCs and M-SCCs processed for Omni-ATAC-seq at following time points: before TAM treatment (-TAM), after 7 days of continuous TAM treatment (+TAM 7d), after 14 days of continuous TAM treatment (+TAM 14d) and after 14 days of TAM treatment followed by 7 days of TAM discontinuation (+TAM 14d/-TAM 7d). Figure modified from Eichelberger et al., 2020.

Next, the dynamic expression of the TFs family ZEB and GRHL was checked in the RNA-seq data generated from the same samples processed for ATAC-seq for all previously considered treatment time points. We observed that the expression of *GRHL* factors was downregulated upon Twist1 activation for M-SCCs but not for E-SCCs, where especially *GRHL2* showed a consistent high expression. ZEB TFs, mainly ZEB1, were upregulated for both E and M samples upon TAM treatment. However, the dynamic changes were different for both sample types: for E-SCCs, *ZEB1* upregulation was mainly observable after 14 days of treatment, whereas for M-SCCs, *ZEB1* was already significantly upregulated following 7 days of TAM treatment (Figure 2.26). Moreover, *ZEB1* expression after 14 days of treatment was more than 2-fold upregulated in M-SCCs compared to E-SCCs, in consonance with our previous observation regarding *ZEB1* expression in HMLE-Twist1-ER EPCAM positive and negative sorted cells (Figure 2.19).

When checking the expression of other genes known to interact with *ZEB1* in a negative feedback loop, as *OVOL2* (Hong et al., 2015) or the MIR200-family members, which have also been described to be involved in cancer development (Humphries et al., 2015; Liu et al., 2019), we confirmed a downregulation of those genes in M-SCCs but a stable expression in E-SCCs upon transitory EMT-induction. This fact was reflected again in the chromatin status at the genomic region encoding *OVOL2* and the gene *MIR200CHG*, hosting the *MIR200*-family members *MIR200C* and *MIR141*, where the ATAC signal revealed a chromatin closing trend

for M-SCC samples upon TAM treatment but a rather stable chromatin openness for the E-SCC samples (Figures 2.23).

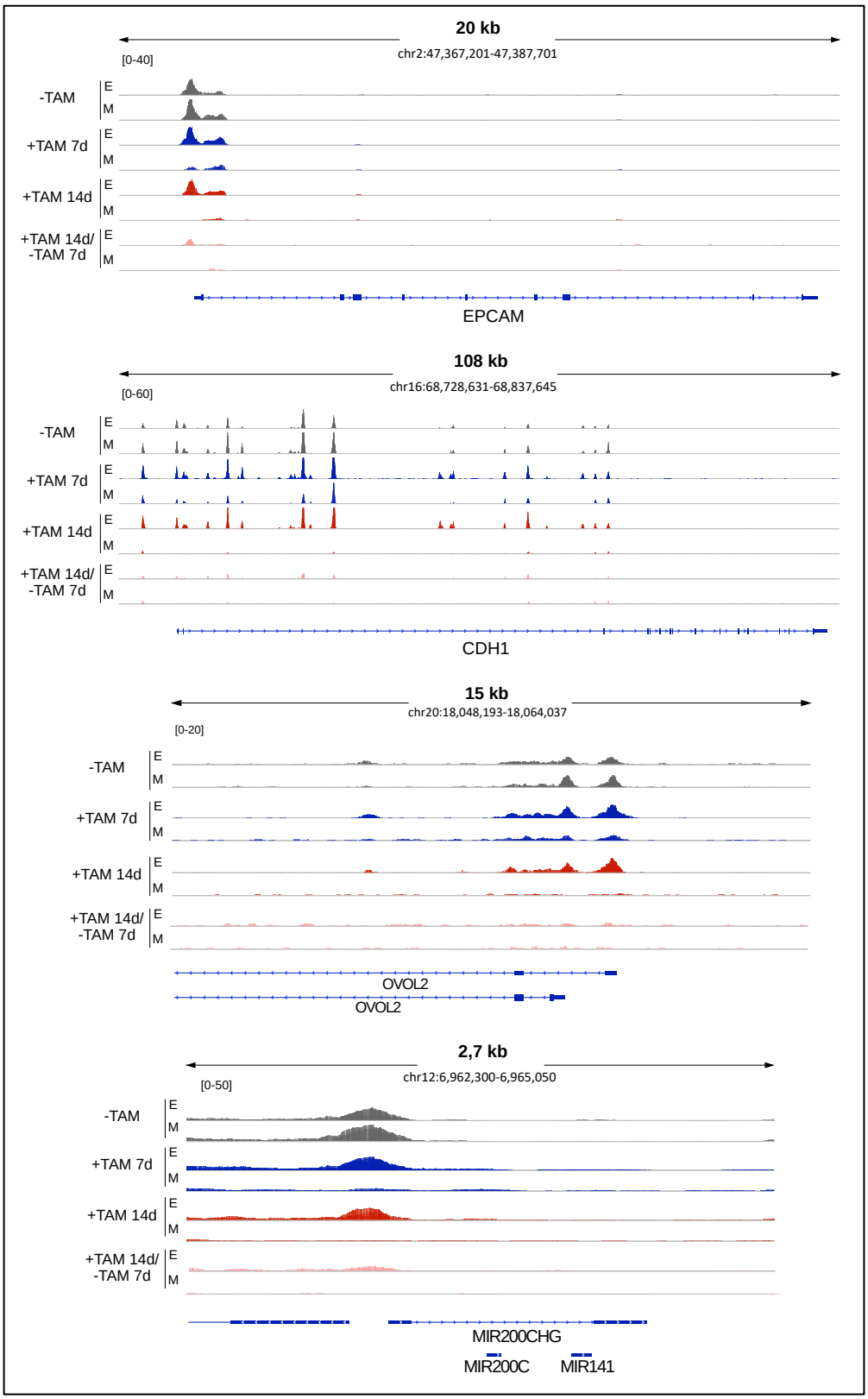

FIGURE 2.23: (Figure caption on next page.)

FIGURE 2.23: (Figure in previous page) IGV regions around *EPCAM*, *CDH1* and *OVOL2* loci, as well as around (some members of the family) *MIR200* (from top to bottom) showing the ATAC signal for E-SCCs (E) and M-SCCs (M) processed for Omni-ATAC-seq untreated (gray), after 7 days of TAM treatment (blue), after 14 days of TAM treatment (red) and after 14 days of TAM treatment followed by 7 days of TAM discontinuation (pink). Each track is the overlay of the tracks corresponding to the three biological replicates processed at each time point for each sample type.

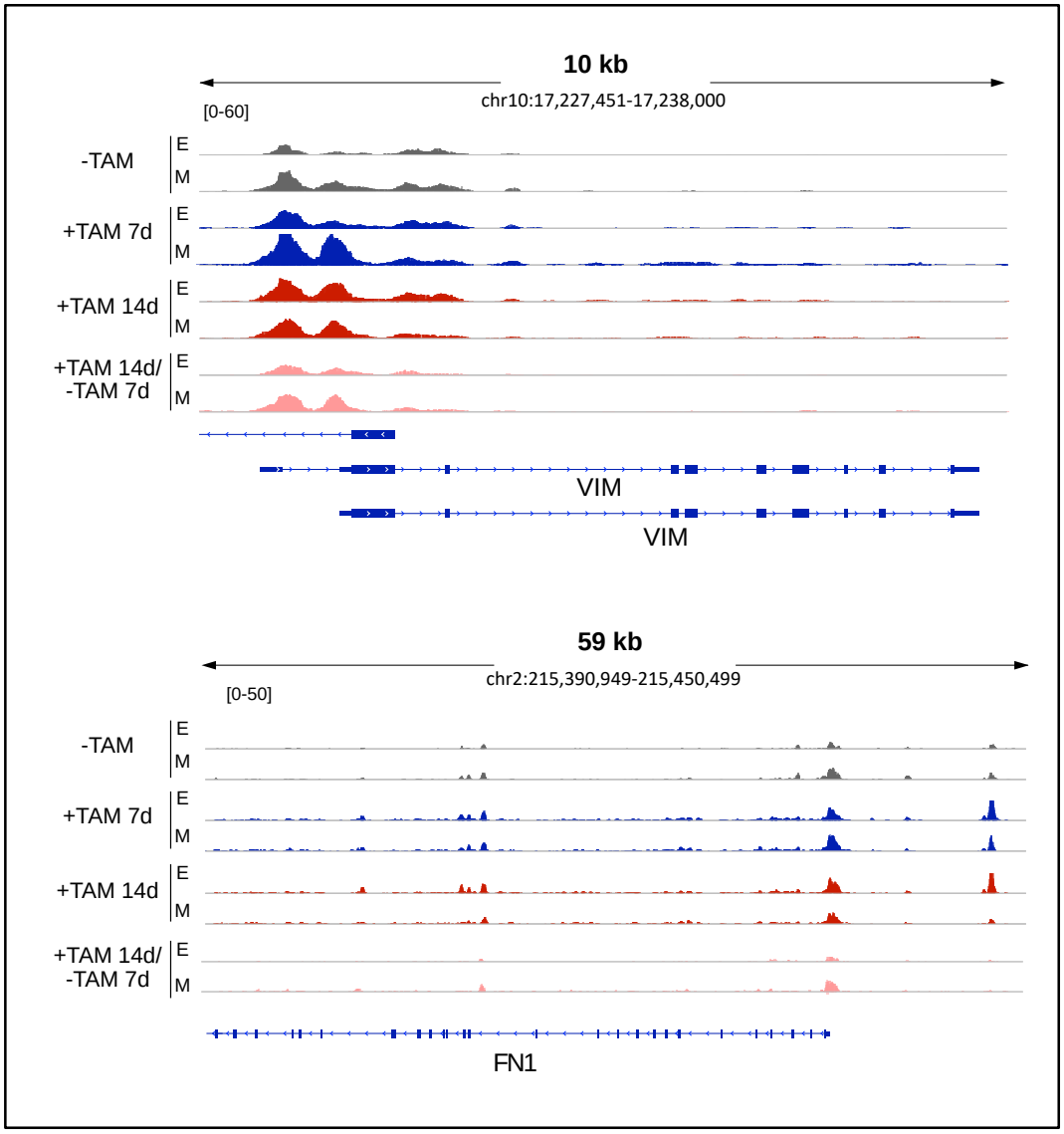

FIGURE 2.24: IGV regions around *VIM* and *FN1* loci (from top to bottom) showing the ATAC signal for E-SCCs (E) and M-SCCs (M) processed for Omni-ATAC-seq untreated (gray), after 7 days of TAM treatment (blue), after 14 days of TAM treatment (red) and after 14 days of TAM treatment followed by 7 days of TAM discontinuation (pink). Each track is the overlay of the tracks corresponding to the three biological replicates processed at each time point for each sample type.

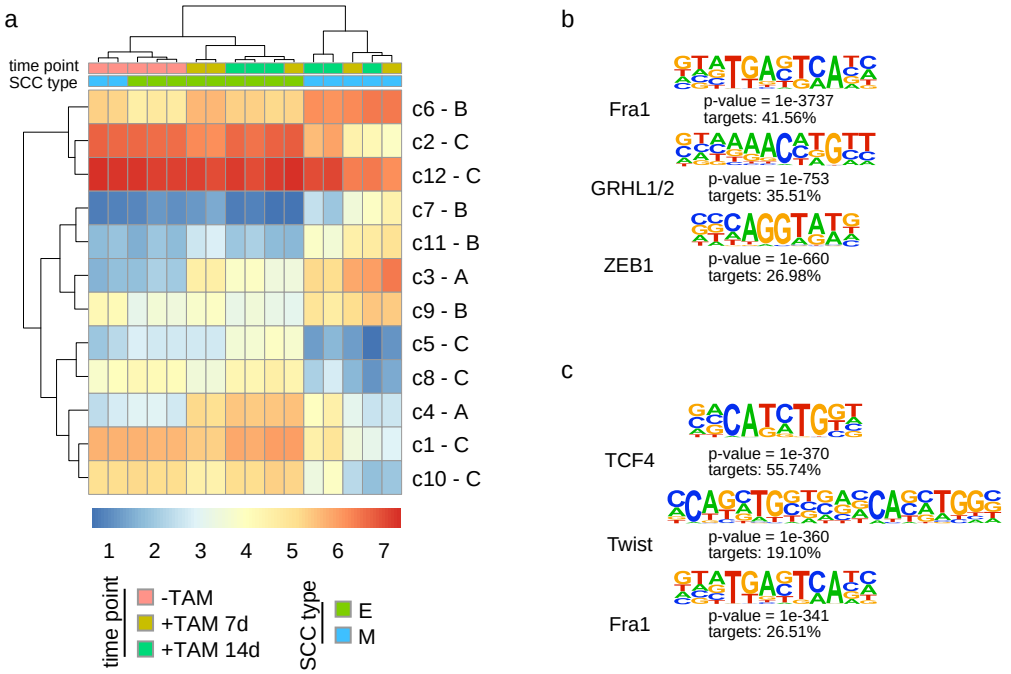

FIGURE 2.25: a, Heatmap of degree of chromatin accessibility showing 12 ATAC-peaks clusters (c1-c12) for E-SCCs (E) and M-SCCs (M) treated as previously described. Clusters can be grouped in three categories: A, chromatin regions opening up upon TAM treatment for both E and M samples; B, chromatin regions gaining accessibility only for M samples but not for E samples; and C, chromatin regions losing accessibility for M samples but not for E samples. b, HOMER *de novo* transcription factor binding prediction analysis showing the top three motifs found in the ATAC peaks subset corresponding to clusters included in group C; c, HOMER *known* transcription factor binding prediction analysis showing the top three motifs found in the ATAC peaks subset corresponding to clusters included in group A. Figures modified from Eichelberger et al., 2020.

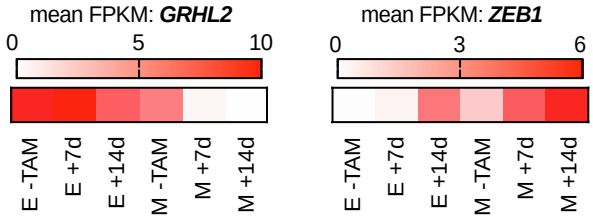

FIGURE 2.26: Heatmap showing the mean FPKM value of RNA-seq data for genes GRHL2 and ZEB1 for 3 E-SCCs (E) and 3 M-SCCs (M) samples treated as previously described. FPKM, fragments per kilobase of transcript per million fragments mapped. Figures modified from Eichelberger et al., 2020.

These results indicate that upon Twist1 activation the negative feedback loop inhibiting the activity of ZEB1 continues operating in E-SCCs, whereas it is deactivated in M-SCCs.

Therefore, epithelial genes like *OVOL2* and *CDH1*, which are ZEB1 targets, are downregulated and, as we observed, also the chromatin at the correspondent loci tends to acquire a "closed" status in M-SCCs during Twist1 activation (Figure 2.23). In contrast, typical mesenchymal genes, as *VIM* or *FN1*, remain highly expressed and the chromatin status at those sites continues being consistently "open" (Figure 2.24). However, the transitory Twist1 activation is, for E-SCC samples, not sufficient to disrupt the ZEB1 feedback loop and this inefficient *ZEB1* expression would ultimately lead to the observed EMT resistance in E samples.

Consequently, we next aimed to determine whether *ZEB1* overexpression (OE) was sufficient to overcome EMT-resistance in E-SCCs. With this purpose, we performed a second ATAC experiment using E-SCCs genetically engineered to conditionally overexpress *ZEB1* (samples listed in Table 2.4). After 7 days of induced *ZEB1* OE, E-SCCs were indeed able to transition to a mesenchymal phenotype. Besides, the chromatin status of the ZEB1 targets *EPCAM*, *CDH1*, *OVOL2* and the *MIR200*-family presented a reduced ATAC signal compared to the control E clones, similar to the M-SCCs behavior (Figure 2.27). The RNA-seq data also showed a downregulation of these epithelial markers for *ZEB1* OE E-SCCs, in consonance with the ATAC results. Nevertheless, when the OE of *ZEB1* was discontinued for 14 days, E-SCCs recovered their initial epithelial morphology and the expression of the aforementioned epithelial markers increased again. These results point out that ZEB1, even when generating epigenetic changes leading to the loss of epithelial features, is not sufficient itself to induce a stable and permanent transition to a mesenchymal state.

Taken together, our results indicate that in the EMT-resistant breast cancer-derived cell subpopulation there are underlying epigenetic mechanisms preventing the TF ZEB1 from repressing the epithelial traits in an irreversible way. Therefore, this subpopulation is characterized by a plasticity allowing cells to show a hybrid epithelial-mesenchymal phenotype. Even more, this dual character, specifically the presence of epithelial features, appears to be involved in the capacity of EMT-resistant cells to give rise to metastases, since a complete, stable and irreversible transition to a mesenchymal state prevents cells from proliferating uncontrollably, as it has been already described (Celiá-Terrassa et al., 2012).

2.2.1 Discussion II

Metastatic outgrowth, together with drug resistance and relapse, continues being one of the major challenges in cancer treatment, largely responsible for compromising patient survival. The epithelial-to-mesenchymal transition (EMT), a fundamental process for new developing tissues and organs, has been amply linked to tumorigenicity and metastasis-initiating capacity in many different types of cancer (Tsai et al., 2012; Krebs et al., 2017; Legras et al., 2017).

In this study, we established the existence of two different tumor cell circulating subpopulations derived from breast cancer patient liquid biopsies considering the expression levels of the surface marker EPCAM, which has been related with increased metastatic initiating capacity *in vivo* (Baccelli et al., 2013): $EPCAM^{high}$ and $EPCAM^{low}$. These two cell subpopulations appear to be different regarding their tumor-initiating capacity and also their capacity to build three-dimensional organoids, in tests carried out both *in vitro* and *in vivo*, presenting $EPCAM^{high}$ cells a greater tumorigenic and metastatic potential in all analyzed contexts.

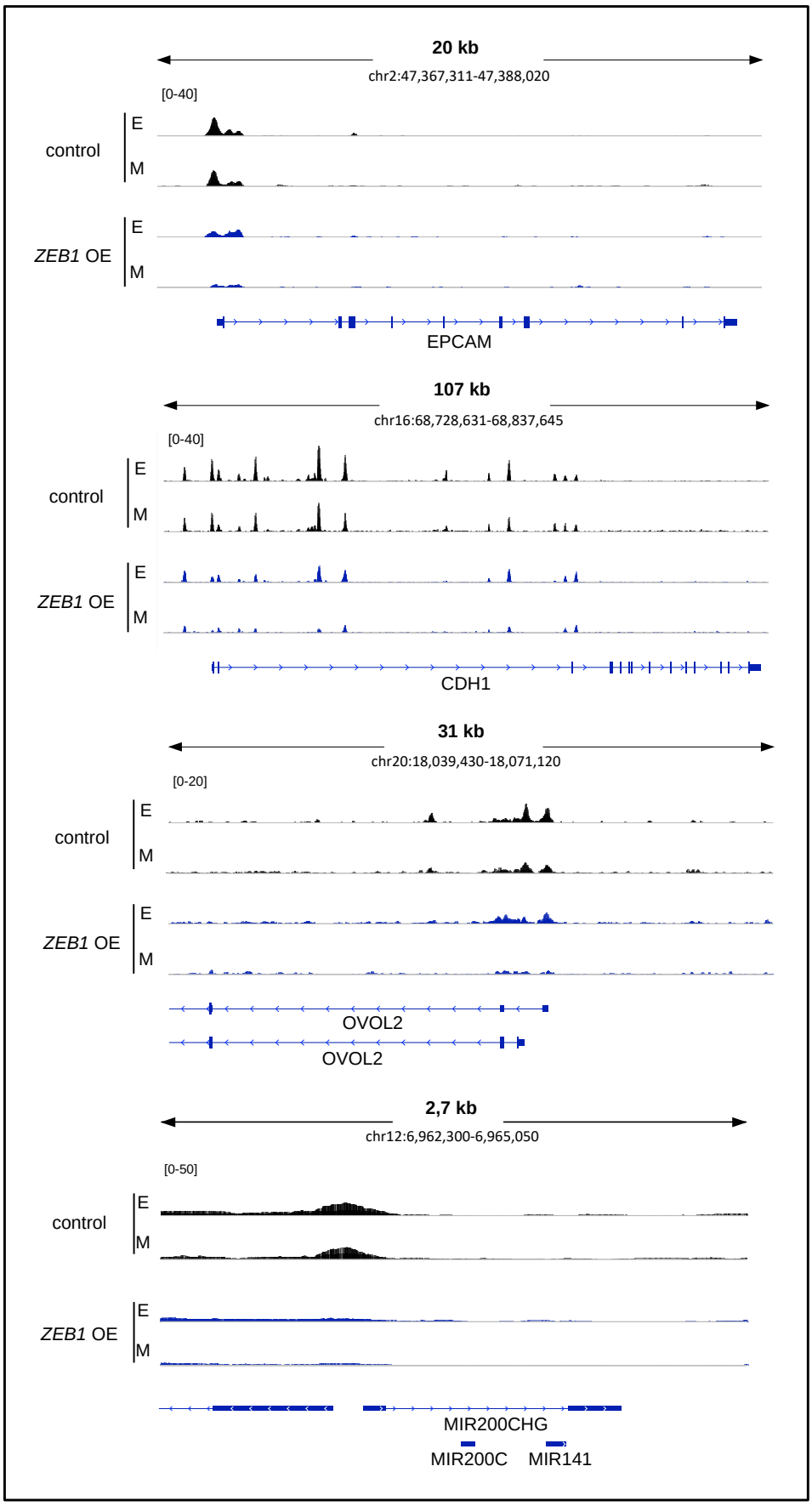

FIGURE 2.27: (Figure caption on next page.)

FIGURE 2.27: (Figure in previous page.) IGV regions around *EPCAM*, *CDH1*, *OVOL2* and *MIR200CHG* (hosting the members of the *MIR200*-family *MIR200C* and *MIR141*) loci (from top to bottom) showing the ATAC signal for E-SCCs (E) and M-SCCs (M) processed for Omni-ATAC-seq. Control clones (black) and ZEB1 overexpression (OE) clones (blue). Each track is the overlay of the tracks corresponding to the different biological replicates processed for each condition and sample type. Replicates processed for each sample type: E-SCC control, 2 replicates; E-SCC *ZEB1* OE, 4 replicates; M-SCC control, 2 replicates; M-SCC *ZEB1* OE, 2 replicates. Note: these data are not included in Eichelberger et al., 2020, pre-published version.

Transcriptional data showed that genes involved in the epithelial-to-mesenchymal transition were enriched in the $EPCAM^{low}$ subpopulation, suggesting that a successful transition to mesenchymal phenotypes is key for preventing an aberrant cellular outgrowth (Celiá-Terrassa et al., 2012). Indeed, the precise role of the EMT in tumor initiation and metastatic outgrowth is still being deciphered (Satoh et al., 2014; Pastushenko et al., 2018), as well as its contribution to tumor chemoresistance (Fischer et al., 2015; Zheng et al., 2015).

We observed, when applying a transient EMT-stimulus to tumor single-cell clones (SCCs) derived from the breast cancer cell line HMLE, genetically engineered to conditionally express the inducing-EMT transcription factor Twist1 (Casas et al., 2011; Schmidt et al., 2015), that a cell fraction was EMT-resistant. Moreover, this EMT-resistant cell fraction acquired a hybrid phenotype during the EMT induction, presenting traits from both epithelial and mesenchymal cells and reversing their status to the initial epithelial features, once the transitory EMT stimulus was over. This fact is in line with the previous observation describing the epithelial marker E-cadherin as required for the occurrence of metastases (Padmanaban et al., 2019) and also with the formerly description of a hybrid epithelial-mesenchymal state as crucial for cellular tumor initiation (Yu et al., 2013; Kröger et al., 2019), hypothesis which is increasingly gaining evidence and support (Teeuwssen et al., 2019; Sinha et al., 2020).

ATAC-seq data, allowed us to take a deeper insight into the epigenetic mechanisms underlying the different behavior of both EMT-resistant (E-SCCs) and EMT-sensitive (M-SCCs) cell fractions. The transcription factor ZEB1, known to be involved in EMT-induction as repressor of epithelial genes (Sánchez-Tilló et al., 2010; Vannier et al., 2013), appeared to be enriched in a subset of ATAC peaks representing chromatin regions getting closed upon EMT induction only for the EMT-sensitive M-SCCs but not for the EMT-resistant E-SSCs. The TF GRHL2, known to bidirectionally interact with ZEB1 in a negative feedback loop and to play an important role in breast cancer initiation (Werner et al., 2013) was also found to be overrepresented in the same subset of ATAC peaks. These results, consistently with the transcriptional data, strongly suggest that inefficient expression of *ZEB1* prevents E-SCCs from undergoing EMT and from acquiring a stable and permanent mesenchymal phenotype and, thus, cellular growth arrest.

However, we also observed that a temporarily sufficient *ZEB1* expression was not enough for overcoming irreversibly EMT resistance: while *ZEB1* was overexpressed in E-SCCs, cells developed mesenchymal traits but as soon as the overexpression was discontinued, E-SCCs recovered their original epithelial features, including chromatin opening in genomic regions

encoding epithelial identity genes, such as *CDH1*, *OVOL2* or *EPCAM*, and regained expression of those epithelial genes. These results suggest the existence of further epigenetic mechanisms in charge of actively maintaining the epithelial traits in the EMT-resistant cell fraction, which are ultimately responsible for the increased capacity to generate distal metastatic outgrowths.

2.2.2 List of processed samples

ID	name	SCC type	time point / condition	replicate
GS541	E1-TAMd1	E	-TAM	R1
GS542	E3-TAMd1	E	-TAM	R2
GS543	E5-TAMd1	E	-TAM	R3
GS544	M2-TAMd1	M	-TAM	R1
GS545	M3-TAMd1	M	-TAM	R2
GS546	M4-TAMd1	M	-TAM	R3
GS547	E1+TAMd7	E	+TAM 7d	R1
GS548	E3+TAMd7	E	+TAM 7d	R2
GS549	E5+TAMd7	E	+TAM 7d	R3
GS550	M2+TAMd7	M	+TAM 7d	R1
GS551	M3+TAMd7	M	+TAM 7d	R2
GS552	M4+TAMd7	M	+TAM 7d	R3
GS553	E1+TAMd14	E	+TAM 14d	R1
GS554	E3+TAMd14	E	+TAM 14d	R2
GS555	E5+TAMd14	E	+TAM 14d	R3
GS556	M2+TAMd14	M	+TAM 14d	R1
GS557	M3+TAMd14	M	+TAM 14d	R2
GS558	M4+TAMd14	M	+TAM 14d	R3
GS559	E1+TAMd14-TAMd7	E	+TAM 14d/-TAM 7d	R1
GS560	E3+TAMd14-TAMd7	E	+TAM 14d/-TAM 7d	R2
GS561	E5+TAMd14-TAMd7	E	+TAM 14d/-TAM 7d	R3
GS562	M2+TAMd14-TAMd7	M	+TAM 14d/-TAM 7d	R1
GS563	M3+TAMd14-TAMd7	M	+TAM 14d/-TAM 7d	R2
GS564	M4+TAMd14-TAMd7	M	+TAM 14d/-TAM 7d	R3
GS1004	M-ZEB#1	M	ZEB1 OE	R1
GS1005	M-GUS#10	M	control	R1
GS1006	E-GUS#2	E	control	R1

TABLE 2.4: continued on next page

ID	name	SCC type	time point / condition	replicate
GS1007	E-ZEB#4	E	ZEB1 OE	R1
GS1008	E-ZEB#10	E	ZEB1 OE	R2
GS1009	E-GUS#4	E	control	R2
GS1010	M-GUS#12	M	control	R2
GS1011	M-ZEB#6	M	ZEB1 OE	R2
GS1012	E-ZEB#9	E	ZEB1 OE	R3
GS1013	E-ZEB#14	E	ZEB1 OE	R4

TABLE 2.4: HMLE-Twist1-ER derived single-cell clone (SCC) samples processed for Omni-ATAC-seq. In all cases, 50k cells were processed. When needed, dead cell removal kit from Miltenyi used to assure a final cell viability of > 90%. First ATAC experiment: samples GS541-564; second ATAC-experiment: samples GS1004-1013.

2.3 Maintenance of Treg cell identity and function during inflammation

To complete our study about chromatin accessibility patterns defining certain diseases, we focused on a non-tumorigenic pathology characterized by inflammation. For it, we were collaborating with investigators from the Experimental Neuroimmunology department in Klinikum rechts der Isar (Technical University of Munich) to unravel the mechanisms preventing the loss of Treg cell identity in the context of inflammation. The results and conclusions of our investigations led to following publication: Garg et al., 2019.

For this project, accumulation of CNS Treg cells during rodent EAE was induced. During the peak of the disease, *Foxp3* expression was observed to be higher in CNS Treg cells compared with spleen (SPL) Treg cells (Figure 2.28.a), indicating the existence of mechanisms involved in the maintenance of Treg cell identity at sites of inflammation. To assess to which extent Treg cells were expressing Blimp1 in inflamed tissues, yellow fluorescent protein (YFP) reporter mice were used. It was observed that most Treg cells in inflamed CNS expressed Blimp1, double as many compared with splenic Treg cells. In contrast, only a small fraction of Treg cells were Blimp1 positive in naive SPL (Figure 2.28.b), indicating that the transcriptional signature characterizing Treg cells is substantially different in inflamed CNS and it is shaped by the upregulation of Blimp1. Other factors known to have an impact in the maintenance or loss of the Treg cell identity during inflammation are proinflammatory cytokines (Koch et al., 2012 and Overacre-Delgoffe et al., 2017). Gene Set Enrichment Analysis (GSEA) showed that genes induced by IFN-γ, IL-12 and IL-27 were enriched, when comparing Treg cells in CNS versus Treg cells in spleen. In particular, 4 genes were found to be common in all three gene sets (Figure 2.28.c), including *Prdm1*, which encodes Blimp1.

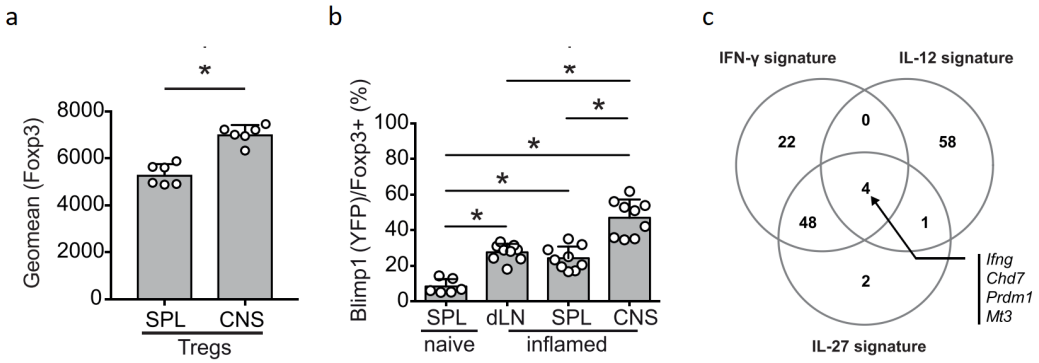

FIGURE 2.28: a) Foxp3 expression assessed by intracellular staining of CNS Treg and SPL Treg cells at the peak of EAE; b) Blimp1 expression of Foxp3+ Treg cells in naive spleen and in inflamed lymph node, spleen and CNS, measured in Blimp1 YFP reporter mice; c) Venn diagram showing the number of upregulated genes in CNS Treg cells versus splenic Treg cells induced by the proinflammatory cytokines IFN-γ, IL-12 and IL-27. Figures from Garg et al., 2019.

To investigate the effects of Blimp1 deficiency in Treg cells recruited to the inflamed CNS, Treg cell conditional Blimp1-deficient mice (*Foxp3 Cre* x *Prdm1*$^{flox/flox}$, but referred to as *Blimp1*$^{\Delta Foxp3}$ in advance) were generated. After induction of EAE through immunization, it was observed that mice deficient in Blimp1 presented more severe disease signs compared to control mice (Figure 2.29.a), showing that Blimp1 expression in CNS Treg cells is crucial for fighting the inflammation in that tissue. Moreover, *Blimp1*$^{\Delta Foxp3}$ mice showed decreased Foxp3 expression and lost IL-10 expression, although both IL-17 and IFN-γ were upregulated, compared with control mice (figure 2.29.b). Based on previous results, it could be argued that the Blimp1 associated loss of IL-10 may be responsible for the altered function of the CNS Treg cells (Bettelli et al., 1998) in inflamed tissue, nevertheless mice with conditional IL-10 deficiency in Treg cells (*IL-10*$^{\Delta Foxp3}$) neither developed more severe EAE, nor showed an increased expression of IL-17 or IFN-γ, compared with wild-type mice. This pointed out Blimp1 leading to a more general de-regulation of Treg cells identity and function at sites of inflammation.

