Dissertation

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

Presented by
Lara Catherine Klett, M.Sc.
born in Freiburg im Breisgau, Germany
Oral examination: 17.05.2021

Disease- and drug-induced chromatin state changes in blood cancers

Referees: Prof. Dr. Karsten Rippe

Prof. Dr. Benedikt Brors

This work was performed from July 2015 to March 2021 under the supervision of Prof. Dr. Karsten Rippe in the Division Chromatin Networks at the German Cancer Research Center (DKFZ) and the BioQuant Center in Heidelberg, Germany.

Declaration

I hereby declare that I have written the submitted dissertation “Disease- and drug-induced chromatin state changes in blood cancers” myself and in this process, have used no other sources or materials than those explicitly indicated. I hereby declare that I have not applied to be examined at any other institution, nor have I used the dissertation in this or any other form at any other institution as an examination paper, nor submitted it to any other faculty as a dissertation.

_____ _____

(Place, Date) Lara Catherine Klett

Für meine Familie

Table of contents

List of publications

During this thesis, I contributed to the below listed publications:

Mallm, J., Iskar, M., Ishaque, N., Klett, L. C., Kugler, S. J., Muino, J. M., Teif, V. B., Poos, A. M., Großmann, S., Erdel, F., Tavernari, D., Koser, S. D., Schumacher, S., Brors, B., König, R., Remondini, D., Vingron, M., Stilgenbauer, S., Lichter, P., Zapatka, M., Mertens, D., Rippe, K. (2019). Linking aberrant chromatin features in chronic lymphocytic leukemia to transcription factor networks. *Molecular Systems Biology*, 15(5), 1–20. https://doi.org/10.15252/msb.20188339

Chaturvedi, A., Herbst, L., Pusch, S., Klett, L., Goparaju, R., Stichel, D., Kaulfuss, S., Panknin, O., Zimmermann, K., Toschi, L., Neuhaus, R., Haegebarth, A., Rehwinkel, H., Hess-Stumpp, H., Bauser, M., Bochtler, T., Struys, E. A., Sharma, A., Bakkali, A., Geffers, R., Araujo-Cruz, M. M., Thol, F., Gabdoulline, R., Ganser, A., Ho, A. D., von Deimling, A., Rippe, K., Heuser, M., Kramer, A. (2017). Pan-mutant-IDH1 inhibitor BAY1436032 is highly effective against human IDH1 mutant acute myeloid leukemia in vivo. *Leukemia, 31*(10), 2020-2028. https://doi.org/10.1038/leu.2017.46

Wierzbinska, J. A., Toth, R., Ishaque, N., Rippe, K., Mallm, J. P., Klett, L. C., Mertens, D., Zenz, T., Hielscher, T., Seifert, M., Küppers, R., Assenov, Y., Lutsik, P., Stilgenbauer, S., Roessner, P. M., Seiffert, M., Byrd, J., Oakes, C. C., Plass, C., Lipka, D. B. (2020). Methylome-based cell-of-origin modeling (Methyl-COOM) identifies aberrant expression of immune regulatory molecules in CLL. *Genome Medicine*, 12(1), 1–19. https://doi.org/10.1186/s13073-020-00724-7

Hanna, B. S., Llao Cid, L., Iskar, M., Roessner, P. M., Klett, L. C., Ioannou, N., Öztürk, S., Mack, N., Kalter, V., Colomer, D., Campo, E., Bloehdorn, J., Stilgenbauer, S., Dietrich, S., Schmidt, M., Gabriel, R., Rippe, K., Feuerer, M., Ramsay, A., Zapatka, M., Lichter, P., Seiffert, M. IL-10 receptor signaling limits CD8 T-cell activation preventing their terminal exhaustion and tumor immune escape. *Immunity*. (under revision)

Muckenhuber, M., Lander, I., Müller-Ott, K., Erdel, F., Mallm, J.-P., Klett, L. C., Bauer, C., Hechler, J., Kepper, N., Rippe, K. Epigenetic signals that direct cell type specific interferon beta response in mouse cells. (submitted)