To further investigate how Blimp1 influences Treg cell identity and function in an inflammatory context, mixed bone marrow chimeric mice were generated, mixing in a 1:1 ratio wild-type and *Blimp1*$^{\Delta Foxp3}$ bone marrow in recipient mice and analyzing them at the peak of EAE. It was observed a reduction of Foxp3 expression in the *Blimp1*$^{\Delta Foxp3}$ Treg cells compared with wild-type Tregs cells (Figure 2.30.a). Moreover, CNS Treg cells lost IL-10 expression and at the same time increased IL-17 expression (Figure 2.30.b). In total 342 genes were found in *Blimp1*$^{\Delta Foxp3}$ CNS Treg cells to be 2-fold upregulated and 409 genes were 2-fold downregulated, compared with wild-type Treg cells.

We performed ATAC-seq in order to understand the transcriptional mechanisms by which Blimp1 controlled CNS Treg cells in an inflamed environment. This was my major contribution to Garg et al., 2019. For it, 24 samples in total were processed, 15 from the spleen and 9 from the CNS (both Treg cells and Tconv cells, as reference). Due to the intrinsic difficulty in collecting these cells, the experiments were carried out with as many cells as we could sort, which in some cases resulted in as few as roughly 1200 cells (Table 2.5 at the end of this section). Unless otherwise indicated, cells were processed as described in Section 4.3.4, scaling

the reagents and the Tn5 transposase up or down depending on the final number of cells collected for each sample.

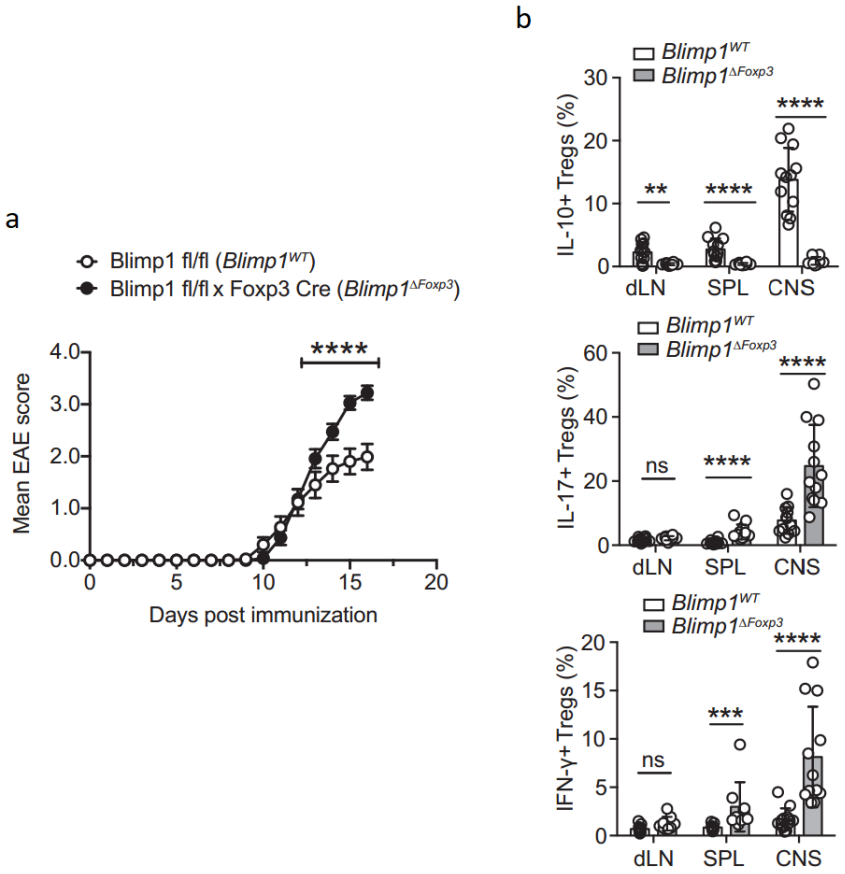

FIGURE 2.29: a) EAE scores for *Blimp1*$^{\Delta Foxp3}$ mice compared with control mice; b) Fraction of Treg cells isolated from dLN, spleen and CNS expressing IL-10, IL-17 and IFN-γ (from the top to the bottom) in *Blimp1*$^{\Delta Foxp3}$ versus control mice. Figures from Garg et al., 2019.

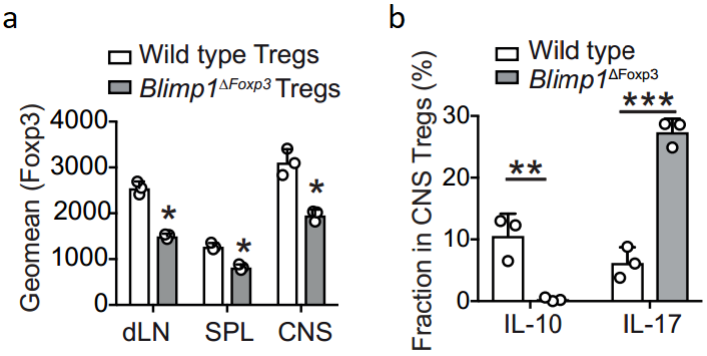

FIGURE 2.30: a) Foxp3 expression of wild-type and *Blimp1*$^{\Delta Foxp3}$ Treg cells in dLN, spleen and CNS of mixed bone marrow chimeric mice ; b) Fraction of CNS Treg cells expressing IL-10 and IL-17 in *Blimp1*$^{\Delta Foxp3}$ versus wild-type. Figures from Garg et al., 2019.

The number of cells obtained from CNS, especially *Blimp1*$^{\Delta Foxp3}$ Treg cells, were for most samples much lower than for splenic samples. These low cell numbers compromised the quality of the data gained from CNS and hindered the comparison with samples derived from spleen. Thus, we decided to conduct the analysis considering only splenic samples, unless otherwise indicated. Samples from the spleen were distributed in four separated PCA clusters depending on the cell type, Tconv or Treg cells, and cell condition, wild-type or *Blimp1*$^{\Delta Foxp3}$ (Figure 2.31).

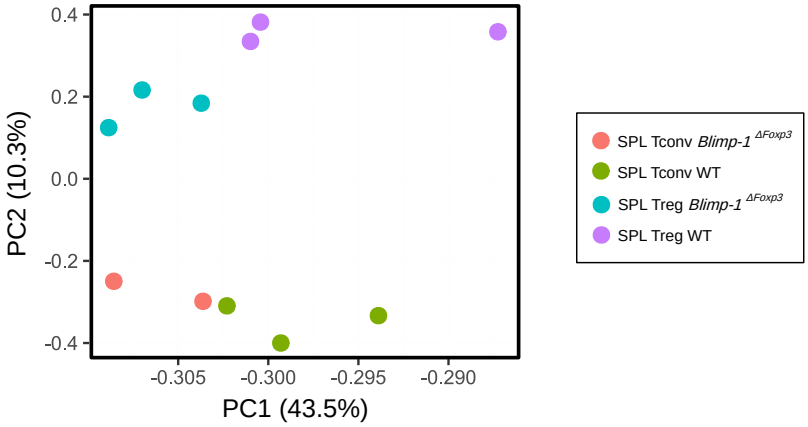

FIGURE 2.31: PCA of EAE chimeric samples from the spleen, wild-type and *Blimp-1*$^{\Delta Foxp3}$ Tconv and Treg cells. Only samples from the first processing batch considered (Table 2.5).

No remarkable differences could be detected regarding the openness of the chromatin when comparing splenic Tconv cells from wild-type mice and mixed bone marrow chimeras, since both conditions presented roughly same ATAC peaks (less than 0.2% of detected ATAC peaks were different, Figure 2.32, left). For Treg cells, we found around 8% of differential ATAC peaks, only detected in wild-type or in *Blimp1*$^{\Delta Foxp3}$ cells (Figure 2.32, right).

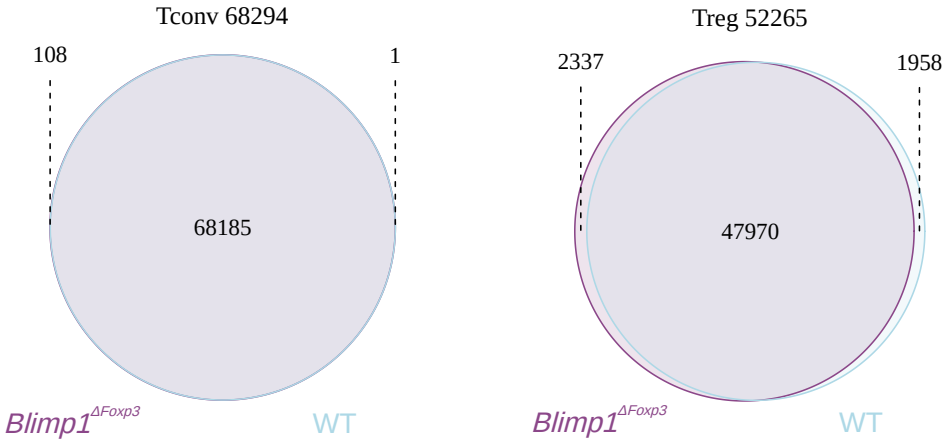

FIGURE 2.32: Venn diagrams showing the number of peaks called by HOMER in splenic wild-type and *Blimp1*$^{\Delta Foxp3}$ Tconv (left) and Treg (right) cells.

We then conducted a transcription factor binding site (TFBS) prediction analysis in those previously found differential ATAC peaks as described in Section 4.4.1.2 and identified groups of TF motifs predominantly enriched in wild-type or *Blimp1*$^{\Delta Foxp3}$ splenic Treg cells at the peak of EAE. In the top of the ranked TFs lists we found the nuclear factor NF-Kβ, Erg and AP1 enriched for wild-type splenic Treg cells differential ATAC peaks and, among the *Blimp1*$^{\Delta Foxp3}$

Treg cells differential peaks Ets, Runx1 or RORγ were present (Figure 2.33), showing that the Blimp1 shortage altered the transcriptional network in Treg cells during inflammation.

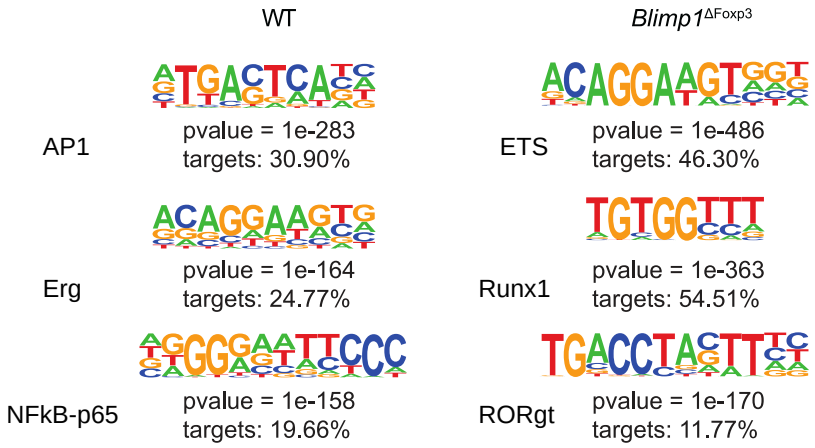

FIGURE 2.33: Top three transcription factor motifs found by HOMER to be enriched in differential ATAC peaks for wild-type (left) and *Blimp1*$^{\Delta Foxp3}$ splenic Treg cells (right). Figure from Garg et al., 2019.

In order to clarify how Blimp1 was directly involved in the transcriptional changes observed in *Blimp1*$^{\Delta Foxp3}$ CNS Treg cells when compared with CNS Treg wild-type cells, we used a publicly available Blimp1 ChIP-seq dataset (Mackay et al., 2016) and focused on genomic regions around genes previously found to be differentially expressed in *Blimp1*$^{\Delta Foxp3}$ versus wild-type CNS Treg cells. In those regions, we aimed to establish a correlation between Blimp1 binding and the openness of the chromatin, which would indicate a potential direct regulation of the correspondent loci by Blimp1. It has been already documented that Blimp1 can serve both as transcriptional repressor or promoter (Martins et al., 2008). For some of the identified top differential expressed genes, Blimp1 binding was observed in a region within 50 kb of the TSS, indicating a direct Blimp1 role in the gene regulation. Some of those genes, as IL-10, were downregulated and some, as Tnfsf8, were upregulated in *Blimp1*$^{\Delta Foxp3}$ CNS Treg cells compared with their wild-type counterparts (Figure 2.34). However, most of the top differentially expressed genes did not present a direct Blimp1 binding, even if differences in the chromatin openness were observed, implying that Blimp1 is not directly involved in the regulation of those genes. In fact, Foxp3, playing a key role in Treg identity and function, belongs to this category of genes: the first gene intron exhibits a significant reduction in ATAC signal for *Blimp1*$^{\Delta Foxp3}$ CNS Treg cells compared with wild-type cells although there is no Blimp1 binding at that location. This fact indicates that the regulation of Foxp3 in *Blimp1*$^{\Delta Foxp3}$ Treg cells, since it is not controlled by the direct binding of Blimp1 in the *Foxp3* locus, must require Blimp1 participation in an indirect manner.

It is known that the expression of Foxp3 is a complex process involving numerous mechanisms (Feng et al., 2014, Huehn et al., 2009). Since our data excluded a direct implication of

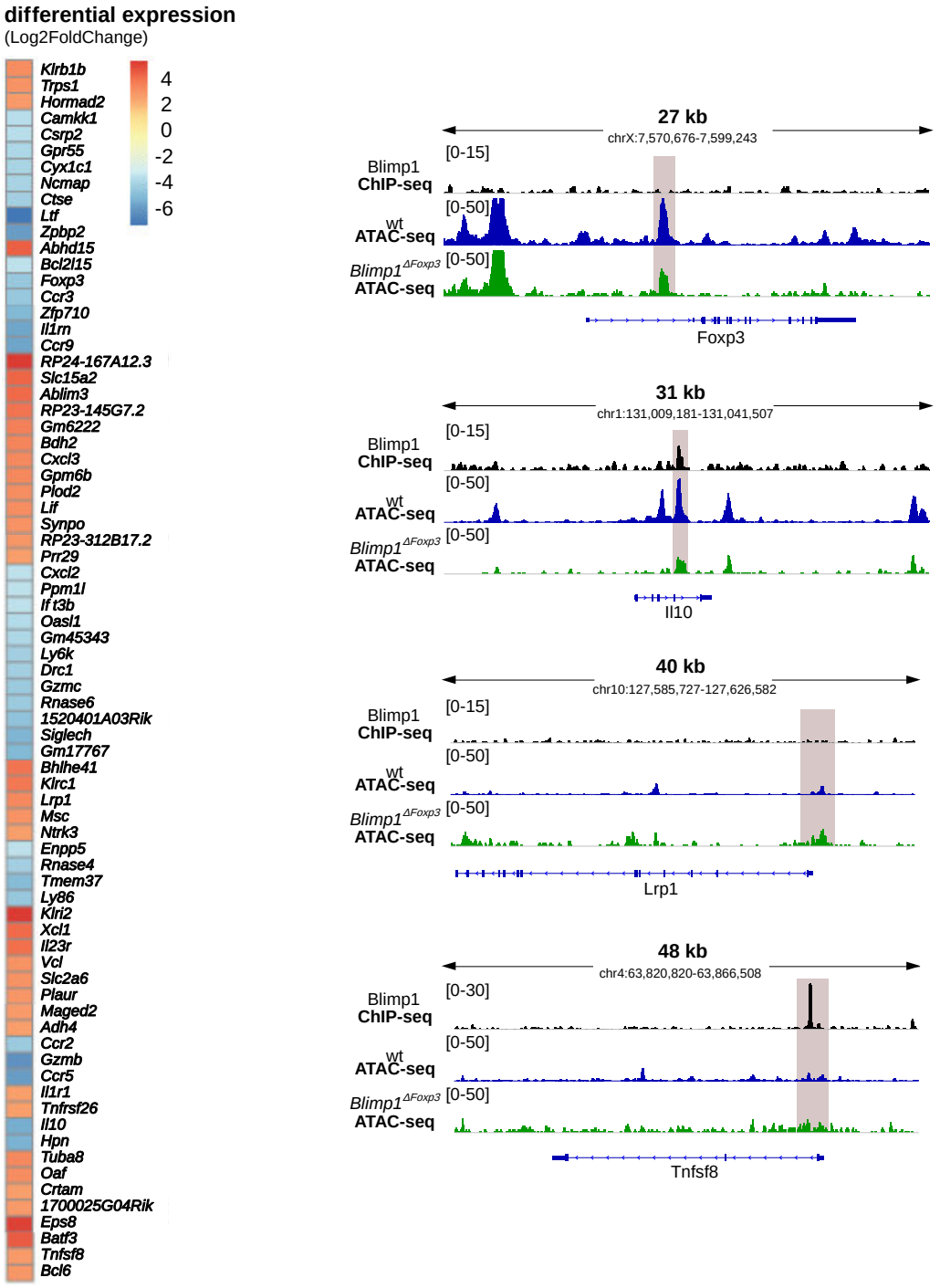

FIGURE 2.34: Left, list of most differential expressed genes in *Blimp1*$^{\Delta Foxp3}$ compared with wild-type CNS Treg cells (figure modified from Garg et al., 2019). Right, IGV regions around the *Foxp3*, *Il-10*, *Lrp1* and *Tnfsf8* loci (from top to bottom) showing Blimp1 binding sites in black (Blimp1 ChIP-seq data from Mackay et al., 2016) and ATAC peaks from wild-type and *Blimp1*$^{\Delta Foxp3}$ CNS Treg cells at the peak of EAE in blue and green, respectively.

Blimp1 in the control of the *Foxp3* locus, we investigated the role of Blimp1 as an indirect regulator of Foxp3. Checking the methylation status of the *Foxp3* locus in CNS cells

from chimeric mice, we confirmed that for Tconv the CNS2 intron was fully methylated and, in contrast, it was fully demethylated for Treg cells, as expected. However, the interrogated intronic element in *Blimp1*$^{\Delta Foxp3}$ CNS Treg cells presented a partial loss of demethylation, which was not observed in their wild-type counterparts. It is known that ten-eleven-translocation (Tet) family of enzymes and DNA-methyltransferases (Dnmts) work oppositely by demethylating or methylating, respectively, CpG islands within the genome (Tang et al., 2015). In this particular case Tet enzymes contribute to the demethylation status of the CNS2 intronic element within the *Foxp3* locus (Yue et al., 2016) and Dnmts generate opposite effect, methylating CpG islands in CNS2 (Feng et al., 2014, Josefowicz et al., 2009). When checking the expression of these two families of enzymes, we observed no remarkable de-regulation for Tet but specifically methyltransferase Dnmt3a showed significant up-regulation in *Blimp1*$^{\Delta Foxp3}$ CNS Treg cells compared with wild-type CNS Treg cells.

Moreover, the analysis of the ATAC peaks around the *Dnmt3a* locus showed an increase in the ATAC signal in Blimp1-deficient CNS Treg cells when compared with the control wild-type CNS Treg cells (Figure 2.35). Nevertheless, no significant difference in ATAC signal was found around the *Dnmt1* locus (Figure 2.36). These observations were consistent with the fact that Dnmt3a mediates *de novo* DNA methylation, while Dnmt1 is responsible for DNA pre-existing methylation maintenance (Lyko, 2018). Together with the fact that Blimp1 was found binding to the *Dnmt3a* locus, as showed by the ChIP-seq data (Mackay et al., 2016), our findings suggest a Blimp1-mediated inhibitory effect over Dnmt3a in an inflammation environment. In contrast, Tet-encoding genes did not show significant differences (or rather a decreased signal) regarding ATAC peaks in Blimp1-deficient Treg cells when compared with control Treg cells, nor presented Blimp1 binding (Figure 2.37).

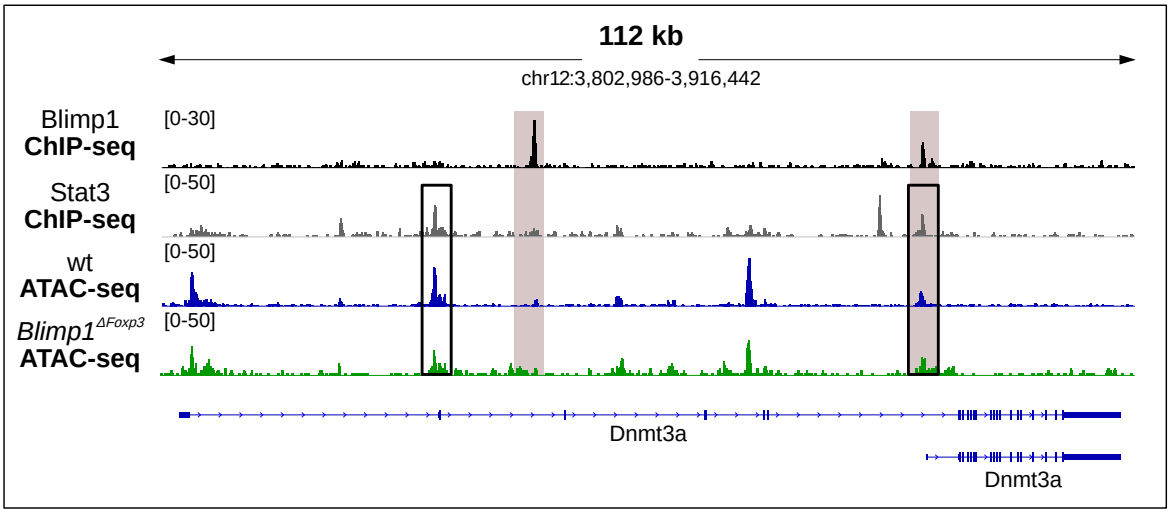

FIGURE 2.35: IGV region around the *Dnmt3a* locus displaying Blimp1 binding sites in black (Blimp1 ChIP-seq data from Mackay et al., 2016), STAT3 binding sites in gray (STAT3 ChIP-seq data from Hirahara et al., 2015) and ATAC peaks from wild-type and *Blimp1*$^{\Delta Foxp3}$ CNS Treg cells at the peak of EAE in blue and green, respectively. Blimp1 binding overlapping with ATAC peaks is highlighted with gray shadow; STAT3 binding overlapping with ATAC peaks is highlighted with a black box.

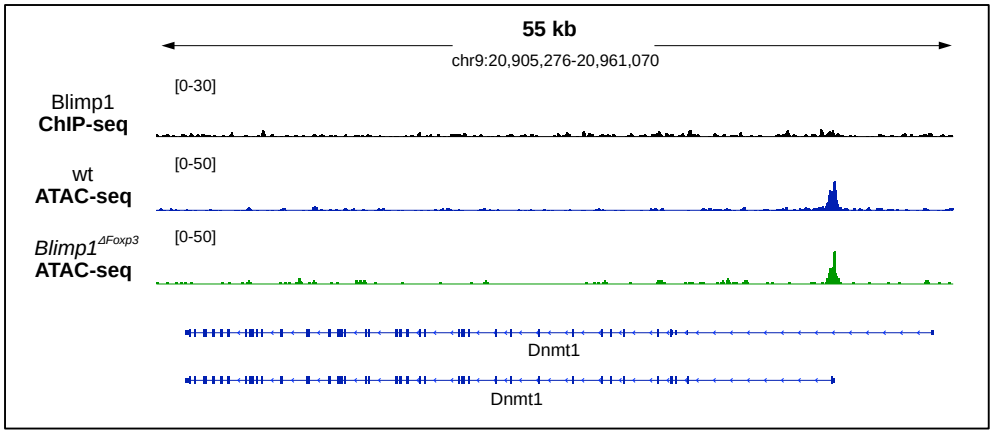

FIGURE 2.36: IGV region around *Dnmt1* locus, showing Blimp-1 binding in black (Blimp-1 ChIP-seq data from Mackay et al., 2016) and ATAC peaks from wild-type and *Blimp-1*$^{\Delta Foxp3}$ CNS Treg cells at the peak of EAE in blue and green, respectively.

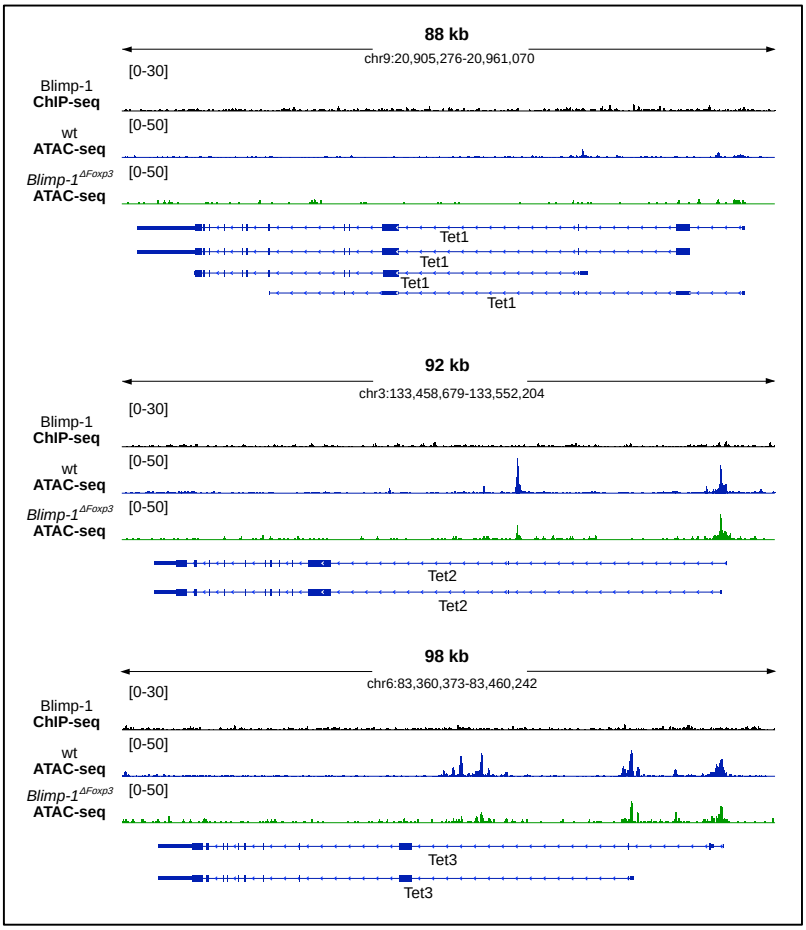

FIGURE 2.37: IGV regions around *Tet1*, *Tet2* and *Tet3* loci (from top to bottom), showing Blimp-1 binding in black (Blimp-1 ChIP-seq data from Mackay et al., 2016) and ATAC peaks from wild-type and *Blimp-1*$^{\Delta Foxp3}$ CNS Treg cells at the peak of EAE in blue and green, respectively.

The pro-inflammatory cytokine interleukin 6 (IL-6) has previously been documented to play an important role in the impaired function of Foxp3$^+$ Treg cells during inflammation (Yang et al., 2008). Our ATAC data contributed to strengthen the assumption that IL-6 was responsible for the de-regulation of Dnmt3a and, through it, the ultimate methylation of the *Foxp3* locus, since *Blimp1*$^{\Delta Foxp3}$ CNS Treg cells showed an increased ATAC signal at Blimp1 binding sites when compared with wild-type CNS Treg cells in the *Dnmt3a* locus (Figure 2.35).

In addition, when the *Dnmt3a* locus was checked for Stat3 binding (data from Hirahara et al., 2015), we observed that some binding sites overlapped with ATAC regions in that locus (Figure 2.35). It is plausible that Dnmt3a is recruited to the CNS2 intronic element by transcription factors binding to it, such as Ets1 (Polansky et al., 2010), since Ets was among the top transcription factors found to be differentially enriched in *Blimp1*$^{\Delta Foxp3}$ CNS Treg cells versus wild-type CNS Treg cells.

Further experiments were conducted to explore to which extent Blimp1 was involved in the Dnmt3a de-regulation induced by IL-6. Notably, when checking the *Stat3* locus, we found Blimp1 binding in a position where our ATAC data indicated an increased chromatin accessibility for *Blimp1*$^{\Delta Foxp3}$ CNS Treg cells when compared with wild-type CNS Treg cells (Figure 2.38), suggesting that Blimp1 could be responsible for the previously observed up-regulation of STAT3 in *Blimp1*$^{\Delta Foxp3}$ CNS Treg cells.

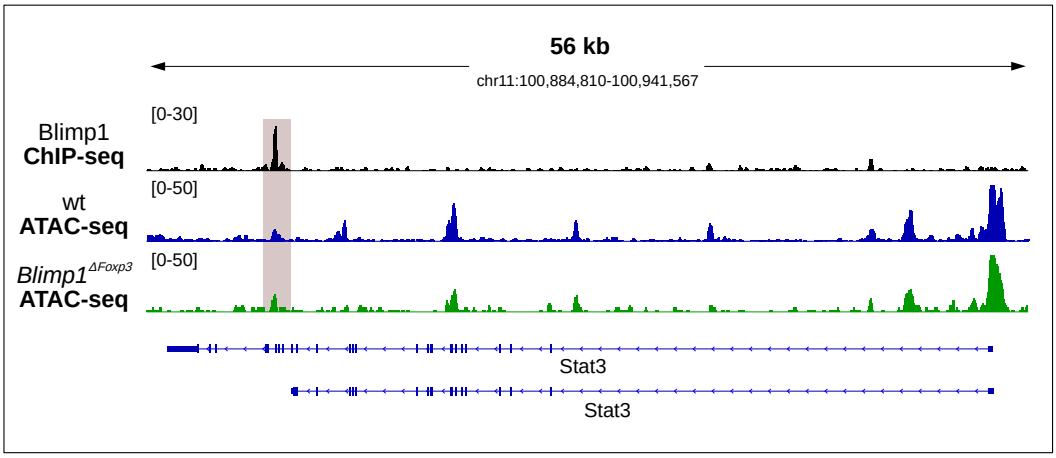

FIGURE 2.38: IGV region around the *Stat3* locus displaying Blimp1 binding sites in black (Blimp1 ChIP-seq data from Mackay et al., 2016) and ATAC peaks from wild-type and *Blimp1*$^{\Delta Foxp3}$ CNS Treg cells at the peak of EAE in blue and green, respectively. Blimp1 binding overlapping with ATAC peaks is highlighted with gray shadow.

Noteworthy, when treating Blimp1 deficient Treg cells pharmacologically with a Dnmt3a inhibitor, the loss of Foxp3 was successfully reversed, as shown by Garg et al., 2019.

Taken together, our data implied that Blimp1 is responsible for maintaining Treg identity and function in an inflammatory environment by suppressing the expression of the methyltransferase Dnmt3a, induced in such a context by the pro-inflammatory cytokine IL-6 through STAT3.

2.3.1 Discussion III

In this study Blimp1 was identified to be indispensable for maintenance of Treg cells identity and function during inflammation. Blimp1 counteracts the mechanisms by which the pro-inflammatory cytokine IL-6 induces, through STAT3, the ultimate methylation of the CNS2 intronic element in the *Foxp3* locus and, thus, its de-regulation. The continuous expression of Foxp3 is crucial for Treg cells normal function. However, it is lost when the demethylated state of CNS2 is reversed. Blimp1 inhibits the effect of the methyltransferase Dnmt3a, responsible for *de novo* methylation, hence preserving the correct expression of Foxp3 in an inflamed environment. Conditional Blimp1-deficient mice presented a more severe inflammatory phenotype upon EAE induction than control mice, associated to the loss of Treg cell identity, since the shortage in Blimp1 implies the loss of expression of the Foxp3 through methylation of the CNS2 intron.

My main contribution to this study was the series of ATAC-seq experiments, as well as the derived data analysis and interpretation, performed with Tconv and Treg cells from spleen and CNS of wild-type and conditional Blimp1-deficient mice, *Blimp1*$^{\Delta Foxp3}$. Data derived from these experiments showed fundamental differences in terms of chromatin accessibility between all four collected cell populations (Figure 2.31). These differences were more pronounced within the group of Treg cells (Figure 2.32). Moreover, ATAC-seq data allowed to deepen in those differences and showed a Blimp1-related altered transcriptional network for *Blimp1*$^{\Delta Foxp3}$ CNS Treg cells compared with their wild-type counterparts (Figure 2.33).

ATAC-seq data in combination with RNA-seq data provided a deeper insight into the epigenetic mechanisms behind the regulation of the key factor Foxp3, responsible for the maintenance of Treg cell identity and function during inflammation. Besides, using publicly available Blimp1 ChIP-seq data (Mackay et al., 2016) for completeness, we observed that Blimp1 could act as a direct but also as an indirect gene regulator, since not in all de-regulated genes Blimp1 bindings were found, despite the presence of an altered chromatin status in the same region for *Blimp1*$^{\Delta Foxp3}$ CNS Treg cells compared with wild-type CNS Treg cells (Figure 2.34).

Particularly the *Dnmt3a* locus, which showed several STAT3 bindings (Hirahara et al., 2015), presented Blimp1 bindings coinciding with ATAC peaks as well, suggesting a direct regulation of Dnmt3a by Blimp1 besides a IL-6-STAT3-dependent regulation (Figure 2.35). Interestingly, Blimp1 was also found binding within the *Stat3* locus and overlapping with ATAC peaks (Figure 2.38), which could indicate a potential regulation of STAT3 by Blimp1.

The motif analysis carried out within the differential ATAC peaks found in *Blimp1*$^{\Delta Foxp3}$ CNS Treg cells versus wild-type CNS Treg cells allows to speculate the mechanisms by which Dnmt3a could potentially be recruited to the CNS2 intronic element within the *Foxp3* locus, since the TF Ets, known to bind to CNS2 (Polansky et al., 2010) was found among the top factors enriched in conditional Blimp1-deficient Treg cells.

Nevertheless, further experiments will be required to unravel the precise manner in which Dnmt3a is recruited to the CNS2 intron and how Blimp1 transcriptionally regulates it and other potentially relevant genetic loci to preserve Treg cells identity and function in an inflamed environment.

2.3.2 List of processed samples

ID	name	type	processing batch	cell number
GS516	Spleen1.1 WT Tconv	SPL Tconv WT	1	20k
GS517	Spleen1.2 KO Tconv	SPL Tconv KO	1	20k
GS518	Spleen1.3 WT Treg	SPL Treg WT	1	20k
GS519	Spleen1.4 KO Treg	SPL Treg KO	1	20k
GS520	Spleen2.1 WT Tconv	SPL Tconv WT	1	20k
GS521	Spleen2.3 WT Treg	SPL Treg WT	1	20k
GS522	Spleen2.4 KO Treg	SPL Treg KO	1	20k
GS523	Spleen3.1 WT Tconv	SPL Tconv WT	1	20k
GS524	Spleen3.2 KO Tconv	SPL Tconv KO	1	20k
GS525	Spleen3.3 WT Treg	SPL Treg WT	1	20k
GS526	Spleen3.4 KO Treg	SPL Treg KO	1	20k
GS527	CNS1.1 WT Tconv	CNS Tconv WT	1	20k
GS528	CNS1.2 KO Tconv	CNS Tconv KO	1	20k
GS529	CNS1.3 WT Treg	CNS Treg WT	1	17k
GS530	CNS1.4 KO Treg	CNS Treg KO	1	4.5k
GS531	CNS2.1 WT Tconv	CNS Tconv WT	1	2618
GS532	CNS2.2 KO Tconv	CNS Tconv KO	1	3295
GS533	CNS2.3 WT Treg	CNS Treg WT	1	4184
GS534	CNS2.4 KO Treg	CNS Treg KO	1	1159
GS830	GS830_9_CNS_Tconv_KO	CNS Tconv KO	2	40k
GS831	GS831_11_SPL_Tconv_KO	SPL Tconv KO	2	40k
GS832	GS832_12_SPL_Treg_KO	SPL Treg KO	2	40k
GS833	GS833_13_SPL_Tconv_WT	SPL Tconv WT	2	40k
GS834	GS834_14_SPL_Treg_WT	SPL Treg WT	2	34k

TABLE 2.5: EAE samples from chimeric mice bone marrow (wild-type, WT, and *Blimp-1*$^{\Delta Foxp3}$ cells, KO) processed for Omni-ATAC-seq.

Chapter 3

General conclusions and future directions

With this work we aimed to contribute to a better understanding regarding the regulatory circuitry behind abnormal health condition. To achieve that, we applied ATAC-seq to study the chromatin accessibility and combined it in some cases with H3K27ac ChIP-seq to explore enhancer and super-enhancer landscapes. In the course of this work we investigated three different pathologies: AML, breast cancer and inflammation.

Chapter 2 is divided in three sections, each one focusing on one of the aforementioned pathologies. Following, the main achievements and conclusions are briefly summarized, as well as suggestions for future research are discussed.

In the first part of this work, we were able to successfully establish the experimental setup to process AML patient-derived samples, including the optimization of protocols to gentle thaw cells, enrich for blasts and further process them for ATAC-seq and H3K27ac ChIP-seq.

We characterized the regulatory landscape of 18 leukemic cell lines in terms of chromatin accessibility and enhancer and super-enhancer landscape (Section 2.1.1). A broad regulatory heterogeneity regarding the considered epigenetic markers was observed among the processed AML cell lines. This fact is in accordance with related literature, where AML heterogeneity at both genetic and epigenetic levels is well documented. Thus, the generated data represent a valuable tool as *in vitro* model for future research.

For exploring *in vivo* the effects induced by the standard therapy used to treat the AML condition, we worked with the PDX mouse model (Section 2.1.2). The results of our experiments suggest that the regulatory differences caused by the application of the standard treatment over several consecutive cycles is only minor and that these divergences become more pronounced at later treatment stages, rather than after just one cycle of therapy. Indeed, the chromatin changes induced by the therapy consisted mainly of regions opening up, i.e., increasing the accessibility, as treatment progressed.

Regarding AML patient-derived samples, we were able, in a preliminary experiment which served as proof of principle, to capture with our experimental approach fundamental regulatory differences between first diagnosis and the corresponding relapse from the same patient. In total, we processed samples derived from 12 AML patients. Since the number of viable tumor cells which could be finally sorted was very sample dependent, we prioritized ATAC-seq and RNA-seq, and performed H3K27ac ChIP-seq in a subgroup of sample for which enough blasts could be isolated.

In a preliminary analysis of the generated ATAC-seq data, we confirmed the already documented AML heterogeneity, not being possible to identify clusters of samples attending to the disease status, when performing a dimensionality reduction analysis. Moreover, this analysis suggested that samples belonging to the same patient tended to group together, independently to the stage of the disease (Section 2.1.3.3). This observation made us change the analysis strategy, so that we focus on analyzing patients independently to each other, to afterwards try to find common regulatory patterns shared by all the cohort. Following this strategy, we managed to identify some common relevant genes for a patient subgroup in relapse (Section 2.1.3.4). Some of these candidate genes have been indeed reported in the scientific literature to play a role in AML progression or response to therapy, as *MUL1*, *GADD45G* and *CSF2RB*, confirming the validity of our approach to determine possible significant genes worthy to be further explored regarding their implication in the context of AML. Indeed, within the generated list of potential AML-relevant genes, some of them with no well documented link with AML yet, such as *FCER1A*, *IKBKG* or *RGCC*, could serve as promising candidates to be further studied making use of the AML cell lines previously characterized in this work. Nevertheless, none of the genes or pathways identified were common to all the patients of the cohort, evidencing once more the intrinsic heterogeneity characteristic of this disorder.

Definitely, a clear goal to be addressed in future research regarding AML would be to use this collection of data as input in a broader project aiming to validate potential candidate genes found to be relevant in the AML relapse stage, using for it the AML cell lines characterized in the present work.

Certainly there are other types of cancer in which relapse or other characteristic tumor features represent major challenges hindering successful treatment. For this reason, in the second part of this work, we studied breast cancer, where the appearance of metastases constitute a frequent threat, affecting to 20-30% of all diagnosed patients and seriously threatening the long-term patient survival.

We aimed to contribute to the research leaded by colleagues from the Institute of Stem Cell Research, (Helmholtz Center Munich) exploring differences in chromatin accessibility patterns through ATAC-seq in two breast cancer-derived cell subpopulations which showed distinct tumor initiating capacity but were also different regarding their ability to irreversibly undergo the epithelial-to-mesenchymal transition (Section 2.2). When applying a transient EMT-inducing stimulus, one of the cell fractions, E-SSC, not only was resistant to undergo EMT, but also showed a hybrid phenotype, displaying both epithelial and mesenchymal features as long as the stimulus persisted, returning to the epithelial phenotype once it was over, whereas the other fraction, M-SCC, was EMT sensitive and able to shift to a stable and irreversible mesenchymal phenotype. Applying ATAC-seq to both cell fractions allowed us to get a deeper insight into the regulatory particularities of these two distinct cell types, since we were able to observe fundamental differences in terms of chromatin accessibility among them. Interestingly, the transcription factor ZEB1, known to play a role in EMT induction by repressing certain epithelial genes, was found to be enriched in a subset of ATAC peaks corresponding to chromatin regions closing down upon EMT for the EMT-sensitive cell fraction only, but not for the EMT-resistant cells.

Therefore, our results in the context of this research allowed us to identify distinct dynamics regarding the chromatin accessibility among both EMT-resistant and -sensitive cell

fractions and contributed to strengthen the hypothesis that an insufficient *ZEB1* expression was responsible for preventing the EMT-resistant cells to undergo EMT and acquire a stable mesenchymal phenotype, characterized by cellular growth arrest. This kind of cells able to withstand EMT actively maintaining epithelial traits, thus, present an increased ability to cause distal metastatic outgrowths.

Nevertheless, cancer is not the only pathology where studying chromatin structure patterns can offer a better understanding of the regulatory machinery underneath. Indeed, in the third part of this work, we had the opportunity to collaborate with investigators from the Experimental Neuroimmunology department in Klinikum rechts der Isar (Technical University of Munich) to study regulatory T cells in the context of inflammation (Section 2.3).

In this case, the transcription factor Blimp1 was previously identified to be indispensable for the maintenance of Treg cells identity and function during inflammation. Data derived from our ATAC-seq experiments, performed in wild-type and Blimp1 deficient EAE mice, in combination with RNA-seq data, contributed to hypothesize potential underlying mechanisms of action, contributing to a better understanding of the question. In this regard, the TF Ets was found enriched in a particular ATAC peak subset characteristic for Blimp1 deficient CNS Treg cells, suggesting that it could be relevant for the maintenance of the Treg identity. Even though more experiments would be necessary to fully unravel the mechanisms of action by which Treg cells are able to maintain their function, our approach including ATAC-seq (in combination with RNA-seq and ChIP-seq data) provided valuable information.

Definitely, as showed throughout this work, the study of chromatin accessibility patterns in disease condition can provide valuable information and serve as basis to elaborate hypothesis about the regulatory circuitry involved in the studied pathology. Even more, this knowledge in combination with information about the enhancer and SE landscape could be used to narrow down the wide spectrum of possible candidate genes critical for the disease condition which could potentially be used as target for developing more effective therapies.

Chapter 4

Materials and Methods

4.1 Materials

4.1.1 AML cell lines

name	cat. number	species	origin	FAB subtype
EOL-1	DSMZ ACC 386	homo sapiens	acute myeloid (eosinophilic) leukemia	M6
HL-60	DSMZ ACC 3	homo sapiens	acute myeloid leukemia	M3
K-562	DSMZ ACC 10	homo sapiens	chronic myeloid leukemia in blast crisis	-
KASUMI-1	DSMZ ACC 220	homo sapiens	acute myeloid leukemia	M2
KG-1a	DSMZ ACC 421	homo sapiens	acute myeloid leukemia	M6
ME-1	DSMZ ACC 537	homo sapiens	acute myeloid leukemia	M4
MM-1	DSMZ ACC 252	homo sapiens	acute monocytic leukemia	M5
MM-6	DSMZ ACC 124	homo sapiens	acute monocytic leukemia	M5
MOLM-13	DSMZ ACC 554	homo sapiens	acute myeloid leukemia	M5
MV4-11	DSMZ ACC 102	homo sapiens	acute monocytic leukemia	M5
NB-4	DSMZ ACC 207	homo sapiens	acute promyelocytic leukemia	M3
OCI-AML3	DSMZ ACC 582	homo sapiens	acute myeloid leukemia	M4
OCI-AML5	DSMZ ACC 247	homo sapiens	acute myeloid leukemia	M4
PL-21	DSMZ ACC 536	homo sapiens	acute myeloid leukemia	M3
SD-1	DSMZ ACC 366	homo sapiens	acute lymphoblastic leukemia	-
SKM-1	DSMZ ACC 547	homo sapiens	acute myeloid leukemia	M5
THP-1	DSMZ ACC 16	homo sapiens	acute monocytic leukemia	M5
U-937	DSMZ ACC 5	homo sapiens	histiocytic lymphoma	-

TABLE 4.1: List of all AML cell lines used in this work. These AML cell lines are commercially available (Leibniz Institute DSMZ-German Collection of Microorganisms and Cell Cultures GmbH) and were provided by AG Spiekermann - (Medizinische Klinik und Poliklinik III, LMU). Cell lines K-562, SD-1 and U-937 correspond to chronic myeloid leukemia (CML), the first one, and other types of lymphoid leukemias, the second and third ones. The last column includes information about the AML subtype according to the FAB classification (Bennet et al., 1985).

4.1.2 PDX samples

sample ID	PDX ID	time point	additional information
GS354	27224	control	-
GS355	27226	control	-
GS356	28675	control	-
GS357	28676	control	-
GS358	27227	C1	-
GS359	27077	C1	-
GS360	26697	control	slow growing clone
GS361	28666	C3	-
GS362	28669	C3	-
GS378	32751	C3	-
GS379	32814	C3	-
GS420	29686	control	-
GS421	29688	control	-
GS422	26699	C1	-
GS423	26617	C3	-
GS424	27093	C1	slow growing clone
GS425	28658	C1	-
GS426	25831	C3	slow growing clone
GS427	28650	C1	-
GS428	26677	C1	-
GS429	30940	C3	-

TABLE 4.2: List of all PDX samples processed for ATAC-seq. All samples come originally from one AML patient-derived sample isolated at the time of first relapse. PDX samples were generated and provided by the Research Unit Apoptosis in Hematopoietic Stem Cells, Helmholtz Centre Munich.

4.1.3 AML patient-derived samples

patient	sample ID	disease status
patient 1 (P1)	HS09	primary diagnosis
	HS08	relapse
patient 2 (P2)	HS14	primary diagnosis
	HS15	relapse
patient 3 (P3)	HS16	primary diagnosis
	HS17	relapse
patient 4 (P4)	HS18	primary diagnosis
	HS19	relapse
patient 5 (P5)	HS20	primary diagnosis
	HS21	relapse
patient 6 (P6)	HS22	primary diagnosis
	HS23	relapse
patient 7 (P7)	HS25	primary diagnosis
	HS26	relapse
patient 8 (P8)	HS28	primary diagnosis
	HS29	relapse
patient 9 (P9)	HS31	primary diagnosis
	HS32	relapse
patient 10 (P10)	HS37	primary diagnosis
	HS38	relapse
patient 11 (P11)	HS40	primary diagnosis
	HS41	relapse
patient 12 (P12)	HS43	primary diagnosis
	HS44	relapse

TABLE 4.3: List of AML patient-derived samples. All human samples were provided by AG Götze - Klinikum rechts der Isar, III. Medizinische Klinik, TUM - and AG Metzeler and AG Subklewe - Medizinische Klinik und Poliklinik III, LMU. A detailed data sheet containing additional information can be found at the end of this work in Appendix A.

4.1.4 Consumables and other material

4.1.4.1 Reagents and kits

reagent	manufacturer	cat. number
Agilent High Sensitivity DNA Kit	Agilent	5067-4626
AMPure XP beads	Beckman Coulter	A63881
Buffer RLT Plus	Qiagen	1053393
CD33 MicroBeads, human	Miltenyi Biotec	130-045-501
Dead Cell Removal Kit	Miltenyi Biotec	130-090-101
Digitonin	Promega	G9441
DNase I	Thermo Fisher Scientific	EN0521
Dynabeads Protein A	Thermo Fisher Scientific	10002D
Formaldehyde	Pierce	28906
Illumina DNA Prep Tagmentation	Illumina	FC-121-1030C
MinElute PCR Purification Kit	Qiagen	28004
NEBNext Ultra II - DNA library Prep Kit	Biolabs	E7645S
NEBNext Master Mix	Biolabs	M0541S
Propidium Iodide (PI)	Thermo Fisher Scientific	BMS500PI
Protease Inhibitor Cocktail	Sigma	04693159001
Proteinase K	Thermo Fisher Scientific	EO0491
Qubit dsDNA HS assay	Invitrogen	Q32854
RNase A	Thermo Fisher Scientific	EN0531
SensiMix SYBR No-ROX Kit	Bioline	QT650-05

TABLE 4.4: List of the most relevant reagents and kits, as well as corresponding manufacturers.

4.1.4.2 Antibodies

target	host	type	manufacturer	cat. number	purpose
H3K27ac	rabbit	polyclonal	Diagenode	pAb-174-050	ChIP-seq
BV421 CD33	mouse	monoclonal	Biolegend	303415	FACS
PerCP/Cyanine5.5 CD34	mouse	monoclonal	Biolegend	343521	FACS

TABLE 4.5: Detailed list of antibodies.

4.1.4.3 Oligonucleotides for library preparation

A. Customed oligonucleotides for ATAC-seq

ID	sequence
Ad1	AATGATACGGCGACCACCGAGATCTACACTCGTCGGCAGCGTCAGATGTG
Ad2.1	CAAGCAGAAGACGGCATACGAGATTCGCCTTAGTCTCGTGGGCTCGGAGATGT
Ad2.2	CAAGCAGAAGACGGCATACGAGATCTAGTACGGTCTCGTGGGCTCGGAGATGT
Ad2.3	CAAGCAGAAGACGGCATACGAGATTTCTGCCTGTCTCGTGGGCTCGGAGATGT
Ad2.4	CAAGCAGAAGACGGCATACGAGATGCTCAGGAGTCTCGTGGGCTCGGAGATGT
Ad2.5	CAAGCAGAAGACGGCATACGAGATAGGAGTCCGTCTCGTGGGCTCGGAGATGT
Ad2.6	CAAGCAGAAGACGGCATACGAGATCATGCCTAGTCTCGTGGGCTCGGAGATGT
Ad2.7	CAAGCAGAAGACGGCATACGAGATGTAGAGAGGTCTCGTGGGCTCGGAGATGT
Ad2.8	CAAGCAGAAGACGGCATACGAGATCCTCTCTGGTCTCGTGGGCTCGGAGATGT
Ad2.9	CAAGCAGAAGACGGCATACGAGATAGCGTAGCGTCTCGTGGGCTCGGAGATGT
Ad2.10	CAAGCAGAAGACGGCATACGAGATCAGCCTCGGTCTCGTGGGCTCGGAGATGT
Ad2.11	CAAGCAGAAGACGGCATACGAGATTGCCTCTTGTCTCGTGGGCTCGGAGATGT
Ad2.12	CAAGCAGAAGACGGCATACGAGATTCCTCTACGTCTCGTGGGCTCGGAGATGT
Ad2.13	CAAGCAGAAGACGGCATACGAGATATCACGACGTCTCGTGGGCTCGGAGATGT
Ad2.14	CAAGCAGAAGACGGCATACGAGATACAGTGGTGTCTCGTGGGCTCGGAGATGT
Ad2.15	CAAGCAGAAGACGGCATACGAGATCAGATCCAGTCTCGTGGGCTCGGAGATGT
Ad2.16	CAAGCAGAAGACGGCATACGAGATACAAACGGGTCTCGTGGGCTCGGAGATGT
Ad2.17	CAAGCAGAAGACGGCATACGAGATACCCAGCAGTCTCGTGGGCTCGGAGATGT
Ad2.18	CAAGCAGAAGACGGCATACGAGATAACCCCTCGTCTCGTGGGCTCGGAGATGT
Ad2.19	CAAGCAGAAGACGGCATACGAGATCCCAACCTGTCTCGTGGGCTCGGAGATGT
Ad2.20	CAAGCAGAAGACGGCATACGAGATCACCACACGTCTCGTGGGCTCGGAGATGT

TABLE 4.6: List of custom-made oligonucleotides used for ATAC-seq library preparation.

B. Primers sets for ChIP-seq libraries

primer set	manufacturer	cat. number
NEBNext Multiplex Oligos for Illumina (Set 1)	Biolabs	E7335S
NEBNext Multiplex Oligos for Illumina (Set 2)	Biolabs	E7500S

TABLE 4.7: Primers sets used for ChIP-seq library preparation.

4.1.4.4 Other material and equipment

name	manufacturer	cat. number
2100 Bioanalyzer Instrument	Agilent	G2939BA
S220 Focused-ultrasonicator	Covaris	500217
AutoMACS Pro Separator	Miltenyi Biotec	130-092-545
DynaMag-2 Magnet	Thermo Fisher Scientific	12321D
DynaMag-PCR-Magnet	Thermo Fisher Scientific	492025
FACS AriaFusion	Becton Dickinson	*NA*
QuadroMACS Separator	Miltenyi Biotec	130-090-976
Qubit 4 Fluorometer	Invitrogen	Q33238
LS columns	Miltenyi Biotec	130-042-401
microTUBE AFA Fiber	Covaris	520045

TABLE 4.8: List of other relevant material and equipment used in the course of this work.

4.2 Software and Algorithms

software and algorithms	reference	version
AffinityDesigner	*NA*	1.5.3.69
Agilent 2100 Expert Software	*NA*	B.02.09.SI725
Bamliquidator	Lin et al., 2016	1.2.0
bedtools2	*NA*	2.27.1
Bowtie1	Langmead et al., 2009	1.1.2
Cisco AnyConnect	*NA*	4.9.01095
Coltron	Federation et al., 2018	1.0.2
FACSDiva, BD	*NA*	8.0
FastQC	*NA*	0.11.7
FIMO	Grant et al., 2011	4.91
Galaxy	Afgan et al., 2018	*
HOCOMOCO	Kulakovskiy et al., 2018	v11

TABLE 4.9: continued on next page

software and algorithms	reference	version
HOMER	Heinz et al., 2010a	4.9
IGV	Robinson et al., 2011	2.6.1
Inkscape	*NA*	1.0
LaTeX/MiKTeX	*NA*	2.9
NetworkX	Hagberg et al., 2020	1.8.1
Notepad++	*NA*	7.9.1
Picard	*NA*	2.18.16
PuTTY	*NA*	0.72
Python	*NA*	2.7.5
R	*NA*	3.4.1
Regulatory Genomics Toolbox: Hint	Li et al., 2019	2
ROSE	Lovén et al., 2013; Whyte et al., 2013	2
RStudio	*NA*	1.2.1335
Samtools	Li et al., 2009	1.9
SFTP Net Drive Free	*NA*	*NA*
UCSCutils	*NA*	3.4.1

TABLE 4.9: List of all software and tools used for data analysis, data visualization, figure generation or in any other way needed for the completion of this work. Note: main R packages used are presented separately in Table 4.10.
* Galaxy instance of LAFUGA: https://blum.galaxy.lafuga.genzentrum.lmu.de

4.2.1 R packages

R packages	reference	version
clusterProfiler	Yu et al., 2012	3.14.3
ComplexHeatmap	Gu et al., 2016	2.2.0
DESeq2	Love et al., 2014	1.26.0
edgeR	McCarthy et al., 2012	3.28.1

TABLE 4.10: continued on next page

R packages	reference	version
gdata	Warnes et al., 2017	2.18.0
GenomicRanges	Lawrence et al., 2013a	1.38.0
ggplot2	Wickham, 2016	3.3.3
graph3d	Laurent et al., 2020	0.2.0
l2p	Finney et al., 2020	0.0.1
org.Hs.eg.db	Carlson, 2019	3.10.0
pathview	Luo et al., 2013	1.26.0
PCAtools	Blighe, 2019	1.2.0
rgl	Adler et al., 2021	0.105.22
rGREAT	McLean et al., 2010	1.18.0
Rtsne	Krijthe, 2015	0.15
SummarizedExperiment	Morgan et al., 2019	1.16.1
TCseq	Wu et al., 2019	1.10.0
VennDiagram	Chen, 2018	1.6.20

TABLE 4.10: Most relevant R packages required for the analysis performed.

4.3 Methods

4.3.1 Cell culture

For the completion of this work, the only cells I kept in culture and took care of personally were AML cell lines, since HMLE cells used for the collaborative project documented in Section 2.2 were directly provided by our cooperation-partners and ready to be processed.

AML cell lines listed in Table 4.1 were cultured with 80-90% of either alpha-MEM medium or RPMI 1640 medium, supplemented with 10-20% of h.i. FBS, attending to the instructions given by the provider and splitted according to the same instructions.

4.3.2 AML patient-derived samples pre-processing

In all cases, both AML samples, initial diagnosis and relapse belonging to the same patient, were pre-processed together. When available, one freshly isolated healthy donor sample was included in the pre-processing. Nevertheless, the timespan necessary for the pre-processing made infeasible to work with more than three samples in parallel.

4.3.2.1 Cell thawing

Vials containing frozen cells were warmed up for 1 min in water bath at 37°C. Content of vials was then immediately transferred to a 50 ml falcon containing 20 ml of AML blasts medium (alphaMEM + 12.5% h.i. FCS + 12.5% horse serum) and centrifuged (300 g, 10 min, RT). The supernatant was discarded carefully, pipetting it out, and cell pellet was re-suspended in 10 ml of FACS buffer (PBS + 1% FBS + 5 mM EDTA + DNase I - 10 units/ml). In case big clumps were observed, they were removed with the pipette tip, instead of filtering the cell suspension, to avoid losing cells. At this point, cell number and viability were assessed from a 10 µl aliquot of cell suspension which was mixed in a 1:1 ratio with trypan blue and observed under an optical microscope in a Neubauer chamber. This step is important to determine the amount of antibody (ab) to be used for FACS staining as well as the volume of FACS buffer to re-suspend cells in. The rest of the cells were centrifuged again (300 g, 5 min, RT), re-suspended in the corresponding volume of FACS buffer (to achieve a final concentration of 10^6 cells / 100 µl) and transferred to a 15 ml falcon. Cells were then ready for FACS staining.

4.3.2.2 Fluorescence activated cell sorting, FACS

According to the recommendations of the core facility Flow Cytometry at BMC for the sample preparation, AML patient-derived samples were stained with BV421 CD33 ab using 0.5 µg per Million cells in 100 µl, and control healthy donor-derived cells with CD34 ab using 1.25 µg per Million cells in 100 µl. Cells were incubated during 20 min at 4°C in the dark and, after incubation, ab was eluted by adding 3 ml of FACS buffer (PBS + 1% FBS + 5 mM EDTA + DNase I - 10 units/ml) and cells were centrifuged (300 g, 10 min, 4°C) and re-suspended in FACS buffer (from 500 µl to 1 ml, depending on cell number), then filtered and transferred into a FACS sorting tube, previously coated with FACS buffer. Shortly before sorting cells, Propidium Iodide (PI) was added to the cell suspension to exclude dead cells.

Cell sorting was performed on a FACS AriaFusion (Becton Dickinson) with BD FACSDiva software v8.0. Viable CD33+ cells were identified as single, PI-CD33+ cells. Cells were sorted with an 85 µm nozzle at 45 psi directly into FACS buffer (for ATAC- and ChIP-seq processing) and into Qiagen RLT Plus buffer + 1% β-mercaptoethanol (for RNA-seq processing).

4.3.2.3 Magnetic cell sorting, MACS and autoMACS

Alternatively, for AML patient samples with high cell number and some PDX samples (for these the cell number was never a limitation), magnetic sorting was applied to enrich for blasts.

In all cases, a positive selection was performed, using CD33 MicroBeads (human, Miltenyi Biotech) either manually with the QuadroMACS Separator (LS columns) or with the AutoMACS Pro Separator (program POSSEL_S, prioritizing the cell recovery) - both form Miltenyi Biotec. Cell labeling was carried out following the instructions of the manufacturer with minor changes: cell suspension was first filtered in a 40 µl strainer to avoid clumps and then pelleted and re-suspended in 60 µl of MACS buffer (PBS, 5% FCS h.i. and 2 mM EDTA) to which 40 µl of CD33 MicroBeads were added and then incubated for 15 min in the dark at 4°C. Cells were then washed to remove beads and finally re-suspended in 500 µl of MACS

buffer previous to proceed with the positive magnetic separation as indicated by the manufacturer.

When necessary, the Dead Cell Removal Kit (Miltenyi Biotec) was also used to deplete both dead, as well as dying cells.

4.3.3 Fast-ATAC-seq

Fast-ATAC-seq was done as described by Corces et al., 2016, which is an improved ATAC-seq protocol based on Buenrostro et al., 2013 to perform better when processing hematopoietic human cells. Briefly, 50×10^3 cells (viability > 90%) were pelleted (500 rcf, 4°C, 5 min) and washed once with PBS + protease inhibitor (EDTA free); after pelleting them again, supernatant was discarded using two pipetting steps being careful to not disturb the cell pellet. Cells were re-suspended in 50 µl of transposase mixture (25 µl 2x TD buffer, 2.5 µl Tn5 transposase -both from Illumina Nextera DNA Library Preparation Kit-, 0.25 µl 2% digitonin -Promega- and 22.25 µl H_2O). Reactions were incubated for 30 min at 37°C in a ThermoMixer shaking at low speed (300 rpm) and DNA was then purified using Qiagen PCR clean-up MinElute kit (Qiagen) and eluted in 10 µl of elution buffer (10 mM Tris pH 8.0). Transposed DNA was subsequently amplified in 50 µl reactions with custom primers as described by Buenrostro et al., 2015. After 4 cycles of amplification, libraries were then monitored with qPCR using for it 5 µl of PCR sample in a 15 µl reaction with the same primers. qPCR output was checked for the ΔRN; 0.25 ΔRN cycle number was used to estimate the number of additional cycles of the PCR reaction needed for the remaining PCR samples. Amplified libraries were purified with the Qiagen PCR clean-up MinElute Kit and size selected for fragments less than 600 bp using the Agencourt AMPure XP beads (Beckman Coulter). Library concentration was measured using the Qubit dsDNA HS Assay Kit in a Qubit 3.0 Fluorometer (Invitrogen) and library quality was finally assessed with the Agilent High Sensitivity DNA Kit on Agilent 2100 Bioanalyzer (Agilent Technologies). Libraries were deep sequenced on a HiSeq 1500 system according to the standard Illumina protocol for 50 bp single-end reads at the Laboratory for Functional Genome Analysis (LAFUGA) in the Gene Center (LMU).

4.3.4 Omni-ATAC-seq

Improved ATAC-seq protocol, based on Corces et al., 2017, which uses detergents after cell lysis to remove mitochondria from the transposition reaction, as well as PBS to improve the signal-to-noise ratio. In brief, 50×10^3 cells (viability > 90%) were pelleted (500 rcf, 4°C, 5 min) and re-suspended in ATAC re-suspension buffer (10 mM Tris-HCl pH 7.4, 10 mM NaCl, 3 mM $MgCl_2$) supplemented with 0.1% NP-40, 0.1% Tween-20 and 0.01% digitonin for lysis, incubated on ice for 3 min and then 1 ml of ATAC re-suspension buffer supplemented with 0.1% Tween-20 was added and the whole reaction was centrifuged (500 rcf, 4°C, 10 min) to collect nuclei. Nuclei were subsequently re-suspended in 50 µl of transposase reaction containing 25 µl of in-house made (Appendix B) 2× TD buffer (20 mM Tris-HCl pH 7.6, 10 mM $MgCl_2$, 20% dimethyl formamide, H_2O), 2.5 µl Tn5 transposase, 5.25 µl H_2O, 16.5 µl PBS, 0.25 µl of 2% digitonin and 0.5 µl of 10% Tween-20. Reactions were incubated for 30 min at 37°C in a ThermoMixer shaking at high speed (900 rpm) and DNA was then purified using Qiagen PCR clean-up MinElute kit. The library preparation from transposed DNA was performed as described in the previous section and finally libraries were deep sequenced in LAFUGA (50 bp, single-end reads).

Note: for adherent cells, as it was the case in the collaboration described in Section 2.2, cells were trypsinized for 5 min and then collected using a cell scraper and further pelleted as previously described. When cell viability was observed to be <90%, dead cells were depleted using the Dead Cell Removal Kit (Miltenyi Biotec), as described by the manufacturer.

Note II: as indicated in Section 2.1.3.2, the protocol version applied for processing samples was initially Fast-ATAC-seq, until we switched to the more recent and efficient Omni-ATAC version. Samples processed with the Fast-ATAC version comprised: all AML cell lines, all PDX samples and the AML patient-derived samples HS08 and HS09. The rest of the samples were processed with the Omni-ATAC version of the protocol.

4.3.5 Chromatin immunoprecipitation of histone modifications

Histone modifications ChIP-seq protocol is based on the original protocol by Dahl et al., 2008. Suspension cells were harvested and then cross-linked in fixation solution pre-warmed at 22°C, consisting either in cell medium (for AML cell lines) + 1% v/v methanol-free formaldehyde (ThermoFischer Scientific) or PBS + 10% FBS (for PDX cells, blast cells isolated from patients or mice cells isolated from mixed bone marrow chimeras) + 1% v/v methanol-free formaldehyde under gentle agitation for exactly 10 min at 23°C. Glycine at a final concentration of 125 mM was added and cells were further incubated for 5 min at RT to quench fixation under gentle agitation. Cells were washed twice with PBS + 10% v/v FCS after pelleting for 5 min at 1600 g and, at the final step, after discarding the supernatant, cell pellets consisting in 1-2 Mio cells were flash frozen (when working with limited cell number or with adherent cells, the procedure was slightly different, see next section and note at the end of this section).

Frozen cell pellets were re-suspended in 100 µl of lysis buffer (50 mM Tris-HCl pH 8.0, 10 mM EDTA, 1% SDS and 1x cOmplete EDTA-free protease inhibitor cocktail tablets, Roche) and incubated 5 min on ice. Then cells were transferred to a Covaris microTUBE AFA fiber pre-slit snap-cap (Covaris) for sonication in the Covaris S220 with the following settings: duty cycle, 2%; peak incident power, 105 Watts; cycles per burst, 200; temperature 4 °C; degassing mode, continuous; processing time was adjusted every time depending on the amount of cells sonicated to get a shearing with average sizes from 100 to 500 bp. After sonication, 900 µl of IP buffer (10 mM Tris-HCl pH 7.5, 1 mM EDTA, 0.5 mM EGTA, 1% Triton X-100, 0.1% SDS, 0.1% Na-deoxycholate, 140 mM NaCl and 1x protease inhibitor cocktail) was added to cells to reduce SDS concentration. Cells were centrifuged 10 min at 4°C using maximal speed. The supernatant containing diluted shared chromatin was transferred to a clean low-binding tube and the corresponding volume to 50×10^3 cells was set apart for DNA size check. The rest was aliquoted to have the chromatin of 300×10^3 cells per aliquot and flash frozen.

10 µl of Dynabeads Protein A (Thermo Fischer Scientific) were transferred to a 0.2 ml PCR tube. Beads were washed twice with 100 µl of ice cold IP buffer. Every time beads were bound to a magnetic rack (Thermo Fisher Scientific) previous to discard the supernatant. Finally, beads were re-suspended in 100 µl of IP buffer and 2.4 µg of anti-H3K27ac antibody (Diagenode, Lot. A.7071-001P) was added to the tube. Antibody was bound to the beads incubating them at 4°C during 1.5 h in a rotating wheel at 35 rpm. Beads-antibody complexes were washed once with IP buffer, re-suspended finally in 150 µl of sheared chromatin and incubated for 4 h at 4°C in a rotating wheel with 35 rpm. Immunoprecipitated chromatin

was then washed 4 times with IP buffer and one last time with washing buffer (10 mM Tris-HCl pH 8.0, 10 mM EDTA) for 10 min at 4°C and 20 rpm in the rotating wheel. Chromatin was eluted from the beads-antibody complexes re-suspending them in 70 µl of elution buffer (10 mM Tris-HCl pH 8.0, 300 mM NaCl, 5mM EDTA, 0.5% SDS) containing 20 µg of RNase A (Thermo Fisher Scientific) for 30 min at 37°C in a ThermoMixer (900 rpm) and then adding 40 µg of Proteinase K (Thermo Fisher Scientific) and incubating 1 h at 55°C followed by 8 h to overnight at 65°C at 900 rpm. Beads were finally bound to a magnetic rack (Thermo Fisher Scientific) and supernatant was transferred to a clean low-binding tube. Additional 30 µl of elution buffer were added to the beads for 1 min and then both eluates were combined and incubated with 20 µg of Proteinase K for 1 h at 55°C with 600 rpm to digest remaining proteins. Purification of DNA was performed using Agencourt AMPure XP beads (Beckman Coulter) at a sample-to-beads ratio of 1:2, according to the manufacturer's instructions, eluting in 11 µl of elution buffer (10 mM Tris pH 8.0). Samples were stored at -20°C until library preparation was performed.

Note: for adherent cells, as it was the case in the collaboration documented in Section 2.2, cells were cultured in 10 cm cell culture dishes and rinsed twice with RT PBS previous to fixation. Then 8 ml of pre-warmed fixation solution consisting in cell medium + 1% v/v methanol-free formaldehyde were added and dishes were incubated for 10 min at RT, as previously described. Glycine was then added at a final concentration of 125 mM to stop fixation and cells were further incubated for 5 min. Cells were rinsed twice with 10 ml of PBS supplemented with 10% v/v FCS and finally 1 ml of wash buffer was added to the dish and cells were harvested using a silicon cell scraper and transferred to a 15-ml falcon, containing 9 ml of wash buffer. Cells were pelleted for 5 min with 1600 g, supernatant was then discarded and cells were re-suspended in 1 ml of wash buffer, transferred to a 1.5-ml tube and pelleted again with the same conditions than previously. Finally, the supernatant was discarded and cell pellets were flash-frozen ant stored at -80°C, when not immediately further processed.

4.3.6 Low-input chromatin immunoprecipitation of histone modifications

For limited number of starting material ($\leq 10^6$ cells), the previously mentioned protocol was slightly modified. Cells were fixed as described before and then fixed cell pellet was re-suspended in 75 µl of lysis buffer (50 mM Tris-HCl pH 8.0, 10 mM EDTA, 0.3% SDS, 1x protease inhibitor cocktail) and incubated 5 min on ice. Cell lysate was transferred to a Covaris micro TUBE AFA Fiber and sonicated to shear the chromatin using same settings than previously reported for 25 to 30 min, depending on the initial number of cells processed. After sonication, sheared chromatin was transferred to a new tube and 75 µl of dilution buffer (1 mM EGTA, 300 mM NaCl, 2% Triton x-100, 0.2% sodium deoxycholate, 1x protease inhibitor cocktail) and the necessary amount of IP buffer to be able to make aliquots of 300 k cells in 150 µl were added to dilute SDS concentration. Chromatin was centrifuged 10 min at 4°C using maximal speed and the supernatant was aliquoted into clean PCR tubes.

Per IP, 10 µl of Dynabeads A and 2.4 µg of anti-H3K27ac ab were used. Pre-washed beads were incubated with ab for 1.5 h at 4°C in a rotating wheel (35 rpm). Chromatin from 300 k cells was pre-cleared by incubating it with 10 µl of pre-washed beads for the same time under same conditions. Pre-coated beads were then incubated with the pre-cleared chromatin for 4 h (4°C, 35 rpm). Immunoprecipitated chromatin was washed 6 times with IP buffer (increasing the number of washing steps contributed to reduce the background) and one

last time with washing buffer. If possible, two IPs were pooled together at this point to increase the DNA yield and de-crosslinking and further cleaning were performed as previously described.

4.3.7 ChIP library preparation and sequencing

ChIP-seq libraries were prepared with NEBNext Ultra II DNA Library Prep Kit for Illumina (New England BioLabs) according to the manual. For libraries purification and size selection (fragments size from 100 to 600 bp), AMPure XP beads were used adjusting the ratio sample:beads following manufacturer's recommendations. Sample concentration was measured using the Qubit dsDNA HS Assay kit in a Qubit 3.0 Fluorometer (Invitrogen) and library quality was finally assessed with the Agilent High Sensitivity DNA Kit on Agilent 2100 Bioanalyzer (Agilent Technologies).

Libraries were sequenced on a HiSeq 1500 system according to the standard Illumina protocol for 50bp single-end reads at LAFUGA in the Gene Center (LMU).

4.3.8 Bulk RNA library preparation

Bulk RNA-seq was performed by L. Wange (Anthropology and Human Genomics, Faculty of Biology, LMU Munich) following a bulk version of the mcSCRB-seq (molecular crowding single-cell RNA barcoding and sequencing) method (Bagnoli et al., 2018).

For RNA library preparation, samples were treated with 20 µg Proteinase K (Thermo Fisher Scientific, cat. no. AM2546) and 1 µl 25 mM EDTA at 50°C for 15 min followed by heat inactivation (75°C, 10 min). Samples were then cleaned using magnetic beads at a ratio of 1:2 of lysate to beads and treated with DNase I (Thermo Fisher Scientific, cat. no. EN0525) at 20°C for 10 min and heat inactivation at 65°C for 5 min. DNAse I treatment was performed on-beads by re-suspending the beads in a mix containing DNAse I, DNAse I Buffer as well as bead binding buffer (beads binding buffer 2x (total volume 5 ml): 1.1 g PEG, 1 ml 5M NaCl, 50 µl 1M, pH 8.0 Tris-HCl, 5 µl 10% Igepal, 25 µl sodium azide, double distilled H_2O) to reduce cleanup steps.

Samples were thus cleaned with the beads already present in the sample and RNA was eluted with the reverse transcription mix, consisting of 25 units Maxima H-enzyme (Thermo Fisher Scientific, cat. no. EP0753), 1x Maxima H-Buffer (Thermo Fisher Scientific), 1 mM each dNTPs (Thermo Fisher Scientific), 1 µM template-switching oligo (IDT), 0.2 µM barcoded oligo-dT primers (IDT). The reaction was incubated at 42°C for 90 min.

Following cDNA synthesis, the samples were pooled, cleaned and concentrated with magnetic beads at a 1:1 ratio and eluted in 17 µl of double distilled H2O. Residual primers were then digested using Exonuclease I (Thermo Fisher Scientific, cat. no. EN0581) at 37°C for 20 min followed by heat inactivation at 80°C for 10 min.

Pre-amplification mix, consisting of 1x KAPA HiFi Ready Mix (Roche, cat. no. 7958935001) and 0.2 µM SingV6 primer (IDT), was added to the sample pool and incubated as follows: initial denaturation at 98°C for 3 mins, denaturation at 98°C for 15 sec, annealing at 65°C for 30 sec, elongation at 68°C for 4 min, and a final elongation at 72°C for 10 min. Denaturation, annealing, and elongation were repeated for 14 cycles. DNA was then cleaned using magnetic beads at a 1:0.8 ratio of DNA to beads and eluted with 10 µl of double distilled H2O. The

quantity was assessed using a Quant-iT PicoGreen dsDNA assay kit (Thermo Fisher Scientific, cat. no. P11496) and the quality was assessed using an Agilent 2100 Bioanalyzer with a High Sensitivity DNA analysis kit (Agilent Technologies, cat. no. 5067-4626).

Following quality checks, tagmentation was performed using a Nextera XT Kit (Illumina, cat. no. FC-131-1096) with 0.8 ng/µl of the DNA in a 20 µl reaction. Next, 5 µl of 5x Phusion HF Buffer (NEB, cat. no. M0530S) were added to the reaction and incubated at RT for 5 min. The Nextera Index PCR was then setup by adding 15 µl of PCR mix consisting 0.375x Phusion HF Buffer, 0.8 U Phusion Polymerase, 0.2 mM dNTPs, 0.1 µM P5NextPT5 primer (IDT) and 0.1 µM i7 index primer (IDT) to a final volume of 40 µl. Index PCR was incubated as follows: gap fill at 72°C for 3 min, initial denaturation at 95°C for 30 sec, denaturation at 95°C for 10 sec, annealing at 63°C for 30 sec, elongation at 72°C for 1 min, and a final elongation at 72°C for 5 min. Denaturation, annealing, and elongation were repeated for 13 cycles.

Libraries were then loaded onto a 2% Agarose E-Gel EX (Invitrogen, cat. no. G401002) and were excised and cleaned using the Monarch DNA Gel Extraction Kit (NEB, cat. no. T1020). Finally, libraries' quality was quantified and qualified using the Agilent High Sensitivity DNA Kit on Agilent 2100 Bioanalyzer.

4.4 Analysis

Note: for details about the software and packages used in the analysis (references, versions and documentation), refer to Section 4.9, where all software used is listed.

4.4.1 ATAC-seq data analysis

4.4.1.1 Primary ATAC-seq analysis: reads alignment and peak calling

ATAC-seq data were first uploaded to the Galaxy server and the public server was used for demultiplexing them and generating the corresponding fastq files. Demultiplexed fastq files were then downloaded to the BMC HPC cluster and further analysis was performed there. Single-end reads were aligned to the human genome hg38 using Bowtie1 with options *-q -n 2 –best –chunkmbs 2000 -p 32 -m 1 -S*. Duplicated reads were removed using picard MarkDuplicates and unwanted reads, like those mapping to mitochondrial DNA, were filtered out from the output sam file. Needed files for further analysis were also generated: bam files, bed files and BigWig files. A first quality control was carried out by visual inspection of the mapped reads loading the BigWig file into the IGV genome browser. Moreover, Fastqc reports were generated as quality control for the sequencing data. Next, ATAC peaks over input background were identified using HOMER *findPeaks.pl* with option *-style factor*. The generated bed output from all samples (when analyzing groups of samples) were merged using bedtools and genomic feature annotation of merged found peaks was performed using HOMER *annotatePeaks.pl*.

Note: the primary analysis is common for all samples processed by ATAC-seq.

4.4.1.2 Transcription factor binding sites prediction

Transcription factor binding sites (TFBS) prediction analysis was performed usually within previously identified differential ATAC peaks (for whichever considered different conditions)

with HOMER *findMotifsGenome.pl*, mostly using the HOCOMOCOv11 motifs database and focusing mainly on the known motifs with the parameter *-mknown*.

4.4.1.3 PDX samples analysis

For the analysis of the PDX samples processed for ATAC-seq at three different treatment time points (as detailed in Section 2.1.2) primary analysis, as previously described, was first applied. The information generated in the *annotatePeaks* file was then used for creating in R a triple Venn diagram representing the total number of ATAC peaks found for each sample group. Then, HOMER-normalized ATAC data were used for dimensionality reduction analysis applying t-SNE and a 2D plot as well as a 3D plot were generated in R. A *SummarizedExperiment* object, including all the experimental metadata, was created from the raw *annotatePeaks* file and it was subsequently RPKM-normalized using the *edgeR* package. This *SummarizedExperiment* object was then used for a time course clustering analysis using the function *timeclust* from the *TCseq* R package.

4.4.1.4 AML patient-derived samples analysis

As for PDX samples, AML human samples processed for ATAC-seq were primarily analyzed as previously described. An R *SummarizedExperiment* object, including all the experimental metadata, was created from the raw *annotatePeaks* file. Previous to PCA, a scree-test was performed to assess the contribution of each individual principal component to the data variability, as well as the cumulative proportion of total variance. The correspondent screeplot including the first fifteen PCs was generated, as well. PCA plots were created with the R package *PCAtools* using the function *pairsplot* for PC1-PC6, which presented a cumulative proportion of total variance < 75%. This analysis was performed for the total amount of human samples and for a subset not including healthy donor samples.

4.4.2 ChIP-seq data analysis

4.4.2.1 Primary ChIP-seq analysis: reads alignment and peak calling

Similarly than for ATAC-seq data, H3K27ac ChIP-seq data were first uploaded to the Galaxy server for demultiplexing. Demultiplexed fastq files were then downloaded to the BMC HPC cluster for further analysis. Single-end reads were aligned to the human genome hg38 using Bowtie1 with options *-q -n 2 –best –chunkmbs 2000 -p 32 -m 1 -S*. Duplicated reads were removed using picard MarkDuplicates and unwanted reads were filtered out. bam, bed and BigWig files were generated and BigWig files were used to check visually data quality in the IGV genome browser. Fastqc reports for further quality control of the sequenced data were also generated. Next, H3K27ac ChIP-seq peaks over input background were identified using HOMER *findPeaks.pl* with option *-style histone* and genomic feature annotation of found peaks was performed using HOMER *annotatePeaks.pl*. When analyzing groups of samples, previous to the annotation, bed files resulting from peak calling were merged using bedtools.

Note: the primary analysis is common for all samples processed by H3K27ac ChIP-seq.

4.4.2.2 Super-enhancers analysis of AML patient-derived samples with ROSE2

For the analysis of super-enhancers (SEs), ROSE2 (Rank Ordering of Super-Enhancers), available at https://github.com/BradnerLab/pipeline, was used to map H3K27ac SEs. This analysis was applied to five paired patient-derived samples and two healthy donor samples, for which it was possible to collect enough cells for performing H3K27ac ChIP-seq. Bed files containing enhancers positions were generated as previously described and then converted into gff format. For each individual patient and condition, this gff file and the corresponding bam file with aligned reads information were used as input for ROSE2 with default parameters, except for the exclusion of regions closer than 2500 bp to TSSs, to consider only enhancer and not promoter regions. A 12.5 kb stitching window was defined to connect proximal clusters of H3K27ac peaks into adjacent enhancer regions.

4.4.3 Integrated data analysis of AML patient-derived samples

AML patient-derived samples were processed for three different assays, namely, ATAC-seq, (bulk) RNA-seq and H3K27ac ChIP-seq. In this section, the integrated data analysis performed to combine in a meaningful way all data generated is described.

4.4.3.1 Integration of ATAC- and RNA-seq data

Note: this analysis was conducted in patients P2-P12, due to the lack of RNA-seq data for patient P1.

Previously, two R *SummarizedExperiment* objects including all relevant metadata, were created from the HOMER raw *annotatePeaks* ATAC-seq derived file as well as from the RNA-seq derived count matrix, directly provided by L. Wange (Anthropology and Human Genomics, Faculty of Biology, LMU Munich).

Both *SummarizedExperiment* R objects were then normalized by using the DESeq2 R package, after organizing and structuring the original information with the *GenomicRanges* R package. For each single patient, differential ATAC peaks between both disease stages, primary diagnosis and relapse, were determined. To select differential peaks, an arbitrary threshold was set at a minimal fold enrichment of 2 for one condition over the other. Simultaneously, the same strategy was applied to the RNA-seq data to determine differentially expressed genes. Then the *rGREAT* (Genomic Regions Enrichment of Annotations Tool) R package was applied to find genes associated to the relevant genomic regions provided by the previously found differential ATAC peaks, considering all possible features regardless to their distance to TSS. The generated genes list was then filtered for keeping only differentially expressed genes for both disease stages. Next, the *l2p* (list-to-pathway) R package was applied, for gene set enrichment analysis over the genes found associated to the selected differential genomic regions, with the following databases: GO, KEGG, PANTH, REACTOME, C5 and C6. For this analysis, the genome annotation *org.Hs.e.db* was considered. Obtained pathways were then filtered for pval < 0.05. Bed files containing the determined genomic features of interest (chromosome, start position and end position) were also created for both disease stages, to be used as input for transcription factor binding sites detection.

TF motif prediction was performed as previously described, and lists of known and *de novo* HOMER predicted motifs were generated for each patient, associated to the found differential genomic features between both disease stages.

For detection of transcription factor binding sites (TFBSs), the *Regulatory Genomics Toolbox* python library tool *HINT* (Hmm-based IdeNtification of Transcription factor footprints) was used with standard parameters for ATAC-seq generated data, considering the human genome hg38. As input, bed files were generated containing the previously determined differential genomic features for each particular condition and patient. Motif matching for footprints was then assessed using the HOCOMOCO database. For it, a merged bed file was generated with the bedtools function *intersect* for each patient and each disease status, containing the previous HINT output and the found differential genomic features. Next, the HINT function *rgt-hint differential* was applied to compare and determine differential TF binding sites in both disease stages for each patient.

4.4.3.2 Global data integration of ATAC-, H3K27ac ChIP- and RNA-seq data: transcriptional regulatory networks

For the six patients we had a both ATAC- and H3K27ac ChIP-seq data (P1, P7, P9, P10, P11 and P12), we performed data integration using the coltron phyton package, available at https://pypi.org/project/coltron, for building transcriptional regulatory networks. It quantifies the degree of inward (in) and outward (out) regulation of SE-regulated TFs. Based on motif analysis, in-degree for a given TF is defined as the number of nodes with motifs found in any enhancer assigned to regulate that TF, that is, the number of SE-regulated TFs bound to its proximal SE. Out-degree is defined as the number of nodes in the network which is regulated by an enhancer containing a motif for a certain TF, or in other words, the number of TF-associated SEs bound by a given SE-regulated TF (Chen et al., 2020; Ott et al., 2018). The coltron python package presents an increased capacity to predict motif-based TF binding, since it scans TF motifs within nucleosome-free-regions instead of considering whole SE domains (Ramsey et al., 2010). The coltron package relies on following dependencies: Bamliquidator (Lin et al., 2016, 1.2.0), Samtools (Li et al., 2009, 0.1.19), FIMO (Grant et al., 2011, 4.91) and Networks (Hagberg et al., 2020, 1.8.1). Coltron was applied to each sample in a recursive script with default parameters, as described in the package documentation, with following inputs: *-e*, ROSE generated enhancer table (*"_AllEnhancers.table.txt"*), *-b*, bam file containing the identified super-enhancers, *-g* HG38, *-s*, bed file containing the ATAC peaks to use as subregions for motif search. Among the coltron outputs, the file *"_DEGREE_TABLE.txt"*, containing the identified TFs, the number of in- and out-degree, as well as the number of total connections, was used for identifying relevant TFs associated to either primary diagnosis or relapse. For doing so, a matrix was created containing the samples in the columns and in the rows all TFs found for any of the analyzed samples. For each sample and each TF, the number of total connections was considered, being zero if the factor was not detected in a particular sample. The threshold for counting a specific TF was set arbitrarily at >60 total connections, so TFs counting with less than 60 total connections were discarded. The matrix was then subset for each patient and TFs unique for either diagnosis or relapse were ranked according to their number of total connections. For each sample, the top six unique TFs (not appearing in the other disease stage from the same patient) were finally selected and common factors among either the primary diagnosis or the relapse from the analyzed patients subset were evaluated. For evaluating the expression of the common TFs either in diagnosis or relapse over the entire patient cohort, a heatmap was created from the corresponding RNA-seq data using the R package *ComplexHeatmap* (Gu et al., 2016).

For pathway analysis, the HOCOMOCO database was subset for each individual sample,

selecting only the motifs corresponding to the previously assessed top differential TFs. Then peak annotation using HOMER *annotatePeaks.pl* was performed considering as input the ATAC peaks bed file generated during the primary analysis and as database, the corresponding HOCOMOCO subset. This resulted in a txt file consisting of the annotation of those peaks containing the motifs of interest. From the txt file containing the peak annotation, only genes whose TSSs were closer than 50 kb to any peak, were considered. For each patient and each disease stage, a txt file was created, including the list of genes associated to the peaks containing the motifs of the previously identified differential TFs.

Next, excluding samples from patient P1 (for which no RNA-seq data were available), for the rest of the samples from the patients' subgroup, the already mentioned gene lists were further filtered so that only differentially expressed genes for both diseases stages were kept in each case. Finally, the R package l2p was applied for pathway analysis. Additionally, emerged pathways were screened for terms of especial interest, like apoptosis-related terms, and common pathways shared by different patients at the same disease stage were interrogated.

Appendix A

Supplementary information

A.1 AML patient-derived samples processed - info sheet

sample ID/ clinical ID	disease status	time to relapse	tissue origin	blast count	karyotype	age	gender	sorting marker	S-P sep	experiments performed	additional considerations
HS09 17-1043	D		BM	85%	normal	52	F	CD33	Hep	H3K27ac ChIP-seq Fast-ATAC-seq	Mol: NPM1, FLT3, ITD
HS08 17-1080	R	< 1 y	BM	48%	normal			CD33	Hep	H3K27ac ChIP-seq Fast-ATAC-seq	*NA*
HS14 16-820	D		BM	*NA*	normal	54	F	CD33	Hep	Omni-ATAC-seq RNA-seq	Mol: NPM1
HS15 17-1079	R	1 y	BM	*NA*	normal			CD33	Hep	Omni-ATAC-seq RNA-seq	*NA*

TABLE A.1: continued on next page

sample ID/ clinical ID	disease status	time to relapse	tissue origin	blast count	karyotype	age	gender	sorting marker	S-P sep	experiments performed	additional considerations
HS16 15-776	D		BM	*NA*	normal	59	M	CD33	Hep	Omni-ATAC-seq RNA-seq	Mol: no know mutations
HS17 16-900	R	1 y	BM	*NA*	normal			CD33	Hep	Omni-ATAC-seq RNA-seq	*NA*
HS18 15-1309	D		BM	71%	normal	74	M	CD33	*NA*	Omni-ATAC-seq RNA-seq	Mol: NPM1-, FLT3-ITD- FLT3-TKD-, MLL-PTD-
HS19 16-2607	R	1 y	BM	38%	normal			CD33	*NA*	Omni-ATAC-seq RNA-seq	Mol: NPM1-, FLT3-ITD-, FLT3-TKD-, KMT2A-PTD- IDH1-, IDH2-
HS20 16-1225	D		BM	44%	complex	62	*NA*	CD33	EDTA	Omni-ATAC-seq RNA-seq	Mol: FLT3-TKD-, NPM1-, FLT3-ITD-, KMT2A-PTD-
HS21 16-3628	R	< 1 y	BM	56%	complex			CD33	EDTA	Omni-ATAC-seq RNA-seq	patient never reached complete remission
HS22 16-1692	D		BM	68%	normal	33	*NA*	CD33	*NA*	Omni-ATAC-seq RNA-seq	Mol: NPM1+, FLT3-TKD+ IDH1/IDH2-, FLT3-ITD- KMT2A-PTD-
HS23 17-0097	R	1 y	BM	39%	normal			CD33	*NA*	Omni-ATAC-seq RNA-seq	Mol: NPM1+, FLT3-TKD+ KMT2A-PTD-, IDH1-, IDH2-

TABLE A.1: continued on next page

sample ID/ clinical ID	disease status	time to relapse	tissue origin	blast count	karyotype	age	gender	sorting marker	S-P sep	experiments performed	additional considerations
HS25 16-3421	D		BM	72%	normal	75	*NA*	CD33	*NA*	H3K27ac ChIP-seq Omni-ATAC-seq RNA-seq	Mol: NPM1-A+, FLT3-ITD (6bp), FLT3-TKD-, KMT2A-PTD- IDH1+, IDH2-, CEBPA-
HS26 17-4178	R	1 y	BM	18%	normal			CD33	*NA*	H3K27ac ChIP-seq Omni-ATAC-seq RNA-seq	Mol: NPM1+, IDH1+, IDH2-, FLT3-TKD-
HS28 09-2782	D		BM	74%	*NA*	63	F	CD33	Hep	Omni-ATAC-seq RNA-seq	*NA*
HS29 16-0193	R	7 y	BM	76%	*NA*			CD33	Hep	Omni-ATAC-seq RNA-seq	*NA*
HS31 15-4027	D		PB	94%	*NA*	77	M	CD33	Hep	H3K27ac ChIP-seq Omni-ATAC-seq RNA-seq	*NA*
HS32 16-1976	R	1 y	PB	50%	*NA*			CD33	Hep	H3K27ac ChIP-seq Omni-ATAC-seq RNA-seq	*NA*
HS37 10-0573	D		PB	8%	*NA*	37	F	CD33	Hep	H3K27ac ChIP-seq Omni-ATAC-seq RNA-seq	*NA*
HS38 15-0646	R	6 y	BM	68%	*NA*			CD33	Hep	H3K27ac ChIP-seq Omni-ATAC-seq RNA-seq	*NA*

TABLE A.1: continued on next page

sample ID/ clinical ID	disease status	time to relapse	tissue origin	blast count	karyotype	age	gender	sorting marker	S-P sep	experiments performed	additional considerations
HS40 16-0845	D		BM	93%	*NA*	58	F	CD33	EDTA	H3K27ac ChIP-seq Omni-ATAC-seq RNA-seq	*NA*
HS41 17-2100	R	1 y	BM	99%	*NA*			CD33	EDTA	H3K27ac ChIP-seq Omni-ATAC-seq RNA-seq	*NA*
HS43 16-3916	D		BM	96%	*NA*	65	M	CD34	EDTA	H3K27ac ChIP-seq Omni-ATAC-seq RNA-seq	*NA*
HS44 18-2177	R	2 y	BM	82%	*NA*			CD34	EDTA	H3K27ac ChIP-seq Omni-ATAC-seq RNA-seq	*NA*

TABLE A.1: List of AML human-derived samples processed. All AML human-derived samples used in this work were provided by AG Götze - Klinikum rechts der Isar, III; Medizinische Klinik, TUM - and AG Metzeler, AG Spiekermann and AG Subklewe - Medizinische Klinik und Poliklinik III, LMU. Samples are grouped by pairs (primary diagnosis and relapse) isolated from the same patient. Abbreviations: D, first diagnosis; R, relapse; y, year(s); BM, bone marrow; PB, peripheral blood; F, female; M, male; S-P sep, serum-plasma separation; Hep, heparin; Mol, molecular genetics; *NA*, information not available.

A.2 Healthy donor samples processed - info sheet

sample ID	tissue of origin	age	gender	sorting marker	experiments performed
HS01	FH	*NA*	*NA*	CD34	H3K27ac ChIP-seq
HS03	FH	*NA*	*NA*	CD34	H3K27ac ChIP-seq
HS24	BM	30	F	CD34	Omni-ATAC-seq RNA-seq
HS27	BM	*NA*	*NA*	CD34	Omni-ATAC-seq RNA-seq
HS33	BM	71	M	CD34	Omni-ATAC-seq RNA-seq
HS36	BM	66	F	CD34	Omni-ATAC-seq RNA-seq
HS39	BM	42	F	CD34	Omni-ATAC-seq RNA-seq
HS42	BM	60	F	CD34	Omni-ATAC-seq RNA-seq
HS45	BM	66	F	CD34	Omni-ATAC-seq RNA-seq

TABLE A.2: List of healthy donor samples processed. Healthy donor human samples used in this work were provided by AG Götze - Klinikum rechts der Isar, III; Medizinische Klinik, TUM - and AG Subklewe - Medizinische Klinik und Poliklinik III, LMU. Abbreviations: FH, femoral head; BM, bone marrow; F, female; M, male; *NA*, information not available.

A.3 AML patient-derived samples - additional information

sample ID	number of frozen cells [$\times 10^6$]	cell viability after thawing	final number of CD33+ cells recovered [$\times 10^6$]
HS09	10*	100%*	1.07
HS08	1.7*	100%*	0.465
HS14	100	58%	0.2
HS15	45	14%	0.066
HS16	38	55%	0.12
HS17	300	36%	6
HS18	30	34%	0.187
HS19	30	35%	0.176
HS20	40	58%	2.1
HS21	30	50%	1.4
HS22	30	20%	0.302
HS23	30	58%	0.182
HS25	48	76%	8.6
HS26	12	84%	0.674
HS28	30	15%	0.015
HS29	24	88%	0.17
HS31	34	82%	11
HS32	65	74%	0.45
HS37	21	83%	0.66
HS38	42	62%	3.2
HS40	20	81%	3.2
HS41	21	76%	3.1
HS43	50	19%	0.66
HS44	33	89%	0.32

TABLE A.3: Additional information about the AML patients-derived samples received. Samples are grouped by pairs (primary diagnosis and relapse, in this order) isolated from the same patient. * Samples HS09 and HS08 were shared with another research group and we received the indicated number of cells already thawed with depleted dead cells; these paired samples were the only ones sorted with AutoMACS as proof of principle, all the rest samples were FACS sorted.

A.4 Supplementary information concerning the analysis of AML patient-derived samples

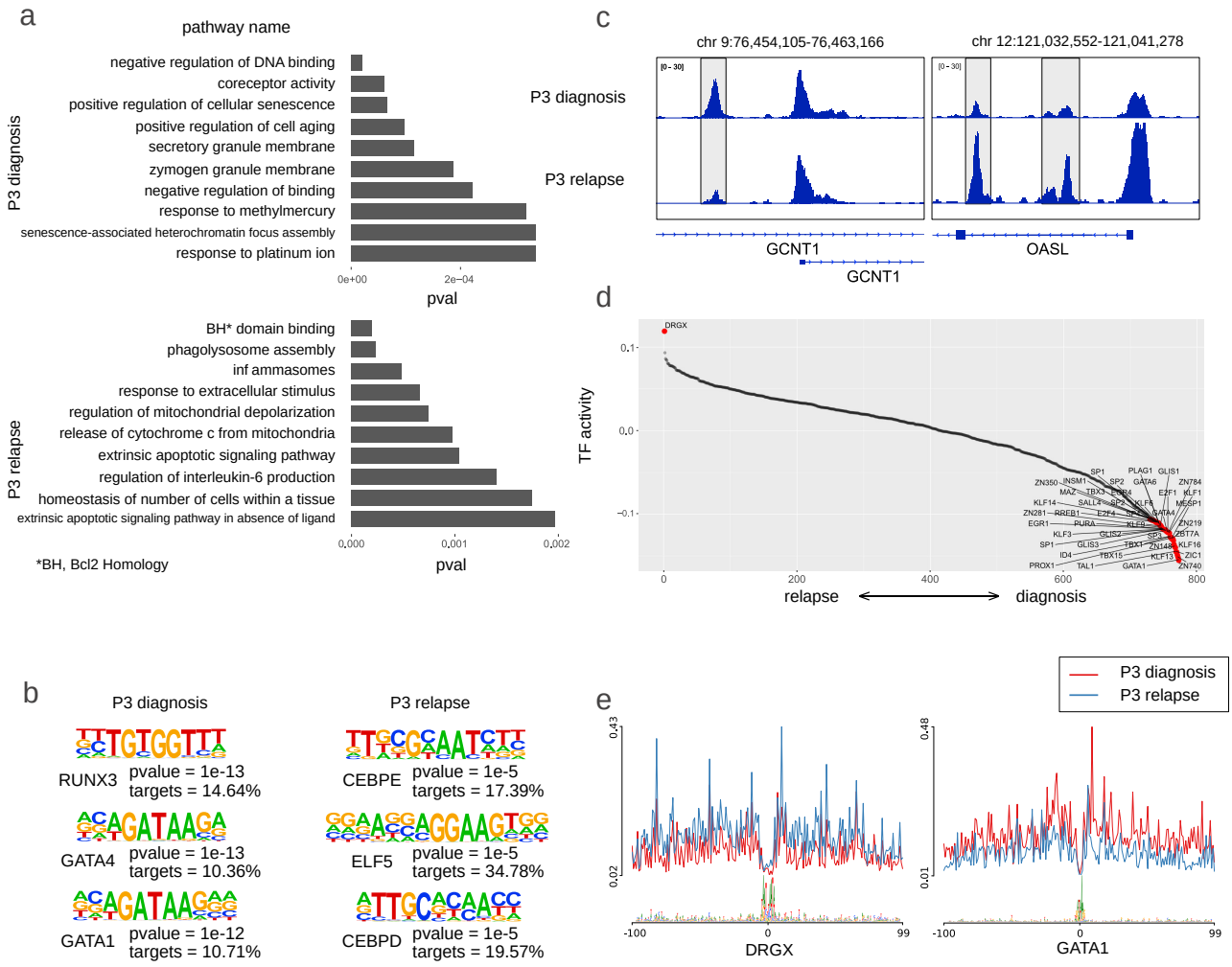

FIGURE A.1: Integrative analysis of ATAC-seq and RNA-seq data for patient 3 (P3). a, *l2p* analysis showing the top 10 pathways associated with differential genomic features in combination with differentially expressed genes for diagnosis (top) and relapse (bottom). b, Top three HOMER *known* predicted transcription factor motifs for assessed differential genomic features. c, IGV tracks showing differential ATAC peaks around loci *GCNT1* and *OASL* for diagnosis and relapse. d, Waterfall plot based on HINT TF differential activity. Each dot represents a TF; TFs showing significant differences in activity score (pval < 0.05) are labeled and highlighted in red. e, Detail of ATAC-seq profiles of diagnosis and relapse for motifs DRGX and GATA1.

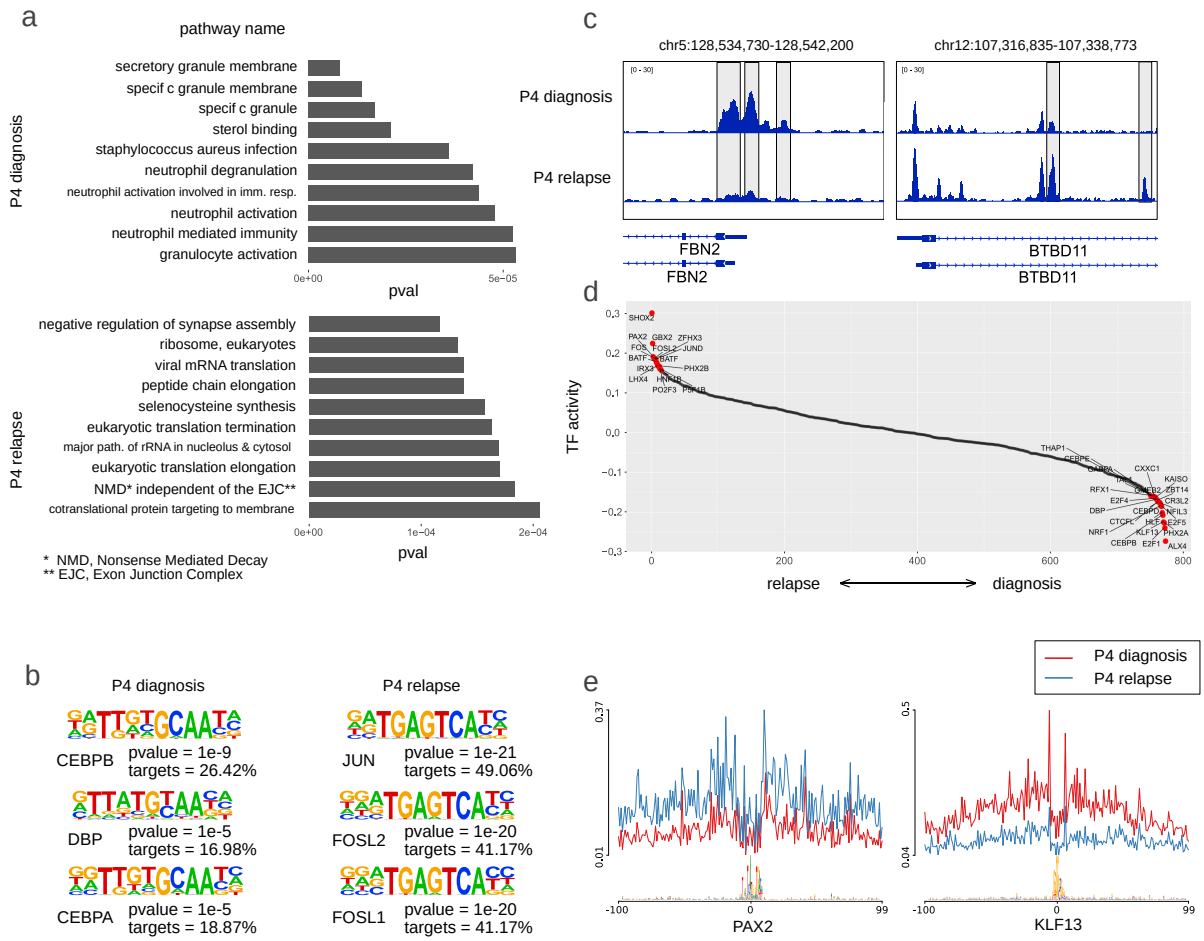

FIGURE A.2: Integrative analysis of ATAC-seq and RNA-seq data for patient 4 (P4). a, *l2p* analysis showing the top 10 pathways associated with differential genomic features in combination with differentially expressed genes for diagnosis (top) and relapse (bottom); abbreviations: *imm.*, immune; *resp.*, response; *path.*, pathway. b, Top three HOMER *known* predicted transcription factor motifs for assessed differential genomic features. c, IGV tracks showing differential ATAC peaks around loci *FBN2* and *BTBD11* for diagnosis and relapse. d, Waterfall plot based on HINT TF differential activity. Each dot represents a TF; TFs showing significant differences in activity score (pval < 0.05) are labeled and highlighted in red. e, Detail of ATAC-seq profiles of diagnosis and relapse for motifs PAX2 and KLF13.

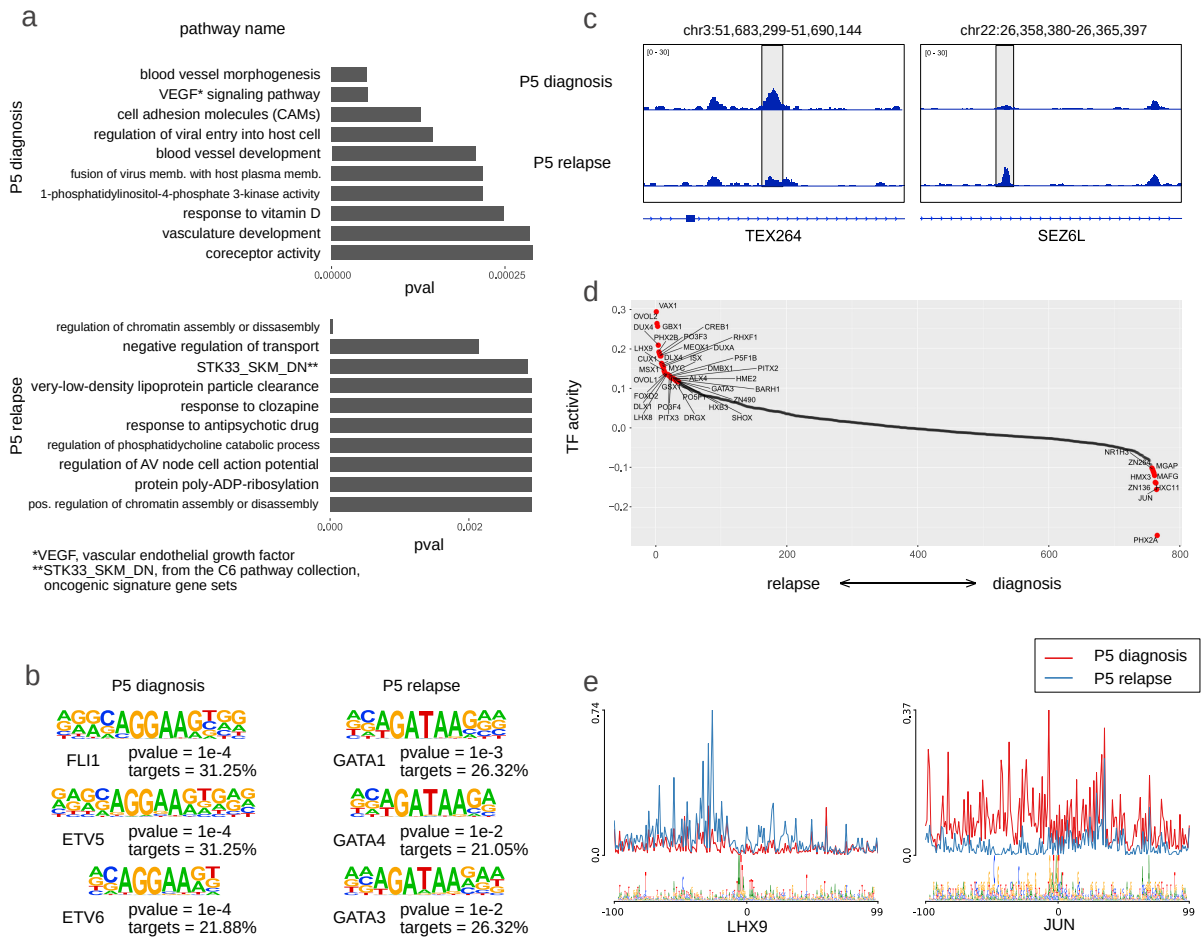

FIGURE A.3: Integrative analysis of ATAC-seq and RNA-seq data for patient 5 (P5). a, *l2p* analysis showing the top 10 pathways associated with differential genomic features in combination with differentially expressed genes for diagnosis (top) and relapse (bottom); abbreviations: *memb.*, membrane; *pos.*, positive. b, Top three HOMER *known* predicted transcription factor motifs for assessed differential genomic features. c, IGV tracks showing differential ATAC peaks around loci *TEX264* and *SEZ6L* for diagnosis and relapse. d, Waterfall plot based on HINT TF differential activity. Each dot represents a TF; TFs showing significant differences in activity score (pval < 0.05) are labeled and highlighted in red. e, Detail of ATAC-seq profiles of diagnosis and relapse for motifs LHX9 and JUN.

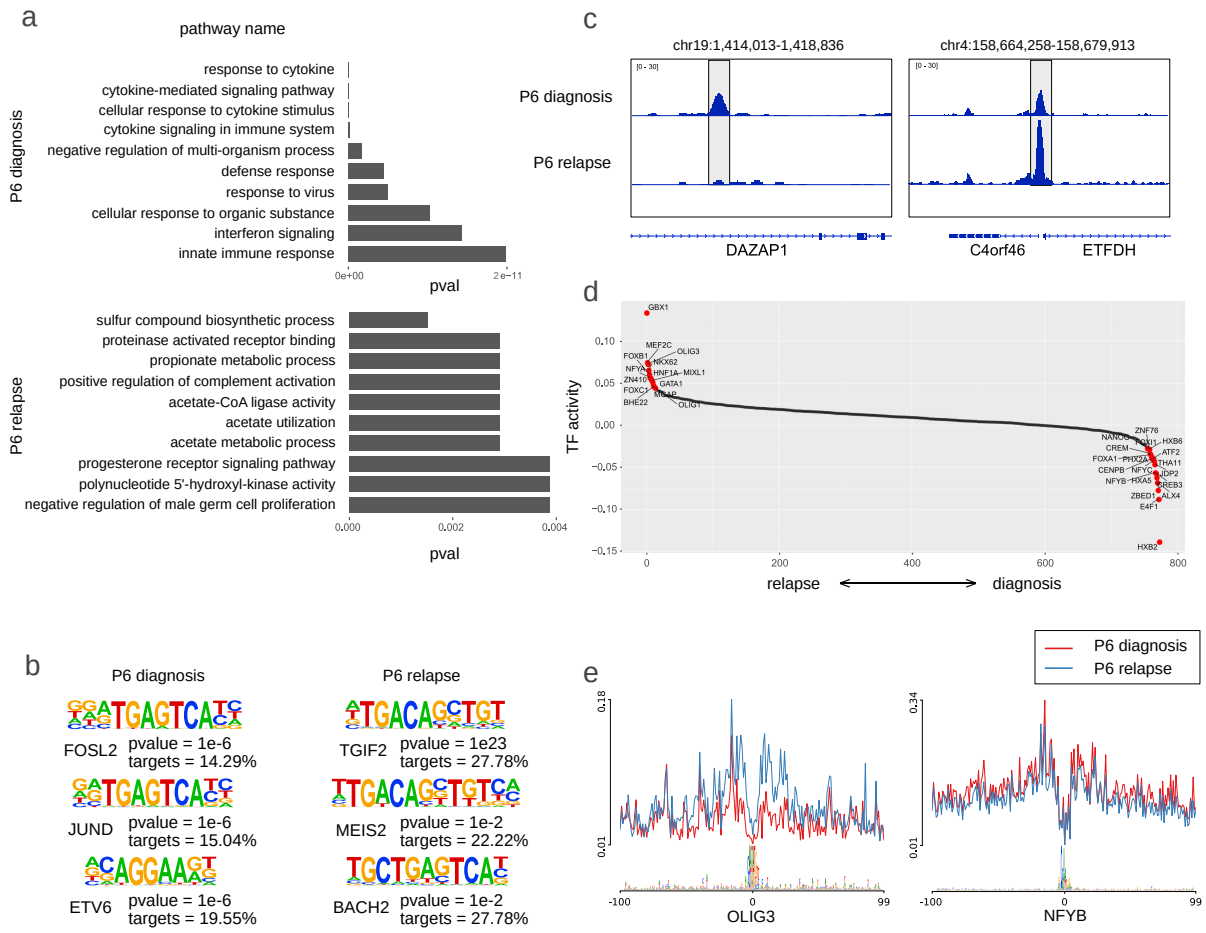

FIGURE A.4: Integrative analysis of ATAC-seq and RNA-seq data for patient 6 (P6). a, *l2p* analysis showing the top 10 pathways associated with differential genomic features in combination with differentially expressed genes for diagnosis (top) and relapse (bottom). b, Top three HOMER *known* predicted transcription factor motifs for assessed differential genomic features. c, IGV tracks showing differential ATAC peaks around loci *DAZAP1* and *ETFDH* for diagnosis and relapse. d, Waterfall plot based on HINT TF differential activity. Each dot represents a TF; TFs showing significant differences in activity score (pval < 0.05) are labeled and highlighted in red. e, Detail of ATAC-seq profiles of diagnosis and relapse for motifs OLIG3 and NFYB.

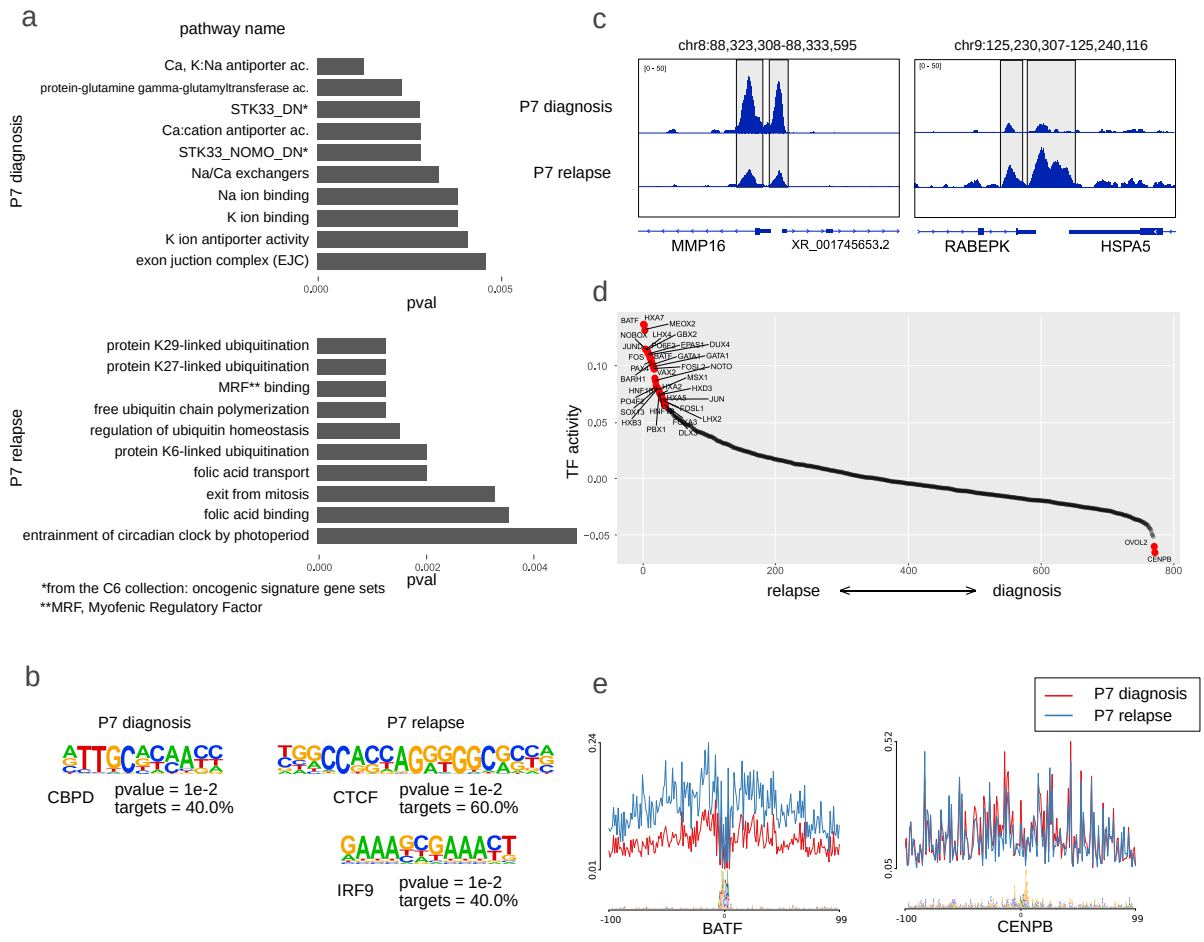

FIGURE A.5: Integrative analysis of ATAC-seq and RNA-seq data for patient 7 (P7). a, *l2p* analysis showing the top 10 pathways associated with differential genomic features in combination with differentially expressed genes for diagnosis (top) and relapse (bottom); abbreviations: *ac.*, activity. b, HOMER *known* predicted transcription factor motifs for assessed differential genomic features. c, IGV tracks showing differential ATAC peaks around loci *MMP16*, *RABEPK* and *HSPA5* for diagnosis and relapse. d, Waterfall plot based on HINT TF differential activity. Each dot represents a TF; TFs showing significant differences in activity score (pval < 0.05) are labeled and highlighted in red. e, Detail of ATAC-seq profiles of diagnosis and relapse for motifs BATF and CENPB.

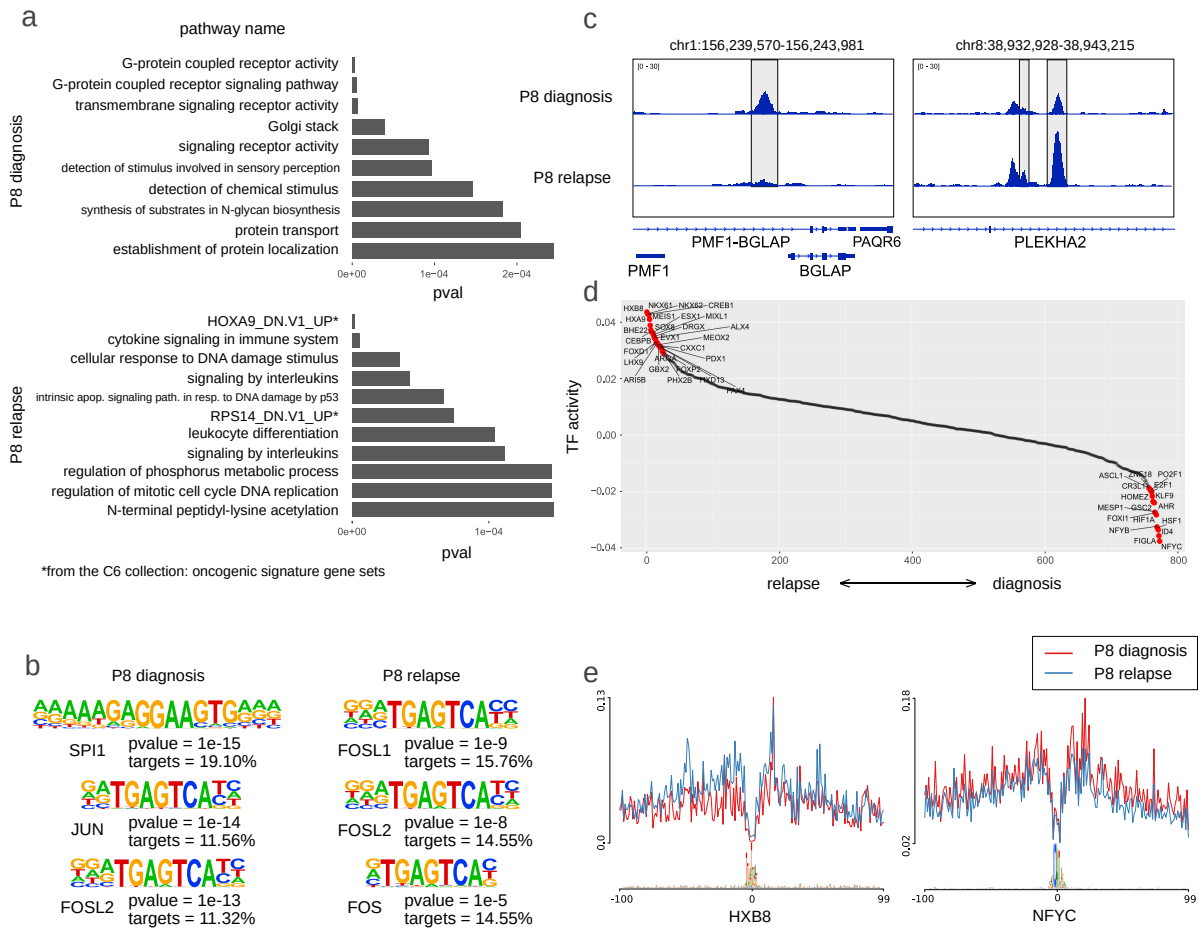

FIGURE A.6: Integrative analysis of ATAC-seq and RNA-seq data for patient 8 (P8). a, *l2p* analysis showing the top 10 pathways associated with differential genomic features in combination with differentially expressed genes for diagnosis (top) and relapse (bottom); abbreviations: *apop.*, apoptotic; *path.*, pathway; *resp.*, response. b, Top three HOMER *known* predicted transcription factor motifs for assessed differential genomic features. c, IGV tracks showing differential ATAC peaks around loci *BGLAP* and *PLEKHA2* for diagnosis and relapse. d, Waterfall plot based on HINT TF differential activity. Each dot represents a TF; TFs showing significant differences in activity score (pval < 0.05) are labeled and highlighted in red. e, Detail of ATAC-seq profiles of diagnosis and relapse for motifs HXB8 and NFYC.

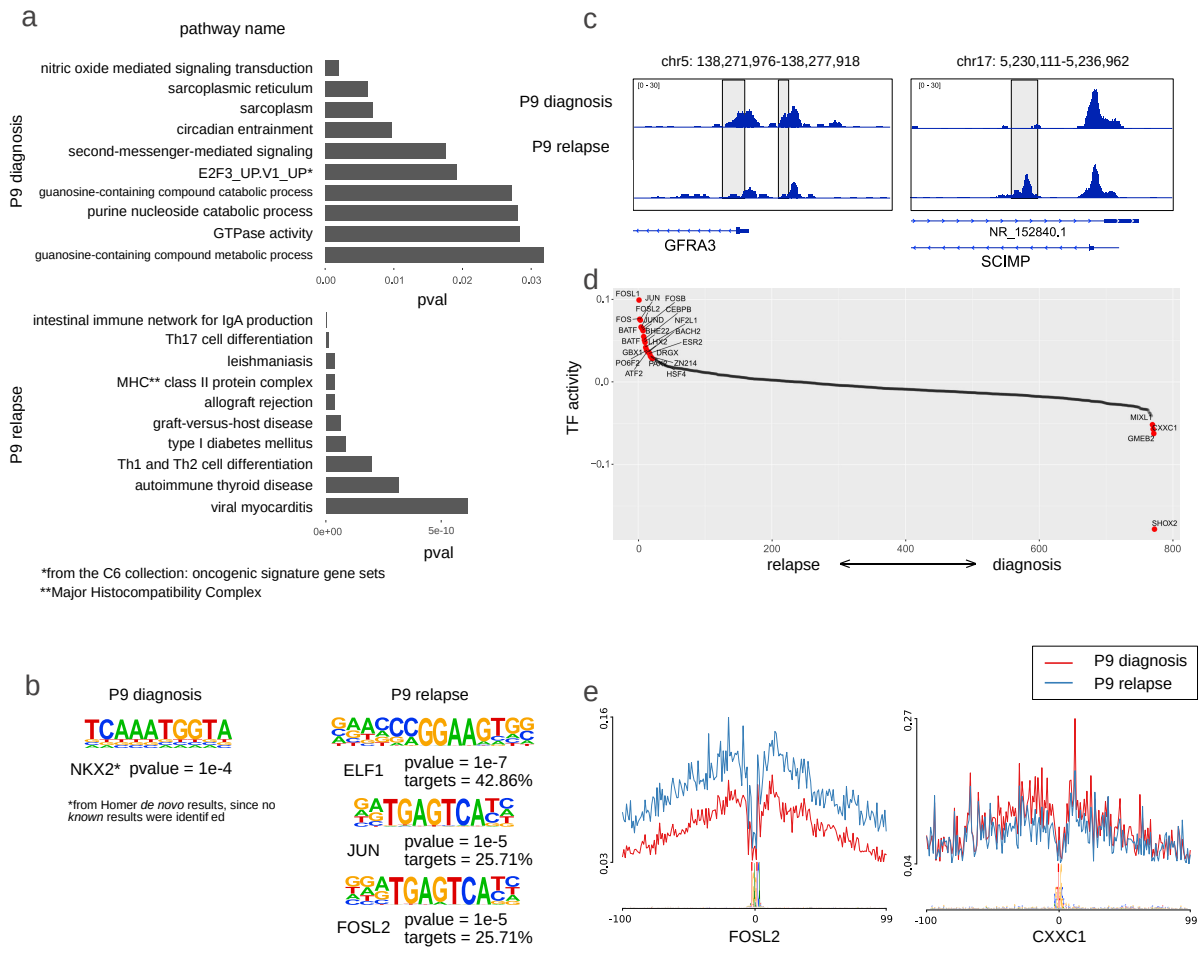

FIGURE A.7: Integrative analysis of ATAC-seq and RNA-seq data for patient 9 (P9). a, *l2p* analysis showing the top 10 pathways associated with differential genomic features in combination with differentially expressed genes for diagnosis (top) and relapse (bottom). b, Top three HOMER *known* predicted transcription factor motifs for assessed differential genomic features. c, IGV tracks showing differential ATAC peaks around loci *GFRA3* and *SCIMP* for diagnosis and relapse. d, Waterfall plot based on HINT TF differential activity. Each dot represents a TF; TFs showing significant differences in activity score (pval < 0.05) are labeled and highlighted in red. e, Detail of ATAC-seq profiles of diagnosis and relapse for motifs FOSL2 and CXXC1.

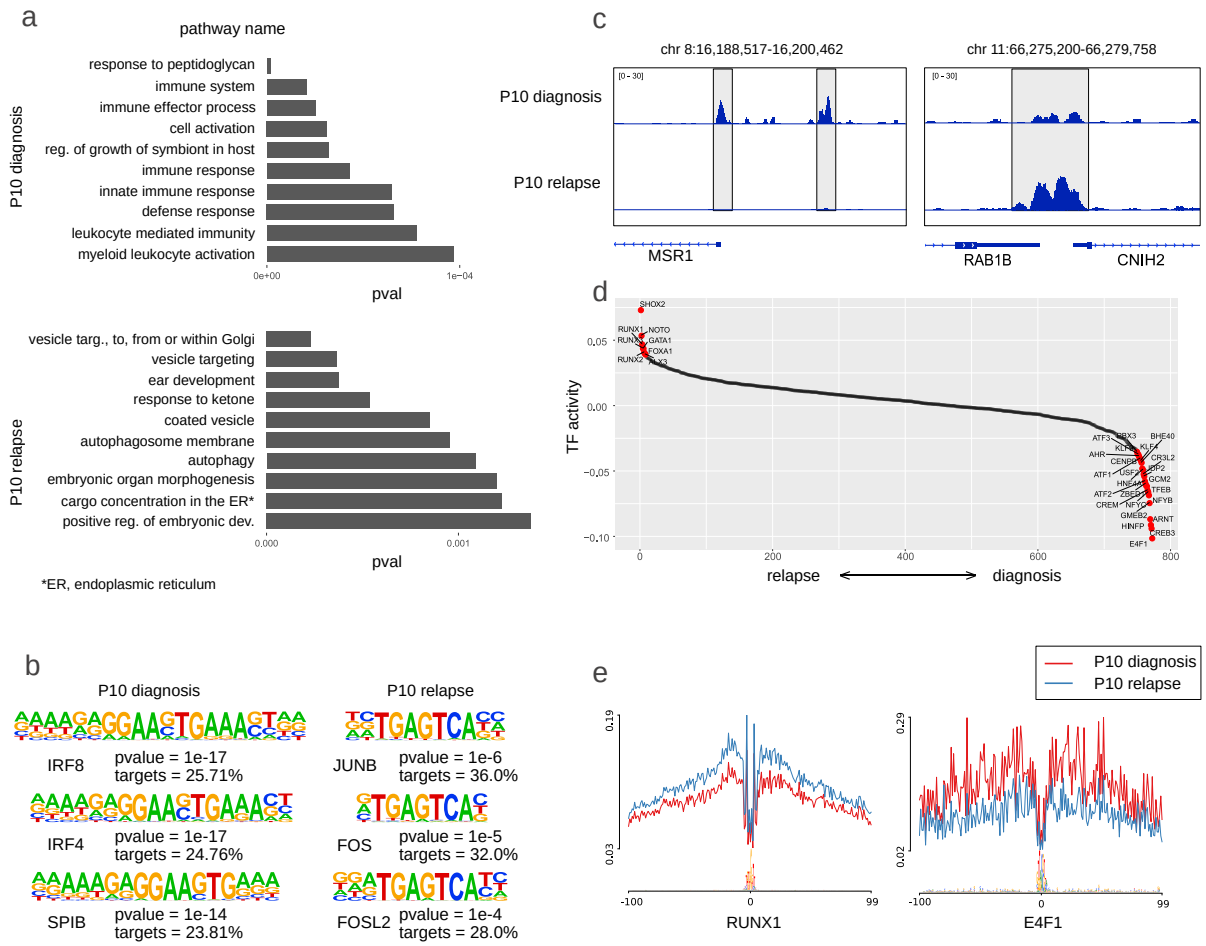

FIGURE A.8: Integrative analysis of ATAC-seq and RNA-seq data for patient 10 (P10). a, *l2p* analysis showing the top 10 pathways associated with differential genomic features in combination with differentially expressed genes for diagnosis (top) and relapse (bottom); abbreviations: *targ.*, targeting; *reg.*, regulation; *dev.*, development. b, Top three HOMER *known* predicted transcription factor motifs for assessed differential genomic features. c, IGV tracks showing differential ATAC peaks around loci *MSR1*, *RAB1B* and *CNIH2* for diagnosis and relapse. d, Waterfall plot based on HINT TF differential activity. Each dot represents a TF; TFs showing significant differences in activity score (pval < 0.05) are labeled and highlighted in red. e, Detail of ATAC-seq profiles of diagnosis and relapse for motifs RUNX1 and E4F1.

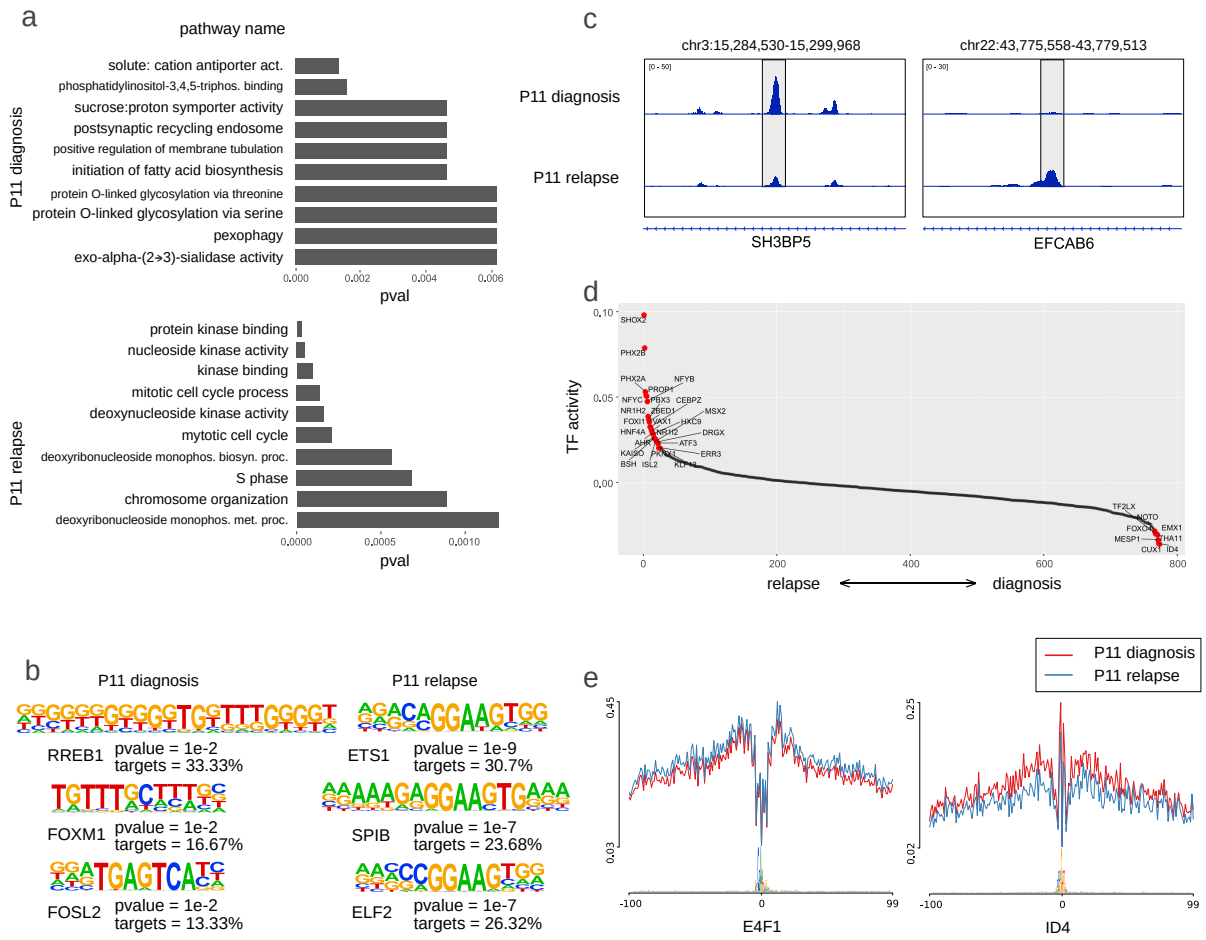

FIGURE A.9: Integrative analysis of ATAC-seq and RNA-seq data for patient 11 (P11). a, *l2p* analysis showing the top 10 pathways associated with differential genomic features in combination with differentially expressed genes for diagnosis (top) and relapse (bottom); abbreviations: *act.*, activity; *triphos.*, triphosphate; *monophos.*, monophosphate; *biosyn.*, biosyntesis; *proc.*, process; *met.*, metylation. b, Top three HOMER *known* predicted transcription factor motifs for assessed differential genomic features. c, IGV tracks showing differential ATAC peaks around loci *SH3BP5* and *EFCAB6* for diagnosis and relapse. d, Waterfall plot based on HINT TF differential activity. Each dot represents a TF; TFs showing significant differences in activity score (pval < 0.05) are labeled and highlighted in red. e, Detail of ATAC-seq profiles of diagnosis and relapse for motifs E4F1 and ID4.

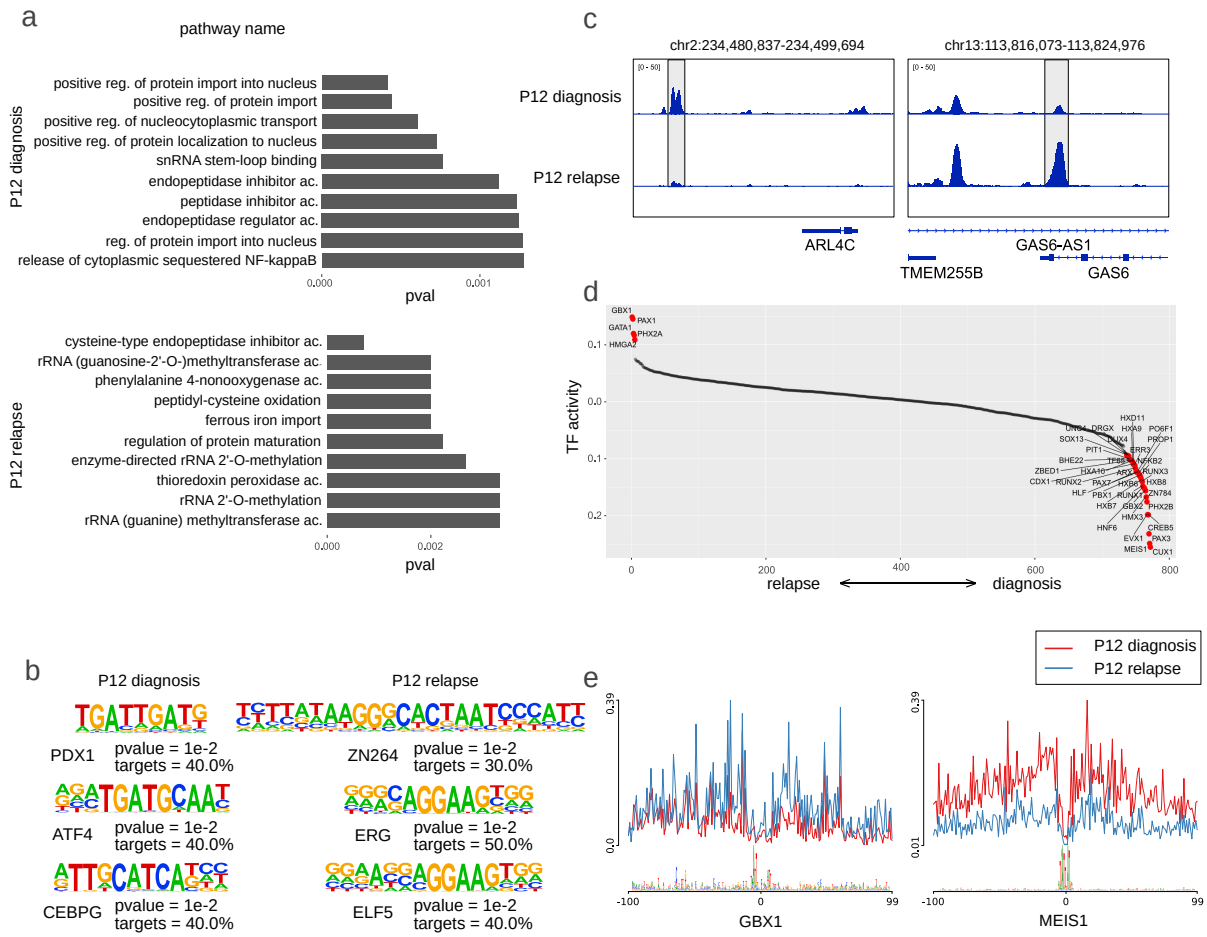

FIGURE A.10: Integrative analysis of ATAC-seq and RNA-seq data for patient 12 (P12). a, *l2p* analysis showing the top 10 pathways associated with differential genomic features in combination with differentially expressed genes for diagnosis (top) and relapse (bottom); abbreviations: *reg.*, regulation; *ac.*, activity. b, Top three HOMER *known* predicted transcription factor motifs for assessed differential genomic features. c, IGV tracks showing differential ATAC peaks around loci *ARL4C* and *GAS6* for diagnosis and relapse. d, Waterfall plot based on HINT TF differential activity. Each dot represents a TF; TFs showing significant differences in activity score (pval < 0.05) are labeled and highlighted in red. e, Detail of ATAC-seq profiles of diagnosis and relapse for motifs GBX1 and MEIS1.

Appendix B

Protocols test

B.1 Testing different parameters to optimize the Omni-ATAC-seq protocol

As described in Section 2.1.3.2, Chapter 2, the results obtained when comparing the Fast-ATAC- and the Omni-ATAC-seq protocols, made us decide to switch protocols and apply from that point on the latest Omni-ATAC version.

Moreover, in order to fine-tune the protocol for optimally performing in our concrete experimental context, we decided to check the impact of some parameters on the data quality, namely:

1. the amount of Tn5 enzyme used for the transposition reaction (as indicated in the protocol, 1 x Tn5, or double as much, 2 x Tn5),
2. the number of cells processed (50 x 10^3 or 10 x 10^3),
3. and the effect of using commercial TD buffer or in-house made TD buffer (as described by Corces et al., 2017)

B.1.1 Tn5 enzyme amount

To analyze the impact that the amount of transposition enzyme has in the reaction, I performed Omni-ATAC-seq on 50 x 10^3 MM-1 cells using the standard amount of Tn5 enzyme, 1 x Tn5, and 2-fold this amount, 2 x Tn5. In both cases, the number of peaks identified was nearby the same: 93.5% of all peaks detected were common (Figure B.1, left), being the small differences associated to peaks with low coverage. Checking the tracks from both experiments (Figure B.1, right), it can be seen how similar they are, pointing out that the amount of transposition enzyme used (in the range we tested) does not have any relevant impact on the results.

B.1.2 Number of cells

The next point we wanted to test was the effect of scaling down the initial number of cells processed. For it, I performed Omni-ATAC-seq on MV4-11 cells, using the standard cell number, 50 x 10^3, and 5-fold less, i.e., 10 x 10^3 cells. In this case we observed a difference of around 7% of more called peaks when processing 50 x 10^3 cells compared to scaling down the amount of starting material. Nevertheless, 90% of all peaks detected were called for both

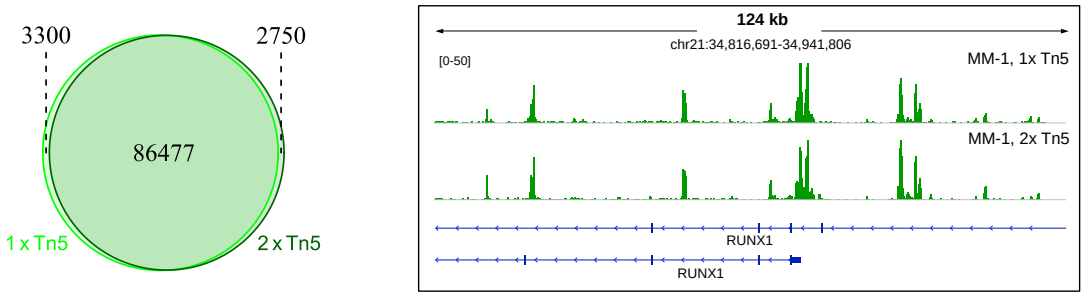

FIGURE B.1: left, Venn diagram showing the number of peaks called by HOMER for Omni-ATAC-seq on MM-1 cells using standard amount of transposition enzyme or 2-fold this amount; right, IGV region around the *RUNX1* locus showing the Omni-ATAC signal for MM-1 cells using the standard amount of transposition enzyme (upper track) or 2-fold this amount (bottom track).

conditions (Figure B.2, left). The biggest impact of scaling down the number of processed cells was found to be the number of PCR cycles needed to appropriately amplify the library: for a standard Omni-ATAC-seq reaction with 50 x 10^3 cells processed, usually 8-9 cycles are needed (four initial cycles prior to the side qPCR reaction + 4 or 5 more cycles -see Section 4.3.4 for a detailed description-), whereas 14 cycles in total were needed when processing 10 x 10^3 cells (Figure B.2, right).

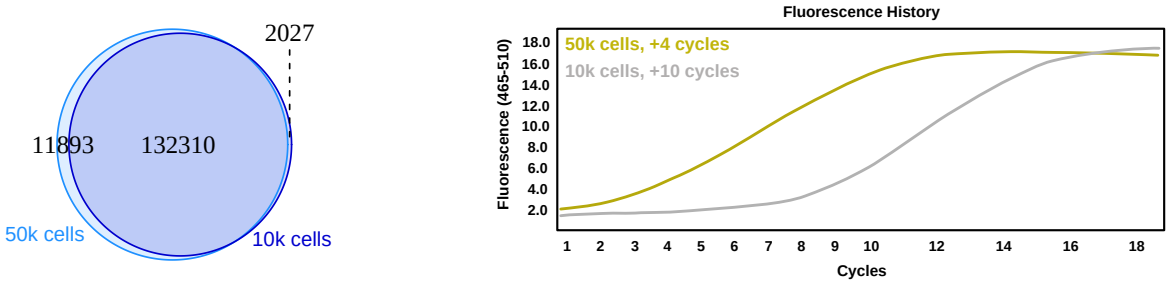

FIGURE B.2: left, Venn diagram showing the number of peaks called by HOMER for Omni-ATAC-seq on MV4-11cells using standard number of cells, 50 x 10^3 and 5-fold less, i.e., 10 x 10^3 cells.; right, plot showing the fluorescence curve from the side-qPCR reaction to estimate the needed number of cycles the libraries should be further amplified after four initial amplification cycles.

B.1.3 TD buffer

The last test I carried out was motivated by the convenience of using in-house made TD buffer, if performing the same, instead of using the commercial one. In this case, I did not sequence the generated libraries, since the pre-sequencing controls, especially the side-qPCR reaction plot, were enough for us to judge the performance of the buffer. The qPCR reaction showed that the same number of cycles were needed for amplification, no matter which buffer was used (Figure B.3).

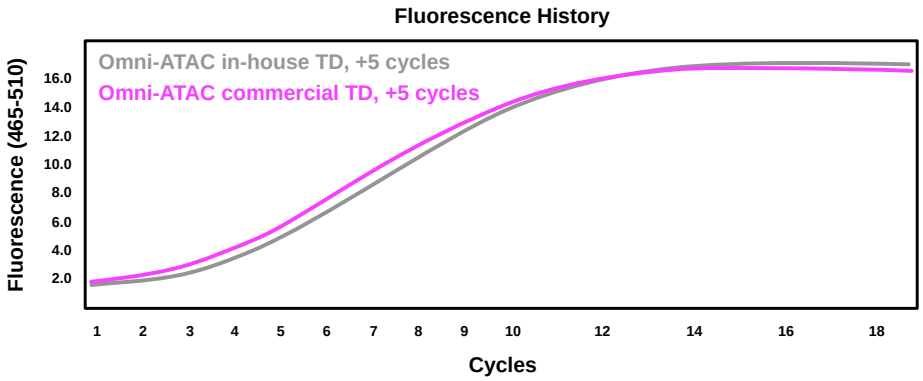

FIGURE B.3: Plot showing the fluorescence curve from the side-qPCR reaction to estimate the needed number of cycles the libraries should be further amplified after four initial amplification cycles. Conditions tested: Omni-ATAC with commercial TD buffer or in-house made buffer.

B.1.4 Conclusions

After all tests and comparisons carried out, the main conclusions drawn and their implications were:

- The Omni-ATAC protocol performs significantly better in terms of number of peaks detected and data quality

  $\Longrightarrow$ use preferentially this version of the protocol for ATAC-seq experiments

- Almost no impact on the results was observed by using the recommended amount of transposition enzyme or double as much of it

  $\Longrightarrow$ use always the standard amount of enzyme, scaling it down when less cells are processed to maintain the ratio Tn5 enzyme-DNA constant

- No differences were observed when using commercial TD buffer or in-house made buffer

  $\Longrightarrow$ perform transposition reaction using self-made TD buffer

- Minimal reduction of found peaks were observed when reducing the number of cells used as starting material from 50 x 10^3 to 10 x 10^3 cells

  $\Longrightarrow$ process, when possible, always 50 x 10^3 cells, being acceptable to use fewer in case of limited cells availability

Appendix C

Scripts

During the course of this work, I contributed to the improvement and fine tuning of some scripts. Among them, the most used scripts were the *bash* scripts for primary analysis, which includes alignment and cleaning-up, as well as peak calling, used for both ATAC- and ChIP-seq data. As the amount of sequencing data increased, we had the need of processing batches of data simultaneously. With this purpose, a loop for parallel-processing of data from an external input file was built-in, increasing the efficiency of the original script and speeding up the running time. Both mentioned scripts are included below.

```
#0_main.sh

#Main script to call the secondary script called 0_alignment.sh, which is the
    actual alignment and cleaning-up script. The external file 1_samples.txt is
    a tab separated two columns file containing in the first column the samples
    ID and in the second one, the path to the correspondent fastq file.

while read ID fastq
do sbatch 0_alignment.sh $ID $fastq
done < 1_samples.txt
```

```
#1_alignment.sh

#load needed modules
module load ngs/FastQC/0.11.7
module load ngs/bowtie1/1.1.2
module load ngs/Homer/4.9
module load ngs/samtools/1.9
module load ngs/UCSCutils/3.4.1

ID=$1
fastq=$2
genome=    #hg38, mm10

#creating a directory for each sample
mkdir $ID
```

```
#generating fastqc report for all fastq files; if the input is double (because
    the sample was re-sequenced and there are two fastq files), only the first
    one will be considered

if [[ $fastq == *","* ]]; then
var=$(echo $fastq | cut -d',' -f 1)
else
var=$fastq
fi
fastqc $var

#moving all fastqc derived files to the correspondent folder
mv *_fastqc.html $ID/
mv *_fastqc.zip $ID/

#alignment without considering multimapping reads
bowtie -q -n 2 --best --chunkmbs 2000 -p 32 -m 1 -S $genome $fastq
    $ID\/$ID\_m1.sam

#checking unwanted reads
awk ' $3 == "chrM" {chrM++};
END {
print "reads mapping to chrM: " chrM;
}' $ID\/$ID\_m1.sam

awk ' $3 == "chrY" {chrY++};
END {
print "reads mapping to chrY: " chrY;
}' $ID\/$ID\_m1.sam

#filtering out those unwanted reads
sed '/chrM/d;/random/d;/chrUn/d;/chrY/d;/chrEBV/d' $ID\/$ID\_m1.sam >
    $ID\/$ID\_m1_filtered.sam

#generating BAM files
samtools view -bS $ID\/$ID\_m1_filtered.sam | samtools sort -o
    $ID\/$ID\_m1_filtered.bam
samtools index $ID\/$ID\_m1_filtered.bam

#removing duplicates
java -jar picard.jar MarkDuplicates I=$ID\/$ID\_m1_filtered.bam
    O=$ID\/$ID\_m1_filtered_nodups.bam M=$ID\/dups.txt REMOVE_DUPLICATES=true
samtools index $ID\/$ID\_m1_filtered_nodups.bam

#generating HOMER TagDirectory
samtools view $ID\/$ID\_m1_filtered_nodups.bam >
    $ID\/$ID\_m1_filtered_nodups.sam
makeTagDirectory $ID\/$ID\_m1_filtered_nodups/ $ID\/$ID\_m1_filtered_nodups.sam
    -format sam -genome $genome
```

```
#generating bigWig files
makeBigWig.pl $ID\/$ID\_m1_filtered_nodups $genome -webdir . -url . -norm 1e7
    -normLength 100 -fragLength 150 -update
mv $ID\_m1_filtered_nodups.ucsc.bigWig $ID/

#removing sam files
rm -f $ID\/*.sam

#compressing all fastq files
gzip $fastq
```

```
#2_peakCalling.sh

#Script to find ATAC peaks (-style factor) or ChIP-seq peaks (-style histone)
    using HOMER; if several samples are analyzed together, the ouputs will be
    merged and finally the peaks annotation will be generated

#load needed modules
module load ngs/Homer/4.9
module load ngs/bedtools2/2.28.0

#1_samples.txt file must include two columns, tab separated, with sample ID and
    path to the Tag Directory

#parameters
genome=  #hg38, mm10
TDI= #path to input tag directory

TDlist=$(awk '{print $2}' < 1_samples.txt | paste -s -d \ )

#Finding peaks (substitute -style factor by histone, when analyzing ChIP-seq
    peaks)
while read ID TD
do
findPeaks $TD -style factor -o $ID\_peaks.txt -i $TDI
sed 1,40d $ID\_peaks.txt > $ID\_peaks_2.txt
pos2bed.pl $ID\_peaks_2.txt > $ID\_peaks.bed
done < 1_samples.txt

#remove *_2 temporary files
rm *_2.txt

#putting files together
cat *_peaks.bed > multi_ATAC.bed #change name accordingly when ChIP-seq; this
    is valid for all code lines below

#sorting
sortBed -i multi_ATAC.bed > multi_ATAC_sorted.bed
```

```
#merging files
bedtools merge -i multi_ATAC_sorted.bed > multi_ATAC_merged.bed

#annotatePeaks from HOMER
annotatePeaks.pl multi_ATAC_merged.bed $genome -d $TDI $TDlist >
    annotatePeaks_ATAC.txt

#annotatePeaks raw
annotatePeaks.pl multi_ATAC_merged.bed $genome -raw -d $TDI $TDlist >
    annotatePeaks_ATAC_raw.txt
```

All bash scripts were run in the HPC cluster (BMC, LMU).

Appendix D

Collaboration research

Therapy resistance in Acute Myeloid Leukemia linked to deficient *KDM6A* expression

This appendix briefly describes our contribution to a research project performed by AG Spiekermann[1], whose results were finally published (Stief et al., 2019). The purpose of the project was to evaluate the impact of the reduction or loss of *KDM6A* expression in the context of therapy resistance in AML.

Analyzing paired diagnosis and relapse samples derived from AML patients, they previously observed acquisition of *KDM6A* mutations associated with the relapse stage and identified *KDM6A* as a novel relapse-related gene relevant in the progression of AML (Greif et al., 2018). Histone lysine demethylase 6a (*KDM6A*) encodes a histone demethylase, specifically mediating the removal of repressive trimethylation from lysine 27 of histone H3 (H3K27me3) and activating, thus, target gene expression. They observed that reduced or absent *KDM6A* expression was correlated with increased resistance to the standard chemotherapy, consisting in a combination of AraC (cytarabine) and daunorubicin, and could probe that AraC resistance was caused by a reduction of the membrane transporter ENT1 (equilibrative nucleoside transporter 1), encoded by the gene *SLC29A1*, which mediates both influx and efflux of nucleosides across the cytoplasmic membrane.

The experiments we performed to complement the mentioned research project included ATAC-seq and ChIP-seq of following histone markers, H3K27ac (characteristic of promoters and active enhancers), as well as H3K27me3 (distinctive of repressed or silenced chromatin), using several AML cell lines.

On the one hand, I processed the AML sister cell lines MM-1, which expresses normal levels of *KDM6A*, i.e., it is *KDM6A* WT, and MM-6, both derived from the same patient at relapse but presenting the last one, as it is also the case for cell line THP-1, *KDM6A* exon deletion mutations and displaying, thus, reduced *KDM6A* expression; on the other hand, I processed the above mentioned THP-1 cells and the myeloid cell line K-562. K-562 cells were genetically engineered using the CRISPR/Cas9 genome editing strategy for knocking-out the expression of *KDM6A* and, together with the THP-1 cell line, it was also manipulated to re-express *KDM6A* through a doxycycline inducible PiggyBac transposon (PB) KDM6A expression system.

[1] Department of Medicine III, University Hospital, LMU Munich, Munich, Germany

They evaluated the proliferation capacity of both AML sister cell lines MM-1 and MM-6 when subjected to AraC treatment and observed a clear evidence that the *KDM6A* mutant MM-6 cell line was less sensitive and, thus, showed a significant growth advantage compared to *KDM6A* WT MM-1 cells under those circumstances. Moreover, they were able to identify the membrane transporter *ENT1* as key candidate gene involved in the AraC metabolism, since the cytotoxic AraC effect is conditioned to the drug to be transported inside the cell, and could demonstrate that decreased *ENT1* expression was linked to *KDM6A* reduction or loss. In line with these results, both ATAC-seq and H3K27ac ChIP-seq signals for AML cell lines MM-1 and MM-6 at the promoter region of the *SLC29A1* locus, encoding ENT1, show a decrease in the *KDM6A* mutant MM-6 compared with the *KDM6A* WT MM-1, whereas no differences can be detected between both cell lines regarding the repressive marker H3K27me3 (Figure D.1).

Moreover, the H3K27ac ChIP signal in the same genomic region studied previously, i.e., at the locus *SLC29A1*, for cell lines K-562 and THP-1 (genetically engineered as previously described) shows a significantly decrease upon *KDM6A* deletion for the K-562 cell line compared to K-562 WT but it is restored again when *KDM6A* is re-expressed upon doxycycline treatment. The same pattern was confirmed for the cell line THP-1, for which the H3K27ac signal also increased upon *KDM6A* re-expression when doxycycline treatment was applied (Figure D.2).

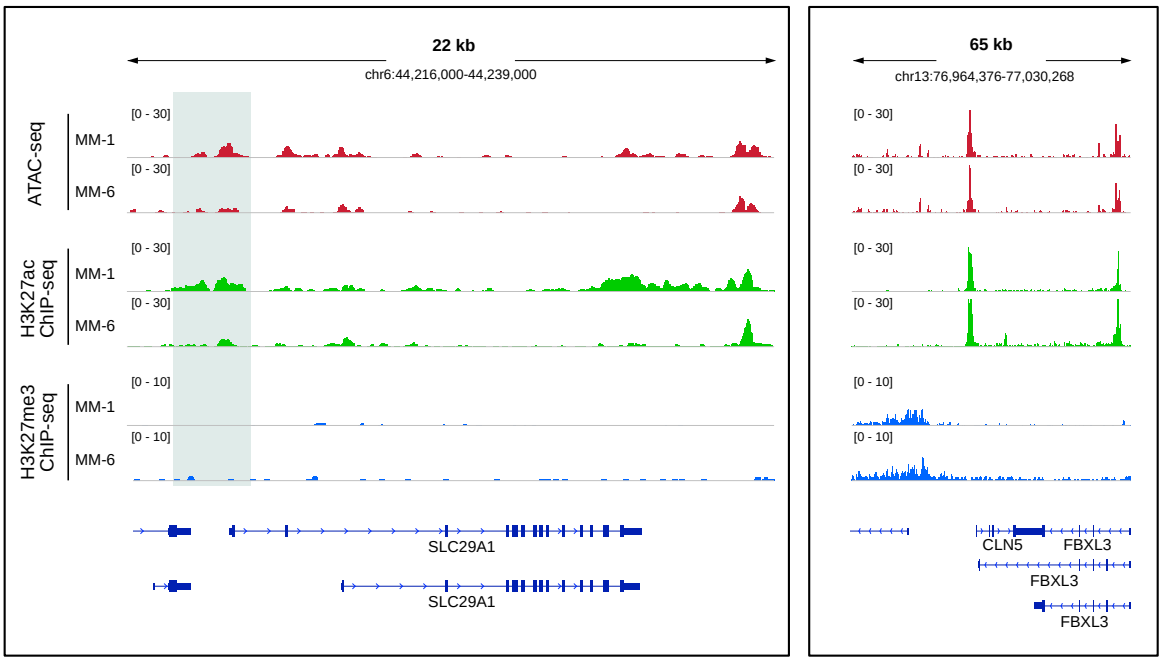

FIGURE D.1: IGV region showing from top to bottom ATAC-, H3K27ac ChIP- and H3K27me3-ChIP-seq signals (in red, green and blue, respectively) for the AML cell lines MM-1 and MM-6 in the region of interest at the *SLC29A1* locus (left panel). As control regions with no significant differences in the signals displayed, regions around the loci *CLN5* and *FBXL3* are shown (right panel). Differential peaks of interest are highlighted at the promoter region of the *ENT1* locus.

Our experiments contributed to confirm their results and supported the reached conclusions with complementary evidences. Taken all together, the research project concluded that *KDM6A* loss contributes to therapy resistance in AML, being the membrane transporter ENT1 involved in this process.

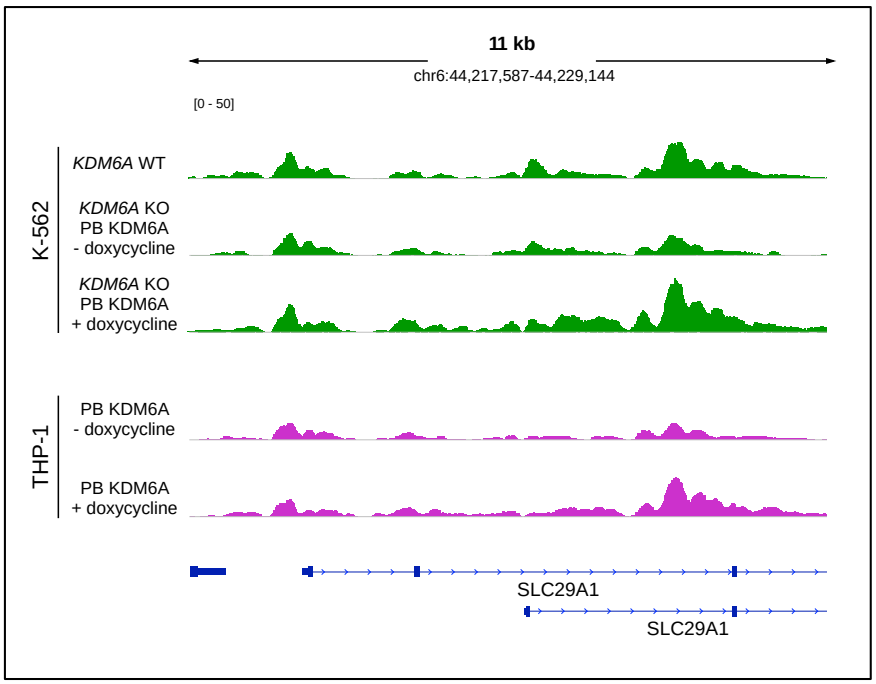

FIGURE D.2: IGV region at *SLC29A1* locus showing H3K27ac ChIP-seq signal corresponding to cell lines K-562 (green) and THP-1 (magenta). For K-562 cells and from top to bottom, displayed tracks correspond to *KDM6A* WT, *KDM6A* KO PB KDM6A before doxycycline treatment and *KDM6A* KO PB KDM6A after doxycycline treatment. For THP-1 cells the first track coincides with PB KDM6A before doxycycline treatment and the second one with cells after doxycycline treatment. Cells were treated with media enriched with 0.5 µg/mL of doxycycline every 24 h for a period of 72 h in total (for more details, see Stief et al., 2019.

Bibliography

Abaza, Y. et al. (Mar. 2017). "Long-term outcome of acute promyelocytic leukemia treated with all-trans-retinoic acid, arsenic trioxide, and gemtuzumab". In: *Blood* 129, pp. 1276–1283. DOI: 10.1182/blood-2016-09-736686.

Adler, Daniel, Duncan Murdoch, et al. (2021). *rgl: 3D Visualization Using OpenGL*. R package version 0.105.22. URL: https://CRAN.R-project.org/package=rgl.

Afgan, E. et al. (July 2018). "The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update". In: *Nucleic Acids Research* 46, W537–W544. DOI: 10.1093/nar/gky379.

Ahn, J.-S. et al. (Dec. 2017). "Assessment of a new genomic classification system in acute myeloid leukemia with a normal karyotype". In: *Oncotarget* 9, pp. 4961–4968. DOI: 10.18632/oncotarget.23575.

Allfrey, V. G., R: Faulkner, and A. E. Mirsky (May 1964). "Acetylation and methylation of histone and their possible role in the regulation of RNA synthesis". In: *Proc Natl Acad Sci USA* 51, pp. 786–794. DOI: 10.1073/pnas.51.5.786.

Allis, C. D. and T. Jenuwein (June 2016). "The molecular hallmarks of epigenetic control". In: *Nature Reviews Genetics* 17, pp. 487–500. DOI: 10.1038/nrg.2016.59.

Altaf, M., N. Saksouk, and J. Coté (May 2007). "Histone modifications in response to DNA damage". In: *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis* 618, pp. 81–90. DOI: 10.1016/j.mrfmmm.2006.09.009.

Alves, C. C. et al. (May 2012). "Leukemia-initiating cells of patient-derived acute lymphoblastic leukemia xenografts are sensitive toward TRAIL". In: *Blood* 119, pp. 4224–4227. DOI: 10.1182/blood-2011-08-370114.

Amin, H. M. et al. (Mar. 2005). "Having a higher blast percentage in circulation than bone marrow: clinical implications in myelodysplastic syndrome and acute lymphoid and myeloid leukemias". In: *Leukemia* 19, pp. 1567–1572. DOI: 10.1038/sj.leu.2403876.

Aran, D. and A. Hellman (July 2013). "DNA methylatin of transcriptional enhancers and cancer predisposition". In: *Cell* 154, pp. 11–13. DOI: 10.1016/j.cell.2013.06.018.

Arrowsmith, C. H. et al. (Apr. 2012). "Epigenetic protein families: a new frontier for drug discovery". In: *Nature Reviews Drug Discovery* 11, pp. 384–400. DOI: 10.1038/nrd3674.

Baccelli, I. et al. (June 2013). "Identification of a population of blood circulating tumor cells from breast cancer patients that initiates metastasis in a xenograft assay". In: *Nature Biotechnology* 6, pp. 539–544. DOI: 10.1038/nbt.2576.

Bagnoli, J. W. et al. (July 2018). "Sensitive and powerful single-cell RNA sequencing using mcSCRB-seq". In: *Nature Communications*. DOI: 10.1038/s41467-018-05347-6.

Bannister, A. J. and T. Kouzarides (Feb. 2011). "Regulation of chromatin by histone modifications". In: *Cell Research* 21, pp. 381–395. DOI: 10.1038/cr.2011.22.

Bao, Y. and X. Shen (Apr. 2007). "Chromatin remodeling in DNA double-strand break repair". In: *Curr Opin Genet Dev* 17, pp. 126–131. DOI: 10.1016/j.gde.2007.02.010.

Becker, P. B. and J. L. Workman (Sept. 2013). "Nucleosome remodeling and epigenetics". In: *Cold Spring Harb Perpect Biol* 5. DOI: 10.1101/cshperspect.a017905.

Belton, J.-M. et al. (May 2012). "Hi-C: a comprehensive technique to capture the conformation of genomes". In: *Methods* 58, pp. 268–276. DOI: 10.1016/j.ymeth.2012.05.001.

Ben-David, U. et al. (Oct. 2017). "Patient-derived xenografts undergo mouse-specific tumor evolution". In: *Nature Genetics* 49, pp. 1567–1576. DOI: 10.1038/ng.3967.

Bennet, J. et al. (Aug. 1976). "Proposals for the classification of acute leukaemias. French-American-British (FAB) cooperative group". In: *British Journal of Haematology* 33, pp. 451–458. DOI: 10.1111/j.1365-2141.1976.tb03563.x.

Bennet, J. M. et al. (Oct. 1985). "Proposed revised criteria for the classification of acute myeloid leukemia. A report of the French-American-British cooperative group". In: *Annals of Internal Medicine* 103, pp. 620–625. DOI: 10.7326/0003-4819-103-4-620.

Berger, S. L. et al. (Apr. 2009). "An operational definition of epigenetics". In: *Genes and Development* 23, pp. 781–783. DOI: 10.1101/gad.1787609.

Bernstein, I. D. (Mar. 2000). "Monoclonal antibodies to the myeloid stem cells: therapeutic implications of CMA-676, a humanized anti-CD33 antibody calicheamicin conjugate". In: *Leukemia* 14, pp. 474–475. DOI: 10.1038/sj.leu.2401663.

Bester, A. C. et al. (Apr. 2018). "An integrated genome-wide CRISPRa approach to functionalize lncRNAs in drug resistance". In: *Cell* 173, pp. 649–664. DOI: 10.1016/j.cell.2018.03.052.

Bettelli, E. et al. (Oct. 1998). "IL-10 is critical in the regulation of autoimmune encephalomyelitis as demonstrated by studies of IL-10- and IL-4-deficient and transgenic mice". In: *The Journal of Immunology* 161, pp. 3299–3306. URL: https://www.jimmunol.org/content/161/7/3299.

Blighe, Kevin (2019). *PCAtools: PCAtools: Everything Principal Components Analysis*. R package version 1.2.0. URL: https://github.com/kevinblighe/PCAtools.

BMC. *Core Facility Flow Cytometry*. https://www.flowcytometry.bmc.med.uni-muenchen.de/our-instruments/cell-sorters/index.html. Last accessed Apr 2021.

Bonnet, D. and J. E. Dick (July 1997). "Human acute myeloid leukemia is organized as a hierarchy that originates from a primitive hematopoietic cell". In: *Nature Medicine* 3, pp. 730–737. DOI: 10.1038/nm0797-730.

Borthakur, G. et al. (Oct. 2014). "Gemtuzumab ozogamicin with fludarabine, cytarabine, and granulocyte colony stimulating factor (FLAG-GO) as fron-line regimen in patients with core binding factor acute myelogenous leukemia". In: *American Journal of Hematology* 89, pp. 964–968. DOI: 10.1002/ajh.23795.

Boultwood, J. and J. S. Wainscoat (July 2007). "Gene silencing by DNA methylation in haematological malignancies". In: *British Journal of Haematology* 138, pp. 3–11. DOI: 10.1111/j.1365-2141.2007.06604.x.

Boyle, A. P. et al. (Nov. 2011). "High-resolution genome-wide in vivo footprinting of diverse transcription factors in human cells". In: *Genome Research* 21, pp. 456–464. DOI: 11.1101/gr.112656.110.

Brown, R. et al. (Nov. 2014). "Poised epigenetic states and acquired drug resistance in cancer". In: *Nature Reviews in Cancer* 14, pp. 747–753. DOI: 10.1038/nrc3819.

Brownell, J. E. et al. (Mar. 1996). "Tetrahymena histone acetyltransferase A: a homolog to yeast Gcn5p linking histone acetylation to gene activation". In: *Cell* 84, pp. 843–851. DOI: 10.1016/s0092-8674(00)81063-6.

Brunning, R. D. et al. (2001). *World Health Organization classification of tumors: pathology and genetics of tumors of haematopoietic and lymphoid tissues*. Lyon, France: IARC Press, pp. 63–67. ISBN: 9283224116.

Bryan, J. C. and E. J. Jabbour (Aug. 2015). "Management of relapsed/refractory acute myeloid leukemia in the elderly: current strategies and developments". In: *Drugs & Aging* 32, pp. 623–637. DOI: 10.1007/s40266-015-0285-6.

Buenrostro, J. D. et al. (Oct. 2013). "Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position". In: *Nature Methods* 10, pp. 1213–1218. DOI: 10.1038/nmeth.2688.

Buenrostro, J. D. et al. (Jan. 2015). "ATAC-seq: a method for assaying chromatin accessibility genome-wide". In: *Current Protocols in Molecular Biology* 109. DOI: 10.1002/0471142727.mb2129s109.

Busch, K. et al. (Feb. 2015). "Fundamental properties of unperturbed haematopoiesis from stem cells *in vivo*". In: *Nature* 518, pp. 542–546. DOI: 10.1038/nature14242.

Byrd, J. C. et al. (Dec. 2002). "Pretreatment cytogenetic abnormalities are predictive of induction success, cumulative incidence of relapse, and overall survival in adult patients with de novo acute myeloid leukemia: results from Cancer and Leukemia Group B". In: *Blood* 100, pp. 4325–4336. DOI: 10.1182/blood-2002-03-0772.

Cai, D. et al. (Jan. 2021). "Identification of six hub genes and analysis of their correlation with drug sensitivity in acute myeloid leukemia through bioinformatics". In: *Translational Cancer Research* 10. ISSN: 2219-6803. URL: https://tcr.amegroups.com/article/view/48013.

Calo, E. and J. Wysocka (Mar. 2013). "Modification of enhancer chromatin: what, how, and why". In: *Molecular Cell*, pp. 825–837. DOI: 10.1016/j.molcel.2013.01.038.

Carlson, Marc (2019). *org.Hs.eg.db: Genome wide annotation for Human*. R package version 3.10.0. DOI: 10.18129/B9.bioc.org.Hs.eg.db.

Casas, E. et al. (Jan. 2011). "Snail2 is an essential mediator of Twist1-induced epithelial-mesen-chymal transition and metastasis". In: *Cancer Research* 71, pp. 245–254. DOI: 10.1158/0008-5472.CAN-10-2330.

Celiá-Terrassa, T. et al. (May 2012). "Epithelial-mesenchymal transition can suppress major attributes of human epithelial tumor-initiating cells". In: *The Journal of Clinical Investigation* 122, pp. 1849–1868. DOI: 10.1172/JCI59218.

Cernilogar, F. et al. (Sept. 2019). "Pre-marked chromatin and transcription factor co-binding shape the pioneering activity of Foxa2". In: *Nucleic Acid Reseach*, pp. 9069–9086. DOI: 10.1093/nar/gkz627.

Chen, A. F. et al. (Aug. 2018). "GRHL2-dependent enhancer switching maintains a pluripotent stem cell transcriptional subnetwork after exit from naive pluripotency". In: *Cell Stem Cell* 23, pp. 226–238. DOI: 10.1016/j.stem.2018.06.005.

Chen, H. (2018). *VennDiagram: Generate High-Resolution Venn and Euler Plots*. R package version 1.6.20. URL: https://CRAN.R-project.org/package=VennDiagram.

Chen, Y. et al. (Sept. 2020). "Core transcriptional regulatory circuitries in cancer". In: *Oncogene*, pp. 6633–6646. DOI: 10.1038/s41388-020-01459-w.

Cheng, Y. et al. (Dec. 2019). "Targeting epigenetic regulators for cancer therapy: mechanisms and advances in clinical trials". In: *Signal Transduction and Targeted Therapy* 62. DOI: 10.1038/s41392-019-0095-0.

Chi, P., C. D. Allis, and G. G. Wang (July 2010). "Covalent histone modifications: miswritten, misinterpreted and mis-erased in human cancers". In: *Nature Reviews Cancer* 10, pp. 457–469. DOI: 10.1038/nrc2876.

Chopra, M. and S. K. Bohlander (Dec. 2019). "The cell of origin and the leukemia stem cell in acute myeloid leukemia". In: *Genes, Chromosomes and Cancer*, pp. 850–858. DOI: 10.1002/gcc.22805.

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

Civin, C. L. et al. (July 1984). "Antigenic analysis of hematopoiesis. III. A hematopoietic progenitor cell surface antigen defined by a monoclonal antibody raised against KG-1a cells". In: *Journal of Immunology*, pp. 157–165.

Collignon, A. et al. (June 2020). "A chemogenomic approach to identify personalized therapy for patients with relapse or refractory acute myeloid leukemia: results of a prospective feasibility study". In: *Blood Cancer Journal* 10.64. DOI: 10.1038/s41408-020-0330-5.

Constantinescu, C. S. et al. (Feb. 2012). "Experimental autoimmune encephalomyelitis (EAE as a model for multiple sclerosis (MS)". In: *British Journal of Pharmacology* 164, pp. 1079–1106. DOI: 10.1111/j.1476-5381.2011.01302.x.

Corces, M. R. et al. (Oct. 2016). "Lineage-specific and single cell chromatin accessibility charts human hematopoiesis and leukemia evolution". In: *Nature Genetics*, pp. 1193–1203. DOI: 10.1038/ng.3646.

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

Creyghton, M. P. et al. (Dec. 2010). "Histone H3K27ac separates active from poised enhancers and predicts developmental state". In: *Proceedings of the National Academy of Sciences of the USA* 107, pp. 21931–21936. DOI: 10.1073/pnas.1016071107.

Cripe, L. D. et al. (May 2002). "Role for c-jun N-terminal kinase in treatment-refractory acute myeloid leukemia (AML): signaling to multidrug-efflux and hyperproliferation". In: *Leukemia* 16, pp. 799–812. DOI: 10.1038/sj.leu.2402457.

Crispino, J. D. and M. S. Horwitz (Feb. 2017). "GATA factor mutations in hematologic disease". In: *Blood* 129, pp. 2103–2110. DOI: 10.1182/blood-2016-09-687889.

Dahl, J. A. and P. Collas (June 2008). "A rapid micro chromatin immunoprecipitation assay (ChIP)". In: *Nature Protocols* 3, pp. 1032–1045. DOI: 10.1038/nprot.2008.68.

Daramola, A. S. et al. (Nov. 2019). "Rara is a druggable super-enhancer regulated dependency in pediatric AML". In: *Blood* 134. DOI: 10.1182/blood-2019-125425.

Deans, C. and K. A. Maggert (Apr. 2015). "What do you mean, "Epigenetic"". In: *Genetics* 199, pp. 887–896. DOI: 10.1534/genetics.114.173492.

Dekker, J. et al. (Feb. 2002). "Capturing chromosome conformation". In: *Science* 295, pp. 1306–1311. DOI: 10.1126/science.1067799.

Derolf, A. R. et al. (Apr. 2009). "Improved patient survival for acute myeloid leukemia: a population-based study of 9729 patients diagnosed in Sweden between 1973 and 2005". In: *Blood*, pp. 3666–3671. DOI: 10.1182/blood-2008-09-179341.

Döhner, H., D. J. Weisdorf, and C. D. Bloomfield (Sept. 2015). "Acute Myeloid Leukemia". In: *The New England Journal of Medicine* 373, pp. 1136–1152. DOI: 10.1056/NEJMra1406184.

Döhner, H. et al. (Jan. 2017). "Diagnosis and management of AML in adults: 2017 ELN recommendations from an international expert panel". In: *Blood* 129, pp. 424–447. DOI: 10.1182/blood-2016-08-733196.

DiNardo, C. D. et al. (Feb. 2016). "Interactions and relevance of blast percentage and treatment strategy among younger and older patients with acute myeloid leukemia (AML) and myelodysplastic syndrome (MDS)". In: *American Journal of Haematology* 91, pp. 227–232. DOI: 10.1002/ajh.24252.

Ding, L., T. J. Ley, and D. Larson et al. (Jan. 2012). "Clonal evolution in relapsed acute myeloid leukaemia revealed by whole-genome sequencing". In: *Nature* 481, pp. 506–510. DOI: 10.1038/nature10738.

Dorritie, K. A., J. A. McCubrey, D. E. Johnson, et al. (June 2013). "STAT transcription factors in hematopoiesis and leukemogenesis: opportunities for therapeutic intervention". In: *Leukemia* 28, pp. 248–257. DOI: 10.1038/leu.2013.192.

Dostie, J. et al. (Oct. 2006). "Chromosome Conformation Capture Carbon Copy (5C): a massively parallel solution for mapping interactions between genomic elements". In: *Genome Research* 16, pp. 1299–1309. DOI: 10.1101/gr.5571506.

Eberharter, A. and P. B. Becker (Mar. 2002). "Histone acetylation: a switch between repressive and permissive chromatin". In: *EMBO Reports* 3, pp. 224–229. DOI: 10.1093/embo-reports/kvf053.

Ehrhardt, H. et al. (Nov. 2012). "Enhanced anti-tumour effects of Vinca alkaloids given separately from cytostatic therapies". In: *British Journal of Pharmacology* 168, pp. 1558–1569. DOI: 10.1111/bph.12068.

Eichelberger, L. et al. (Apr. 2020). "Maintenance of epithelial traits and resistance to mesenchymal reprogramming promote proliferation in metastatic breast cancer". In: *bioRxiv*. DOI: https://doi.org/10.1101/2020.03.19.998823.

Ernst, J. and M. Kellis (July 2010). "Discovery and characterization of chromatin states for systematic annotation of the human genome". In: *Nature Biotechnology* 28, pp. 817–825. DOI: 10.1038/nbt.1662.

Esteller, M. (Apr. 2007). "Cancer epigenomics: DNA methylomes andhistone-modification maps". In: *Nature Reviews Genetics* 8, pp. 286–298. DOI: 10.1038/nrg2005.

Estey, E. H. (Apr. 2013). "Acute myeloid leukemia: 2013 update on risk stratification and management". In: *American Journal of Hematology* 88, pp. 318–327. DOI: 10.1002/ajh.23404.

Federation, A. et al. (June 2018). "Identification of candidate master transcription factors within enhancer-centric transcriptional regulatory networks". In: *BioRxiv*. DOI: 10.1101/345413.

Felsenfeld, G. (Sept. 1996). "Chromatin structure and gene expression". In: *Proc. Natl. Acad. Sci. USA* 93, pp. 9384–9388. DOI: 10.1073/pnas.93.18.9384.

Feng, Y. et al. (Aug. 2014). "Control of the inheritance of regulatory T cell identity by a *cis* element in the *Foxp3* locus". In: *Cell* 158, pp. 749–763. DOI: 10.1016/j.cell.2014.07.031.

Ferrara, F. and C. A. Schiffer (Feb. 2013). "Acute myeloid leukaemia in adults". In: *Lancet*, pp. 484–495. DOI: 10.1016/S0140-6736(12)61727-9.

Finney, R. and G. Nelson (July 2020). "List-to-Pathway: an ultrafast R package for gene set enrichment analysis (version0.0.3". In: *Zenodo*. DOI: 10.5281/zenodo.3942233.

Fischer, K. R. et al. (Nov. 2015). "Epithelial-to-mesenchymal transition is not required for lung metastasis but contributes to chemoresistance". In: *Nature* 527, pp. 472–476. DOI: 10.1038/nature15748.

Fletcher, T. M. and J. C. Hansen (1996). "The nucleosomal array: structure/function relationships". In: *Critical Reviews in Eukaryotic Gene Expression* 6, pp. 149–188. DOI: 10.1615/CritRevEukarGeneExpr.v6.i2-3.40.

Floess, S. et al. (Jan. 2007). "Epigenetic control of the *Foxp3* locus in regulatory T cells". In: *Plos Biology* 5. DOI: 10.1371/journal.pbio.0050038.

Fraga, M. F. et al. (Apr. 2005). "Loss of acetylation at Lys16 and trimethylation at Lys20 of histone H4 is a common hallmark of human cancer". In: *Nature Genetics* 37, pp. 391–400. DOI: 10.1038/ng1531.

Fu, L. et al. (July 2017). "High expression of ETS2 predicts poor prognosis in acute myeloid leukemia and may guide treatment decisions". In: *Journal of Translational Medicine* 15. DOI: 10.1186/s12967-017-1260-2.

Fulco, C. P. et al. (Nov. 2016). "Systematic mapping of functional enhancer-promoter connections with CRISPR interference". In: *Science* 354. DOI: 10.1126/science.aag2445.

Furey, T. S. (Oct. 2012). "ChIP-seq and beyond: new and improved methodologies to detect and characterize protein-DNA interactions". In: *Nature Reviews Genetics* 13, pp. 840–852. DOI: 10.1038/nrg3306F.

Furtado, C. L. M. et al. (Dec. 2019). "Epidrugs: targeting epigenetic marks in cancer treatment". In: *Epigenetics* 14, pp. 1164–1176. DOI: 10.1080/15592294.2019.1640546.

Fyodorov, D. V. et al. (Oct. 2018). "Emerging roles of linker histones in regulating chromatin structure and function". In: *Nature Reviews Molecular Cell Biology* 19, pp. 192–206. DOI: 10.1038/nrm.2017.94.

Galm, O., J. G. Herman, and S. B. Baylin (Jan. 2006). "The fundamental role of epigenetics in hematopoietic malignancies". In: *Blood Reviews* 20, pp. 1–13. DOI: 10.1016/j.blre.2005.01.006.

Gao, J., Y.-H. Chen, and L. C. Peterson (Oct. 2015). "GATA family transcription factors: emerging suspects in hematologic disorders". In: *Experimental Hematology & Oncology*. DOI: 10.1186/s40164-015-0024-z.

Garg, G. et al. (Feb. 2019). "Blimp1 prevents methylation of Foxp3 and loss of regulatory T cell identity at sites of inflammation". In: *Cell Reports* 26, pp. 1854–1868. DOI: 10.1016/.celrep.2019.01.070.

Garson, O. M. et al. (July 1989). "Cytogenetics studies of 103 patients with acute myelogenous leukemia in relapse". In: *Cancer Genetics* 40, pp. 187–202. DOI: 10.1016/0165-4608(89)90024-1.

Gartlgruber, M. et al. (Dec. 2020). "Super enhancers define regulatory subtypes and cell identity in neuroblastoma". In: *Nature Cancer* 2, pp. 114–128. DOI: 10.1038/s43018-020-00145-w.

Gavert, N. and A. Ben-Ze'ev (May 2008). "Epithelial-mesenchymal transition and the invasive potential of tumors". In: *Trends in Molecular Medicine* 14, pp. 199–209. DOI: 10.1016/j.molmed.2008.03.004.

Gawdat, R. M. et al. (Jan. 2017). "Overexpression of the giant FAT1 cadherin gene and its prognostic significance in *de novo* acute leukaemia patients". In: *Comparative Clinical Pathology* 26, pp. 505–512. DOI: 10.1007/s00580-017-2409-3.

Gbadamosi, Bolanle et al. (Sept. 2017). "Treatment outcomes in adults with acute myeloid leukemia (AML) using standard induction "3+7", hypomethylating agents or best supportive care: a single institution experience". In: *Clinical Lymphoma, Myeloma and Leukemia* 17. DOI: 10.1016/j.clml.2017.07.066.

Ge, Y. et al. (Mar. 2007). "The role of the proto-oncogene ETS2 in acute megakaryocytic leukemia biology and therapy". In: *Leukemia* 22, pp. 521–529. DOI: 10.1038/sj.leu.2405066.

Goryshin, I. Y. and W. S. Reznikoff (Mar. 1998). "Tn5 in vitro transposition". In: *Journal of Biological Chemistry* 273, pp. 7367–7374. DOI: 10.1074/jbc.273.13.7367.

Grant, C. E., T. L. Bailey, and W. S. Noble (Feb. 2011). "FIMO: scanning for occurrences of a given motif". In: *Bioinformatics* 27, pp. 1017–1018. DOI: 10.1093/bioinformatics/btr064.

Greif, P. A. et al. (Apr. 2018). "Evolution of cytogenetically normal acute myeloid leukemia during therapy and relapse: an exome sequencing study of 50 patients". In: *Clinical Cancer Research* 24, pp. 1716–1726. DOI: 10.1158/1078-0432.CCR-17-2344.

Grimwade, D. et al. (Oct. 1998). "The importance of diagnostic cytogenetics on outcome in AML: analysis of 1612 patients entered into the MRC AML 10 trial". In: *Blood*. PMID: 9746770.

Grimwade, D. et al. (Sept. 2001). "The predictive value of hierarchical cytogenetic classification in older adults with acute myeloid leukemia (AML): analysis of 1065 patients entered into the United Kingdom Medical Research Council AML11 trial". In: *Blood*. DOI: 10.1182/blood.v98.5.1312.

Gröschel, S. et al. (Jan. 2015). "Mutational spectrum of myeloid malignancies with inv(2)/t(3;3) reveals a predominant involvement of RAS/RTK signaling pathways". In: *Blood*. DOI: 10.1182/blood-2014-07-591461.

Grunstein, M. (Sept. 1997). "Histone acetylation in chromatin structure and transcription". In: *Nature* 389, pp. 349–352. DOI: 10.1038/38664.

Gu, Zuguang, Roland Eils, and Matthias Schlesner (2016). "Complex heatmaps reveal patterns and correlations in multidimensional genomic data". In: *Bioinformatics*.

Guarino, M., B. Rubino, and G. Ballabio (June 2007). "The role of epithelial-mesenchymal transition in cancer pathology". In: *Pathology* 39, pp. 305–318. DOI: 10.1080/00313020701329914.

Guo, D. et al. (May 2021). "GADD45g acts as a novel tumor suppressor and its activation confers new combination regimens for the treatment of AML". In: *Blood*. DOI: 10.1182/blood.2020008229.

Hackl, H., K. Astanina, and R. Wieser (Feb. 2017). "Molecular and genetic alterations associated with therapy resistance and relapse of acute myeloid leukemia". In: *Journal of Hematology & Oncology*. DOI: 10.1186/s13045-017-0416-0.

Hagberg, A., D. Schult, and P. Swart (2020). *NetworkX reference*. Release 2.5. URL: https://networkx.org/.

Hariprakash, J. M. and F. Ferrari (June 2019). "Computational biology solutions to identify enhancers-target gene pairs". In: *Computational and Structural Biotechnology Journal* 17, pp. 821–831. DOI: 10.1016/j.csbj.2019.06.012.

He, H. H. et al. (Mar. 2010). "Nucleosome dynamics define transcriptional enhancers". In: *Nature Genetics* 42, pp. 343–347. DOI: 10.1038/ng.545.

Heintzman, N. and B. Ren (Dec. 2009). "Finding distal regulatory elements in the human genome". In: *Current Opinion in Genetic Development* 19, pp. 541–549. DOI: 10.1016/j.gde.2009.09.006.

Heintzman, N. et al. (Feb. 2007). "Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome". In: *Nature Genetics* 39, pp. 311–318. DOI: 10.1038/ng1966.

Heinz, S. et al. (May 2010a). "Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities". In: *Molecular Cell*. DOI: 10.1016/j.molcel.2010.05.004.

Heinz, S. et al. (Mar. 2010b). "The selection and function of cell type-specific enhancers". In: *Nature Reviews Molecular Cell Biology* 16, pp. 144–154. DOI: 10.1038/nrm3949.

Henikoff, S. (Jan. 2008). "Cucleosome destablilization in the epigenetic regulation of gene expression". In: *Nature Reviews Genetics* 9, pp. 15–26. DOI: 10.1038/nrg2206.

Herz, H.-M., D. Hu, and A. Shilatifard (Mar. 2014). "Enhancer malfunction in cancer". In: *Molecular cell*, pp. 859–866. DOI: 10.1016/j.molcel.2014.02.033.

Hirahara, K. et al. (May 2015). "Asymmetric action of STAT transcription factors drives transcriptional outputs and cytokine specificity". In: *Immunity* 42, pp. 877–889. DOI: 10.1016/j.immuni.2015.04.014.

Hnisz, D. et al. (Oct. 2013). "Super-Enhancers in the control of cell identity and disease". In: *Cell* 155.4, pp. 934–947. DOI: 10.1016/j.cell.2013.09.053.

Hong, L. et al. (Jan. 1993). "Studies of the DNA binding properties of histone H4 amino terminus. Thermal denaturation studies reveal that acetylation markedly reduces the binding constant of the H4 "tail" to DNA". In: *Journal of Biological Chemistry* 268, pp. 305–314.

Hong, T. et al. (Nov. 2015). "An Ovol2-Zeb1 mutual inhibitory circuit governs bidirectional and multi-step transition between epithelial and mesenchymal states". In: *PLoS Computational Biology* 11. DOI: 10.1371/journal.pcbi.1004569.

Hori, S., T. Nomura, and S. Sakaguchi (Feb. 2003). "Control of regulatory T cell development by the transcription factor Foxp3". In: *Science* 299, pp. 1057–1061. DOI: 10.1126/science.1079490.

Horibata, S. et al. (May 2019). "Heterogeneity in refractory acute myeloid leukemia". In: *PNAS* 116, pp. 10494–10503. DOI: 10.1073/pnas.1902375116.

Huehn, J., J. K. Polansky, and A. Hamann (Feb. 2009). "Epigenetic control of FOXP3 expression: the key to a stable regulatory T-cell lineage?" In: *Nature Reviews Immunology* 9, pp. 83–89. DOI: 10.1038/nri2474.

Humphries, B. and C. Yang (Mar. 2015). "The microRNA-200 family: small molecules with novel roles in cancer development, progression and therapy". In: *Oncotarget* 6, pp. 6472–6498. DOI: 10.18632/oncotarget.3052.

Hyndman, K. A. and M. A. Knepper (Oct. 2017). "Dynamic regulation of lysine acetylation: the balance between acetyltransferase and deacetylase activities". In: *American Journal of Physiology; Renal physiology* 313, pp. 842–846. DOI: 10.1152/ajprenal.00313.2017.

Jacobs, J. et al. (July 2018). "The transcription factor grainy head primes epithelial enhancers for spatiotemporal activation by displacing nucleosomes". In: *Nature Genetics* 50, pp. 1011–1020. DOI: 10.1038/s41588-018-0140-x.

Jan, M. et al. (Aug. 2012). "Clonal evolution of preleukemic hematopoietic stem cells precedes human acute myeloid leukemia". In: *Science Translational Medicine* 4. DOI: 10.1126/scitranslmed.3004315.

Josefowicz, S. Z., C. B. Wilson, and A. Y. Rudensky (June 2009). "Cutting edge: TCR stimulation is sufficient for induction of Foxp3 expression in the absence of DNA methyltransferase 1". In: *The Journal of Immunology* 182, pp. 6648–6652. DOI: 10.4049/jimmunol.0803320.

Kallies, A. et al. (Aug. 2009). "Blimp-1 transcription factor is required for the differentiation of effector $CD8^+$ T cells and memory responses". In: *Immunity* 31, pp. 283–295. DOI: 10.1016/j.immuni.2009.06.021.

Kandoth, C. et al. (Oct. 2013). "Mutational landscape and significance across 12 major cancer types". In: *Nature* 502, pp. 333–339. DOI: 10.1038/nature12634.

Kansal, R. (Mar. 2016). "Acute myeloid leukemia in the era of precision medicine: recent advances in diagnostic classification and risk stratification". In: *Cancer Biology & Medicine* 13, pp. 41–54. DOI: 10.28092/j.issn.2095-3941.2016.0001.

Kantarjian, H. M. et al. (Feb. 2021). "Acute myeloid leukemia: current progress and future directions". In: *Blood Cancer Journal* 11.41. DOI: 10.1038/s41408-021-00425-3.

Karantanos, T. and R. J. Jones (2019). *Acute myeloid leukemia stem cell heterogeneity and its clinical relevance*. Vol. 1139, pp. 153–169. DOI: 10.1007/978-3-030-14366-4_9.

Kelly, T. K., D. D. De Carvalho, and P. A. Jones (Oct. 2010). "Epigenetic modifications as therapeutic targets". In: *Nature Biotechnology* 28, pp. 1069–1078. DOI: 10.1038/nbt.1678.

Kim, D. et al. (July 2019). "A principled strategy for mapping enhancers to genes". In: *Scientific Reports* 9.11043. DOI: 10.1038/s41598-019-47521-w.

Kimby, E., P. Nygren, and B. Glimelius (2001). "A systematic overview of chemotherapy effects in acute myeloid leukaemia". In: *Acta Oncology* 40, pp. 231–252. DOI: 10.1080/02841860151116321.

Klein, D. C. and S. J. Hainer (Mar. 2020). "Genomic methods in profiling DNA accessibility and factor localization". In: *Chromosome Research* 28, pp. 69–85. DOI: 10.1007/s10577-019-09619-9.

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

Koch, M. A. et al. (Sept. 2012). "T-bet(+) Treg cells undergo abortive Th1 cell differentiation due to impaired expression of IL-12 receptor β2". In: *Immunity* 37, pp. 501–510. DOI: 10.1016/j.immuni.2012.05.031.

Kogan, A. A. (Jan. 2019). "Exploiting epigenetically mediated changes: acute myeloid leukemia, leukemia stem cells and the bone marrow microenvironment". In: *Advanced Cancer Research* 141, pp. 213–253. DOI: 10.1016/bs.acr.2018.12.005.

Kolitz, Jonathan et al. (Sept. 2018). "Outcomes in Older Patients with Newly Diagnosed, High-risk/Secondary Acute Myeloid Leukemia (sAML) Who Received Consolidation in a Phase 3 Study of CPX-351 versus Conventional 7+3/5+2 Cytarabine and Daunorubicin". In: *Clinical Lymphoma Myeloma and Leukemia* 18, S208 –S209. DOI: 10.1016/j.clml.2018.07.064.

Kornberg, R. D. (May 1974). "Chromatin structure: a repeating unit of histones and DNA". In: *Science* 184, pp. 868–871. DOI: 10.1126/science.184.4139.868.

Kornberg, R. D. and Y. Lorch (Aug. 1999). "Twenty-five years of the nucleosome, fundamental particle of the eukaryotic chromosome". In: *Cell* 98, pp. 285–294. DOI: 10.1016/S0092-8674(00)81958-3.

Kouchkovsky, I. De and M. Abdul-Hay (July 2016). "Acute myeloid leukemia: a comprehensive review and 2016 update". In: *Blood Cancer Journal*. DOI: 10.1038/bcj.2016.50.

Kouzarides, T. (Feb. 2007). "Chromatin modifications and their function". In: *Cell* 128, pp. 693–705. DOI: 10.1016/j.cell.2007.02.005.

Kowalski, P. J., M. A. Rubin, and C. G. Kleer (Sept. 2003). "E-cadherin expression in primary carcinomas of the breast and its distant metastases". In: *Breast Cancer Research* 5, pp. 217–222. DOI: 10.1186/bcr651.

Krebs, A. M. et al. (May 2017). "The EMT-activator Zeb1 is a key factor for cell plasticity and promotes metastasis in pancreatic cancer". In: *Nature Cell Biology* 19, pp. 518–529. DOI: 10.1038/ncb3513.

Kröger, C. et al. (Apr. 2019). "Acquisition of a hybrid E/M state is essential for tumorigenicity of basal breast cancer cells". In: *Proc. Natl. Acad. Sci. U.S.A.* 116, pp. 7353–7362. DOI: 10.1073/pnas.1812876116.

Krijthe, Jesse H. (2015). *Rtsne: T-Distributed Stochastic Neighbor Embedding using Barnes-Hut Implementation*. R package version 0.15. URL: https://github.com/jkrijthe/Rtsne.

Kulakovskiy, I. V. et al. (2018). "HOCOMOCO: towards a complete collection of transcription factor binding models for human and mouse via large-scale ChIP-seq analysis". In: *Nucleic Acid Research* 46, pp. D252–D259. DOI: 10.1093/nar/gkx1106.

Kuo, M. H. and C. D. Allis (Aug. 1998). "Roles of histone acetyltransferases and deacetylases in gene regulation". In: *BioEssays* 20, pp. 615–626. DOI: 10.1002/(SICI)1521-1878(199808)20:8<615::AID-BIES4>3.0.CO;2-H.

Kurdistani, S. K. and M. Grunstein (Apr. 2003). "Histone acetylation and deacetylation in yeast". In: *Nature Reviews Molecular Cell Biology* 4, pp. 276–284. DOI: 10.1038/nrm1075.

Kuznetsova, T. and H. G. Stunnenberg (May 2016). "Cynamic chromatin organization: role in development and disease". In: *International Journal of Biochemistry and Cell Biology* 76, pp. 119–122. DOI: 10.1016/j.biocel.2016.05.006.

L. Botezatu, J. M. Hönes adn et al. (Feb. 2016). "GFI1 as a novel prognostic and therapeutic factor for AML/MDS". In: *Leukemia* 30, pp. 1237–1245. DOI: 10.1038/leu.2016.11.

Langmead, B. et al. (Mar. 2009). "Ultrafast and memory-efficient alignment of short DNA sequences to the human genome". In: *Genome Biology* 10, pp. 693–705. DOI: 10.1186/gb-2009-10-3-r25.

Laurent, Stéphane, B. V. Almende, and vis.js contributors (2020). *graph3d: A Wrapper of the JavaScript Library 'vis-graph3d'*. R package version 0.2.0. URL: https://CRAN.R-project.org/package=graph3d.

Law, J. et al. (Dec. 2020). "Expression and function of modlator of apoptosis (MOAP-1) in blood cancers". In: *Research Square*. DOI: 10.21203/rs.3.rs-109180/v1.

Lawrence, M. et al. (Aug. 2013a). *Software for Computing and Annotating Genomic Ranges*. DOI: 10.1371/journal.pcbi.1003118.

Lawrence, M. S. et al. (July 2013b). "Mutational heterogeneity in cancer and the search for new cancer-associated genes". In: *Nature* 499, pp. 214–218. DOI: 10.1038/nature12213.

Lawrence, M. S. et al. (Jan. 2014). "Discovery and saturation analysis of cancer genes across 21 tumour types". In: *Nature* 505, pp. 495–501. DOI: 10.1038/nature12912.

Lee, T. I. and R. A. Young (Mar. 2013). "Transcriptional regulation and its misregulation in disease". In: *Cell* 152, pp. 1237–1251. DOI: 10.1016/j.cell.2013.02.014.

Legras, A. et al. (Aug. 2017). "Epithelial-to-mesenchymal transition and microRNAs in lung cancer". In: *Cancers* 9, p. 101. DOI: 10.3390/cancers9080101.

Lemischka, I. R., D. H. Raulet, and R. C. Mulligan (June 1986). "Developmental potential and dynamic behavior of hematopoietic stem cells". In: *Cell* 45, pp. 917–927. DOI: 10.1016/0092-8674(86)90566-0.

Levin, M. et al. (Jan. 2021). "Deciphering molecular mechanisms underlying chemoresistance in relapsed AML patients: towards precision medicine overcoming drug resistance". In: *Cancer Cell Int.* 21. DOI: 10.1186/s12935-021-01746-w.

Li, H. et al. (Aug. 2009). "The sequence alignment/map format and SAMtools". In: *Bioinformatics*, pp. 2078–79. DOI: 10.1093/bioinformatics/btp352.

Li, S. et al. (June 2016). "Distinct evolution and dynamics of epigenetic and genetic heterogenetiy in acute myeloid leukemia". In: *Nature Medicine* 22, pp. 792–799. DOI: 10.1038/nm.4125.

Li, Z. et al. (Feb. 2019). "Identification of transcription factor binding sites using ATAC-seq". In: *Genome Biology* 20. DOI: 10.1186/s13059-019-1642-2.

Lichtman, M. A. (Feb. 2013). "A historical perspective on the development of the cytarabine (7 days) and daunorubicin (3 days) treatment regimen for acute myelogenous leukemia: 2013 the 40th anniversary of 7+3". In: *Blood Cells, Molecules and Diseases*, pp. 119–130. DOI: 10.1016/j.bcmd.2012.10.005.

Liebermann, D. A. and B. Hoffman (July 2007). "GADD45 in the response of hematopoietic cells to genotoxic stress". In: *Blood Cells, Molecules, and Diseases* 39, pp. 329–335. DOI: 10.1016/j.bcmd.2007.06.006.

Lin, C. Y. et al. (Jan. 2016). "Active medulloblastoma enhancers reveal subgroup-specific cellular origins". In: *Nature*, pp. 57–62. DOI: 10.1038/nature16546.

Liu, B. et al. (July 2019). "Interplay between miRNAs and host genes and their role in cancer". In: *Briefings in Functional Genomics* 18, pp. 255–266. DOI: 10.1093/bfgp/elz002.

Liu, P. et al. (Jan. 2021). "Nrf2 overexpression increases risk of high tumor mutation burden in acute myeloid leukemia by inhibiting MSH2". In: *Cell Death & Disease*. DOI: 10.1038/s4141-020-03331-x.

Love, M. I., W. Huber, and S. Anders (Dec. 2014). "Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2". In: *Genome Biology* 15. DOI: 10.1186/s13059-014-0550-8.

Lovén, J. et al. (Apr. 2013). "Selective inhibition of tumor oncogenes by disruption of super-enhancers". In: *Cell*, pp. 320–334. DOI: 10.1016/j.cell.2013.03.036.

Lu, Y. et al. (Apr. 2020). "Epigenetic regulation in human cancer: the potential role of epi-drug in cancer therapy". In: *Molecular Cancer* 19. DOI: 10.1186/s12943-020-01197-3.

Luger, K. and T. J. Richmond (Apr. 1998). "The histione tails of the nucleosome". In: *Current Opinion in Genetics & Development*, pp. 140–146. DOI: 10.1016/s0959-437x(98)80134-2.

Luger, K. et al. (Sept. 1997). "Crystal structure of the nucleosome core particle at 2.8 Å resolution". In: *Nature*, pp. 251–260. DOI: 10.1038/38444.

Luo et al. (2013). "Pathview: an R/Bioconductor package for pathway-based data integration and visualization". In: *Bioinformatics* 29.14, pp. 1830–1831. DOI: 10.1093/bioinformatics/btt285.

Luo, Z. and C. Lin (Feb. 2016). "Enhancer, epigenetics, and human disease". In: *Current Opinion in Genetics & Development*, pp. 27–33. DOI: 10.1016/j.gde.2016.03.012.

Lyko, F. (Oct. 2018). "The DNA methyltransferase family: a versatile toolkit for epigenetic regulation". In: *Nature Reviews Genetics* 19, pp. 81–92. DOI: 10.1038/nrg.2017.80.

Mackay, L. K. et al. (Apr. 2016). "Hobit and Blimp1 instruct a universal transcriptional program of tissue residency in lymphocytes". In: *Science* 352, pp. 459–463. DOI: 10.1126/science.aad2035.

Malaise, M. et al. (June 2011). "Stable and reproducible engraftment of primary adult and pediatric acute myeloid leukemia in NSG mice". In: *Leukemia*, pp. 1635–1639. DOI: 10.1038/leu.2011.121.

Mao, G. et al. (Feb. 2008). "Preferential loss of mismatch repair function in refractory and relapsed acute myeloid leukemia: potential contribution to AML progression". In: *Cell Research*, pp. 281–289. DOI: 10.1038/cr.2008.14.

Martino, O. di and J. S. Welch (Dec. 2019). "Retinoic acid receptors in acute myeloid leukemia therapy". In: *Cancers* 11. DOI: 10.3390/cancers11121915.

Martins, G. and K. Calame (Apr. 2008). "Regulation and functions of Blimp-1 in T and B lymphocytes". In: *Annual Review of Immunology* 26, pp. 133–169. DOI: 10.1146/annurev.immunol.26.021607.090241.

McCarthy, Davis J, Yunshun Chen, and Gordon K Smyth (May 2012). "Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation". In: *Nucleic Acids Research* 40, pp. 4288–4297. DOI: 10.1093/nar/gks042.

McGhee, J. D. and G. Felsenfeld (1980). "Nucleosome structure". In: *Annual Review of Biochemistry*, pp. 1115–1156. DOI: 10.1146/annurev.bi.49.070180.005343.

McGinty, R. K. and S. Tan (Mar. 2015). "Nucleosome structure and function". In: *Chemical Reviews*, pp. 2255–2273. DOI: 10.1021/cr500373h.

McKeown, Michael et al. (July 2017). "Superenhancer Analysis Defines Novel Epigenomic Subtypes of Non-APL AML, Including an RARα Dependency Targetable by SY-1425, a Potent and Selective RARα Agonist". In: *Cancer Discovery* 7, pp. CD–17. DOI: 10.1158/2159-8290.CD-17-0399.

McLean, C. Y. et al. (May 2010). "GREAT improves functional interpretation of cis-regulatory regions". In: *Nature Biotechnology*. DOI: 10.1038/nbt.1630.

Méndez-Ferrer, S. et al. (Feb. 2020). "Bone marrow niches in haematological malignancies". In: *Nature Reviews Cancer* 20, pp. 285–298. DOI: 10.1038/s41568-020-0245-2.

Metzeler, K. H. et al. (Aug. 2016). "Spectrum and prognostic relevance of driver gene mutations in acute myeloid leukemia". In: *Blood* 128, pp. 686–698. DOI: 10.1182/blood-2016-01-693879.

Meyers, J. et al. (May 2013). "Medicare fee-for-service enrollees with primary acute myeloid leukemia: an analysis of treatment patterns, survival, and healthcare resource utilization and costs". In: *Applied Health Economics and Health Policy*, pp. 275–286. DOI: 10.1007/s40258-013-0032-2.

Miguel-Escalada, I., L. Pasquali, and J. Ferrer (Aug. 2015). "Transcriptional enhancers: functional insights and role in human disease". In: *Current Opinion in Genetics & Development*, pp. 71–76. DOI: 10.1016/j.gde.2015.08.009.

Miltenyi Biotec. *AutoMACS Pro Separator*. https://www.miltenyibiotec.com/US-en/products/automacs-pro-separator-starter-kit.html. Last accessed Apr 2021.

Minnoye, L. et al. (Jan. 2021). "Chromatin accessibility profiling methods". In: *Nature Reviews Methods Primers* 1. DOI: 10.1038/s43586-020-00008-9.

Mohammad, H. P., O. Barbash, and C. L. Creasy (Mar. 2019). "Targeting epigenetic modifications in cancer therapy: erasing the roadmap to cancer". In: *Nature Medicine* 25, pp. 403–418. DOI: 10.1038/s41591-019-0376-8.

Montalvo-Casimiro, M. et al. (Nov. 2020). "Epidrug repurposing: discovering new faces of old acquaintances in cancer therapy". In: *Frontiers in Oncology* 10. DOI: 10.3389/fonc.2020.605386.

Moore, J. E. et al. (Jan. 2020). "A curated benchmark of enhancer-gene interactions for evaluating enhancer-target gene prediction methods". In: *Genome Biology* 21.17. DOI: 10.1186/s13059-019-1924-8.

Morgan, M. et al. (2019). *SummarizedExperiment*. R package version 1.16.1. DOI: 10.18129/B9.bioc.SummarizedExperiment.

Morita, K. et al. (Oct. 2020). "Clonal evolution of acute myeloid leukemia revealed by high-throughput single-cell genomics". In: *Nature Communications* 11.5327. DOI: 10.1038/s41467-020-19119-8.

Murayama, T. and N. Gotoh (June 2019). "Patient-derived xenograft models of breast cancer and their application". In: *Cells* 8. DOI: 10.3390/cells8060621.

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

Niederriter, A. R. et al. (Nov. 2015). "Super-enhancers in cancers, complex disease, and developmental disorders". In: *Genes* 6, pp. 1183–1200.

NIH. *National Human Genome Research Institute*. https://www.ohsu.edu/knight-cancer-institute/understanding-leukemia. Last accessed Apr 2021.

Ocaña, O. H. et al. (Dec. 2012). "Metastatic colonization requires the repression of the epithelial-mesenchymal transition inducer Prrx1". In: *Cancer Cell* 22, pp. 709–724. DOI: 10.1016/j.ccr.2012.10.012.

OHSU. *OHSU Knight Cancer Institute*. https://www.genome.gov/genetics-glossary/Chromatin. Last accessed Sep 2019.

Oliva, E. N. et al. (Nov. 2018). "The real-world incidence of relapse in acute myeloid leukemia (AML): A systematic literature review (SLR)". In: *Blood* 132. DOI: 10.1182/blood-2018-99-111839.

Ong, C.-T. and V. G. Corces (Apr. 2011). "Enhancer function: new insights into the regulation of tissue-specific gene expression". In: *Nature Reviews Genetics* 12, pp. 283–293. DOI: 10.1038/nrg2957.

O'Shaughnessy, J. (Oct. 2005). "Extending survival with chemotherapy in metastatic breast cancer". In: *The Oncologist* 10, pp. 20–29. DOI: 10.1634/theoncologist.10-90003-20.

Ott, C. J. et al. (Dec. 2018). "Enhancer architecture and essential core regulatory circuitry of chronic lymphocytic leukemia". In: *Cancer Cell* 34, pp. 982–995. DOI: 10.1016/j.ccell.2018.11.001.

Overacre-Delgoffe, A. E. et al. (June 2017). "Interferon-γ drives Treg fragility to promote anti-tumor immunity". In: *Cell* 169, pp. 1130–1141. DOI: 10.1016/j.cell.2017.05.005.

Padmanaban, V. et al. (Sept. 2019). "E-cadherin is required for metastasis in multiple models of breast cancer". In: *Nature* 573, pp. 439–444. DOI: 10.1038/s41586-019-1526-3.

Palmieri, R. et al. (Jan. 2020). "Therapeutic choice in older patients with acute myeloid leukemia: a matter of fitness". In: *Cancers* 12. DOI: 10.3390/cancers12010120.

Papaemmanuil, E. et al. (June 2016). "Genomic Classification and Prognosis in Acute Myeloid Leukemia". In: *N Engl J Med* 374, pp. 2209–2221. DOI: 10.1056/NEJMoa1516192.

Pastushenko, I. et al. (Apr. 2018). "Identification of the tumor transition states occurring during EMT". In: *Nature* 556, pp. 463–468. DOI: 10.1038/s41586-018-0040-3.

Pende, D. et al. (Mar. 2005). "Analysis of the receptor-ligand interactions in the natural killer-mediated lysis of freshly isolated myeloid or lymphoblastic leukemias". In: *Blood* 105, pp. 2066–2073. DOI: 10.1182/blood-2004-09-3548.

Phan, N. N., T. T: Huynh, and Y.-C. Lin (Aug. 2017). "Hyperpolarization-activated cyclic nucleotide-gated gene signatures and poor clinical outcome of cancer patients". In: *Translational Cancer Research* 6. DOI: 10.21037/tcr.2017.07.22.

Polansky, J. K. et al. (Oct. 2010). "Methylation matters: binding of Ets1 to the demethylated Foxp3 gene contributes to the stabilization of Foxp3 expression in regulatory T cells". In: *Journal of Molecular Medicine* 88, pp. 1029–1040. DOI: 10.1007/s00109-010-0642-1.

Pott, S. and J. D. Lieb (Jan. 2015). "What are super-enhancers?" In: *Nature Genetics* 47, pp. 8–12. DOI: 10.1038/ng.3167.

Quang, D. X. et al. (July 2015). "Motif signatures in stretch enhancers are enriched for disease-associated genetic variants". In: *Epigenetics & Chromatin*. DOI: 10.1186/s13072-015-0015-7.

Rada-Iglesias, A. et al. (Dec. 2011). "A unique chromatin signature uncovers early developmental enhancers in humans". In: *Nature* 470, pp. 279–283. DOI: 10.1038/nature09692.

Ramsey, S. A. et al. (Sept. 2010). "Genome-wide histone acetylation data improve prediction of mammalian transcription factor binding sites". In: *Bioinformatics* 26, pp. 2071–2075. DOI: 10.1093/bioinformatics/btq405.

Rhee, H. S. and B. Franklin Pugh (Dec. 2012). "Comprehensive genome-wide protein-DNA interactions detected at single nucleotide resolution". In: *Cell* 147, pp. 1408–1419. DOI: 10.1016/j.cell.2011.11.013.

Robinson, J. T. et al. (Jan. 2011). "Integrative Genomics Viewer". In: *Nature Biotechnology* 29, pp. 24–26. DOI: 10.1038/nbt.1754.

Robson, M. I., A Ringel, and S. Mundlos (June 2020). "Regulatory landscaping: how enhancer-promoter communication is sculpted in 3D". In: *Molecular Cell* 74, pp. 1110–1122. DOI: 10.1016/j.molcel.2019.05.032.

Roeder, R. G. (Feb. 2005). "Transcriptional regulation and the role of the diverse coactivators in animal cells". In: *FEBS Letters* 579, pp. 909–915. DOI: 10.1016/j.febslet.2004.12.007.

Roth, S. Y., J. M. Denu, and C. D. Allis (July 2001). "Histone acetyltranferases". In: *Annual Review of Biochemistry* 70, pp. 81–120. DOI: 10.1146/annurev.biochem.70.1.81.

Rutishauser, R. L. et al. (Aug. 2009). "Transcriptional repressor Blimp-1 promotes $CD8^+$ T cell terminal differentiation and represses the acquisition of central memory T cell properties". In: *Immunity* 31, pp. 296–308. DOI: 10.1016/j.immuni.2009.05.014.

Sánchez, P. V. et al. (2009). "A robust xeno-transplantation model for acute myeloid leukemia". In: *Leukemia* 23, pp. 2109–2117. DOI: 10.1038/leu.2009.

Sánchez-Tilló, E. et al. (June 2010). "ZEB1 represses E-cadherin and induces an EMT by recruiting the SWI/SNF chromatin-remodeling protein BRG1". In: *Oncogene* 29, pp. 3490–34500. DOI: 10.1038/onc.2010.102.

Sandén, C. et al. (Feb. 2020). "Clonal competition within complex evolutionary hierarchies shapes AML over time". In: *Nature Communications* 11.579. DOI: 10.1038/s41467-019-14106-0.

Santamaría, C. M. et al. (Dec. 2009). "High FOXO3a expression is associated with a poorer prognosis in AML with normal cytogenetics". In: *Leukemia Research* 33, pp. 1706–1709. DOI: 10.1016/j.leukres.2009.04.024.

Satoh, K. et al. (Sept. 2014). "Involvement of epithelial-to-mesenchymal transition in the development of pancreatic ductal adenocarcinoma". In: *Journal of Gastroenterology* 50, pp. 140–146. DOI: 10.1007/s00535-014-0997-0.

Schmidt, J. M. et al. (Jan. 2015). "Stem-cell-like properties and epithelial plasticity arise as stable traits after transient Twist1 activation". In: *Cell Reports* 10, pp. 131–139. DOI: 10.1016/j.celrep.2014.12.032.

Schoenfelder, S. and P. Fraser (May 2019). "Long-range enhancer-promoter contacts in gene expression control". In: *Nature Reviews Genetics* 20, pp. 437–455. DOI: 10.1038/s41576-019-0128-0.

Schultz, L. D., F. Ishikawa, and D. L. Greiner (Feb. 2007). "Humanized mice in translational biomedical research". In: *Nature Reviews Immunology* 7, pp. 118–130. DOI: 10.1038/nri2017.

Schwenger, E. and U. Steidl (May 2021). "An evolutionary approach to clonally complex hematologic disorders". In: *Blood Cancer Discovery*, pp. 201–215. DOI: 10.1158/2643-3230.BCD-20-0219.

Shapiro-Shelef, M. et al. (Oct. 2003). "Blimp-1 is required for the formation of immunoglobulin secreting plasma cells and pre-plasma memory B cells". In: *Immunity* 19, pp. 607–620. DOI: 10.1016/s1074-7613(03)00267-x.

Shlush, L. I. et al. (Feb. 2014). "Identification of pre-leukaemic haematopoietic stem cells in acute leukaemia". In: *Nature* 506, pp. 328–333. DOI: 10.1038/nature13038.

Shlyueva, D., G. Stampfel, and A. Star (Mar. 2014). "Transcriptional enhancers: from properties to genome-wide predictions". In: *Nature Reviews Genetics* 15, pp. 272–286. DOI: 10.1038/nrg3682.

Silva-Coelho, P. da et al. (Apr. 2017). "Clonal evolution in myelodysplastic syndromes". In: *Nature Communications* 15099. DOI: 10.1038/ncomms15099.

Sinha, D. et al. (Nov. 2020). "Emerging concepts of hybrid epithelial-to-mesenchymal transition in cancer progression". In: *Biomolecules* 10. DOI: 10.3390/biom10111561.

Smallwood, A. and B. Ren (June 2013). "Genome organization and long-range regulation of gene expression by enhancers". In: *Current Opinion in Cell Biology* 25, pp. 387–394. DOI: 10.1016/j.ceb.2013.02.005.

Spicuglia, S. and L. Vanhille (Mar. 2012). "Chromatin signatures of active enhancers". In: *Nucleus* 3, pp. 126–131. DOI: 10.4161/nucl.19232.

Spiekermann, K. et al. (Aug. 2001). "Constitutive activation of STAT transcription factors in acute myelogenous leukemia". In: *Eur. J. Haematol.* 67, pp. 63–71.

Spitz, F. (Aug. 2012). "Transcription factors: from enhancer binding to developmental control". In: *Nature Reviews Genetics* 13, pp. 613–626. DOI: 10.1038/nrg3207.

Stief, S. M. et al. (June 2019). "Loss of KDM6A confers drug resistance in acute myeloid leukemia". In: *Leukemia* 34, pp. 50–62. DOI: 10.1038/s41375-019-0497-6.

Struhl, K. (May 1998). "Histone acetylation and transcriptional regulatory mechanisms". In: *Genes & Development* 12, pp. 599–606. ISSN: 10.1101/gad.12.5.599.

Sur, I. and J. Taipale (July 2016). "The role of enhancers in cancer". In: *Nature Reviews Cancer* 16, pp. 483–494. DOI: 10.1038/nrc.2016.62.

Surveillance, Epidemiology and End Results Program. *SEER*. https://seer.cancer.gov/statfacts. Last accessed Mar 2021.

Swaninathan, M. et al. (Nov. 2018). "Clinical heterogeneity of AML is associated with mutational heterogeneity". In: *Blood* 132.5240. DOI: 10.1182/blood-2018-99-117287.

Swerdlow, S. H. et al. (May 2016). "The 2016 revision of the World Health Organization classification of lymphoid neoplasms". In: *Blood*, pp. 2375–90. DOI: 10.1182/blood-2016-01-643569.

Tamura, R. E. et al. (June 2012). "GADD45 proteins: central players in tumorigenesis". In: *Current Molecular Medicine*, pp. 634–651. DOI: 10.2174/156652412800619978.

Tang, W. W. C. et al. (June 2015). "A unique gene regulatory network resets the human germline epigenome for development". In: *Cell* 161, pp. 1453–1467. DOI: 10.1016/j.cell.2015.04.053.

Teeuwssen, M. and R. Foddle (Sept. 2019). "Cell heterogeneity and phenotypic plasticity in metastasis formation: the case of colon cancer". In: *Cancers* 11. DOI: 10.3390/cancers11-091368.

Terziyska, N. et al. (Dec. 2012). "In vivo imaging enables high resolution preclinical trials on patients´ leukemia cells growing in mice". In: *PLoS ONE* 7. DOI: 10.1371/journal.pone.0052798.

Testa, J. R. et al. (Sept. 1979). "Evolution of karyotypes in acute nonlymphocytic leukemia". In: *Cancer Research* 39, pp. 3619–3627. PMID: 476688.

Testa, U., E. Pelosi, and G. Castelli (Feb. 2021). "Precision medicine treatment in acute myeloid leukemia is not a dream". In: *Hemato* 2, pp. 131–153. DOI: 10.3390/hemato2010008.

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

Trimboli, A. J. et al. (Feb. 2008). "Direct evidence for epithelial-mesenchymal transition in breast cancer". In: *Cancer Research* 68, pp. 937–945. DOI: 10.1158/0008-5472.CAN-07-2148.

Tsai, J. H. et al. (Dec. 2012). "Spatiotemporal regulation of epithelial-mesenchymal transition is essential for squamous cell carcinoma metastasis". In: *Cancer Cell* 22, pp. 725–736. DOI: 10.1016/j.ccr.2012.09.22.

Turner, B. M. (Aug. 2000). "Histone acetylation and an epigenetic code". In: *BioEssays* 22, pp. 836–845. DOI: 10.1002/1521-1878(200009)22:9<836::AID-BIES9>3.0.CO;2-X.

Vannier, C. et al. (Mar. 2013). "Zeb1 regulates E-cadherin and Epcam (epithelial cell adhesion molecule) expression to control cell behavior in early zebrafish development". In: *The Journal of Biological Chemistry* 288, pp. 18643–59. DOI: 10.1074/jbc.M113.467787.

Vardiman, J. W., N. L. Harris, and R. D. Brunning (Oct. 2002). "The World Health Organization (WHO) classification of the myeloid neoplasms". In: *Blood*, pp. 2292–302. DOI: 10.1182/blood-2002-04-1199.

Verdone, L. et al. (July 2006). "Histone acetylation in gene regulation". In: *Briefings in Functonal Genomics and Proteomics* 5, pp. 209–221. DOI: 10.1093/bfgp/ell028.

Vicente-Dueñas, C. et al. (June 2018). "Epigenetic priming in cancer initiation". In: *Trends in Cancer* 4, pp. 408–417. DOI: 10.1016/j.trecan.2018.04.007.

Vick, B. et al. (Mar. 2015). "An advanced preclinical mouse model for acute myeloid leukemia using patients cells of various genetic subgroups and *in vivo* bioluminescence imaging". In: *PLoS ONE*. DOI: 10.1371/journal.pone.0120925.

Visel, A., E. M. Rubin, and L. A. Pennacchio (Sept. 2009). "Genomic views of distant-acting enhancers". In: *Nature* 461, pp. 199–205. DOI: 10.1038/nature08451.

Vosberg, S. and P. A. Greif (May 2019). "Clonal evolution of acute myeloid leukemia from diagnosis to relapse". In: *Genes, Chromosomes and Cancer* 58, pp. 839–849. DOI: 10.1002/gcc.22806.

Waddington, C. H. (Nov. 1942). "Canalization of development and the inheritance of acquired characters". In: *Nature* 150, pp. 563–565. DOI: 10.1038/150563a0.

Wang, A. and H. Zhong (Dec. 2018). "Roles of the bone marrow niche in hematopoiesis, leukemogenesis, and chemotherapy resistance in acute myeloid leukemia". In: *Hematology* 23, pp. 729–739. DOI: 10.1080/10245332.2018.1486064.

Wang, B. et al. (May 2016). "Mutational spectrum and risk stratification of intermediate-risk acute myeloid leukemia patients based on next-generation sequencing". In: *Oncotarget* 7, pp. 32065–32078. DOI: 10.18632/oncotarget.7028.

Warnes, Gregory R. et al. (2017). *gdata: Various R Programming Tools for Data Manipulation*. R package version 2.18.0. URL: https://CRAN.R-project.org/package=gdata.

Weake, V. M. and J. L. Workman (Apr. 2010). "Inducible gene expression: diverse regulatory mechanisms". In: *Nature Reviews Genetics* 11, pp. 426–437. DOI: 10.1038/nrg2781.

Werner, S. et al. (Aug. 2013). "Dual roles of the transcription factor graiyhead-like 2 (GRHL2) in breast cancer". In: *The Journal of Biological Chemistry* 288, pp. 22993–3008. DOI: 10.1074/jbc.M113.456293.

Whyte, W. A. et al. (Apr. 2013). "Master transcription factors and mediator establish super-enhancers at key cell identity genes". In: *Cell* 153.2, pp. 307–319. DOI: 10.1016/j.cell.2013.03.035.

Wickham, Hadley (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. ISBN: 978-3-319-24277-4. URL: https://ggplot2.tidyverse.org.

World Health Organization. *WHO*. http://www.euro.who.int/__data/assets/pdf_file/0005/97016/4.1.-Incidence-of-childhood-leukaemia-EDITED_layouted.pdf. Last accessed Feb 2020.

World Medical Association. *WMA*. https://www.wma.net/policies-post/wma-declaration-of-helsinki-ethical-principles-for-medical-research-involving-human-subjects/. Last accessed Feb 2021.

Wu, Mengjun and Lei Gu (2019). *TCseq: Time course sequencing data analysis*. R package version 1.10.0. DOI: 10.18129/B9.bioc.TCseq.

Xu, J. et al. (July 2019). "Epigenetic abnormalities in acute myeloid leukemia and leukemia stem cells". In: *Advances in Experimental Medicine and Biology* 1143, pp. 173–189. DOI: 10.1007/978-981-13-7342-8_8.

Yagi, H. et al. (Nov. 2004). "Crucial role of Foxp3 in the development and function of human $CD25^{+}CD4^{+}$ regulatory T cells". In: *International Immunology* 16, pp. 1643–1656. DOI: 10.1093/intimm/dxh165.

Yang, X.-J. (Feb. 2004). "The diverse superfamily of lysine acetyltransferases and their roles in leukemia and other diseases". In: *Nucleic Acids Research* 32, pp. 959–976. DOI: 10.1093/nar/gkh252.

Yang, X. O. et al. (July 2008). "Molecular antagonism and plasticity of regulatory and inflammatory T cell programs". In: *Immunity* 29, pp. 44–56. DOI: 10.1016/j.immuni.2008.05.007.

Yi, S.-J. and K. Kim (May 2018). "Histone tail cleavage as a novel epigenetic regulatory mechanism for gene expression". In: *BMB Reports* 51, pp. 211–218. DOI: 10.5483/bmbrep.2018.51.5.053.

Yu, Guangchuang et al. (2012). "clusterProfiler: an R package for comparing biological themes among gene clusters". In: *OMICS: A Journal of Integrative Biology* 16.5, pp. 284–287. DOI: 10.1089/omi.2011.0118.

Yu, M. et al. (Feb. 2013). "Circulating breast tumor cells exhibit dynamic changes in epithelial and mesenchymal composition". In: *Science* 339, pp. 580–584. DOI: 10.1126/science.1228522.

Yue, X. et al. (Mar. 2016). "Control of Foxp3 stability through modulation of TET activity". In: *Journal of Experimental Medicine* 213, pp. 377–397. DOI: 10.1084/jem.20151438.

Zaret, K. S. and J. S. Carroll (Nov. 2011). "Pioneer transcription factors: establishing competence for gene expression". In: *Genes & Development* 25, pp. 2227–2241. DOI: 10.1101/gad.176826.111.

Zeidner, J. F. et al. (Sept. 2018). "Final results of a randomized multicenter phase II study of alvocidib, cytarabine, and mitoxantrone versus cytarabine and daunorubicin (7+3) in newly diagnosed high-risk acute myeloid leukemia (AML)". In: *Leukemia Research* 72, pp. 92–95. DOI: 10.1016/j.leukres.2018.08.005.

Zentner, G. E., P. J. Tesar, and P. C. Scacheri (Aug. 2011). "Epigenetic signatures distinguish multiple classes of enhancers with distinct cellular functions". In: *Genome Research* 21, pp. 1273–1283. DOI: 10.1101/gr.122382.111.

Zhang, J., Y. Gu, and B. Chen (Mar. 2019). "Mechanisms of drug resistance in acute myeloid leukemia". In: *Onco Targets Therapy* 12, pp. 1937–1945. DOI: 10.2147/OTT.S191621.

Zheng, X. et al. (Nov. 2015). "Epithelial-to-mesenchymal transition is dispensable for metastasis but induces chemoresistance in pancreatic cancer". In: *Nature* 527, pp. 525–530. DOI: 10.1038/nature16064.

Zhou, K., G. Gaullier, and K. Luger (Dec. 2018). "Nucleosome structure and dynamics are coming of age". In: *Nature Structural & Molecular Biology* 26, pp. 3–13. DOI: 10.1038/s41594-018-0166-x.

Zhou, X. et al. (July 2009). "Instability of the transcription factor Foxp3 leads to the generation of pathogenic memory T cells in vivo". In: *Nature Immunology* 10, pp. 1000–1007. DOI: 10.1038/ni.1774.