Summary

Malignant transformation of cells from the lymphoid lineage can occur at various developmental stages. These stages carry distinct patterns of epigenetic modifications as a result of dynamic epigenetic programming during physiological cell maturation that enables stage-specific gene regulation. Transformation frequently involves the disruption of these epigenetic patterns, which can potentially be modulated by drug treatment. Thus, epigenetic profiles of cancer cells comprise both disease-specific modifications as well as signatures that reflect the original cell type and developmental stage of the cell-of-origin from which they arise. Deregulated epigenetic signaling via activity changes of promoters and enhancers is likely to be a major factor for the establishment and maintenance of cancer specific gene expression patterns. The activity of regulatory promoter and enhancer elements and the binding of transcription factors in the genome can be mapped by the assay for transposase-accessible chromatin using sequencing (ATAC-seq). Thus, ATAC-seq is a highly informative readout about the local epigenetic state of chromatin. However, the epigenetic profiles of the non-malignant tumor cell-of-origin needs to be considered to reveal by contrast disease-specific aberrant gene regulation mechanisms and analyze the effect of drug treatment on deregulated chromatin patterns. Here, I dissected deregulated chromatin states in blood cancers by mapping the chromatin accessibility of gene regulatory elements and their relation to gene expression during physiological B-cell development, malignant transformation and drug treatment. First, I established and adjusted the experimental and bioinformatic procedures with quality controls to map chromatin accessible regions by ATAC-seq in primary human cells. By applying this method, I could demonstrate that the chromatin accessibility pattern of the genome changes during B-cell development and can be used to reconstruct the developmental phylogeny of sorted healthy B-cell subsets. Next, I characterized how the chromatin accessibility landscape was disrupted in chronic lymphocytic leukemia (CLL) compared to $CD19^{+}$ B cells of healthy donors. The epigenetic deregulation affected in addition to many enhancer and promoter regions also loci with repressive or bivalent histone modifications. While enhancers with gained accessibility in CLL were enriched for transcription factor binding sites of NFAT and E2A, enhancers with lost accessibility were enriched for binding sites of EBF1. To further refine the definition of the physiological state, the accessibility-based phylogenetic tree of B-cell development was used to infer the maturation stage and the chromatin accessibility pattern of the cell-of-origin for each individual CLL patient. The results of the cell-of-origin *in silico* modeling for individual patients were in agreement with the IGHV-mutational status, an established marker for the maturation stage, and enabled a more detailed classification. The comparison of the CLL state with the direct cell-of-origin allowed to discriminate between developmental changes and disease-specific aberrant accessibility. With this

refined analysis I was able to follow chromatin accessibility changes in CLL patients that underwent standard treatment with ibrutinib, an inhibitor of disease-enhanced B-cell signaling. Regions with CLL-specific aberrant chromatin accessibility approached physiological accessibility levels with the course of treatment. However, a single-cell analysis of the DNA mutations revealed changes in the clonal heterogeneity in two patients during treatment. Furthermore, the presence of an ibrutinib-resistant subclone was detected already during remission. In order to further analyze the druggability of disrupted epigenetic patterns, the effect of treatment with the histone deacetylase inhibitor panobinostat was evaluated in multiple myeloma. Panobinostat induced a global increase in histone acetylation, mainly attributable to a broadening of acetylation within the gene body of expressed genes. No significant changes of accessibility were detected. However, the degree of transcriptional drug response between different patients and multiple myeloma cell lines varied considerably. Only the transcriptionally responsive cell lines showed a G1 cell cycle arrest and an enrichment of early senescence pathway genes in the panobinostat responding genes.

In summary, this thesis contributes to explore novel aspects of the deregulation of gene regulatory elements in blood cancers, which holds great potential for clinical application. The approach of identifying disease-specific changes in the chromatin accessibility and transcription factor occupancy while accounting for the tumor cell-of-origin profiles might also help to better understand epigenetic deregulation and the effect of drugs in other tumor entities.

Zusammenfassung

Während des gesamten Reifungsprozesses der Lymphopoese können Zellen verschiedener Entwicklungsstadien entarten und zur Bildung maligner Neoplasien führen. Der epigenetische Status des Genoms verändert sich während dieses physiologischen Reifungsprozesses um die für ein bestimmtes Entwicklungsstadium wichtigen Genregulationsprogramme zu ermöglichen. Die Transformation von Zellen geht oft einher mit der Modifikation dieser physiologischen epigenetischen Muster. Diese Veränderungen sind jedoch potentiell reversibel. Tumorzellen tragen sowohl die epigenetischen Muster des malignen Status als auch Muster des ursprünglichen Zelltyps sowie den Entwicklungsstatus der Tumor-initiierenden Zelle. Deregulierte epigenetische Signale spielen wahrscheinlich durch Aktivitätsänderungen von Promotoren und Enhancern eine große Rolle bei der Etablierung und Erhaltung von tumorspezifischen Genexpressionsmustern. Die Aktivität von regulatorischen Elementen und die Bindung von Transkriptionsfaktoren im Genom kann mittels des Assays für Transposase-zugängliches Chromatin durch Sequenzierung (ATAC-seq) erfasst werden. Damit ist ATAC-seq eine sehr informative Methode um den lokalen epigenetischen Status des Chromatins auszulesen. Um krankheitsspezifische Veränderungen von Genregulationsmechanismen zu identifizieren und den Effekt von Medikamenten auf diese deregulierten Chromatinmuster zu analysieren, muss allerdings auch das epigenetische Profil der nicht malignen Ursprungszelle des Tumors berücksichtigt werden. In der vorliegenden Arbeit habe ich deregulierte Chromatinzustände im Blutkrebs analysiert, indem ich die Chromatinzugänglichkeit an genregulatorischen Elementen und deren Relation zur Genexpression während der physiologischen B-Zell Entwicklung, malignen Transformation und unter Behandlung kartierte. Um Regionen von hoher Chromatinzugänglichkeit in primären humanen Zellen mittels ATAC-seq zu identifizieren, wurden zunächst die experimentellen und bioinformatischen Verfahren mit Qualitätskontrollschritten etabliert und angepasst. Mithilfe dieser Methoden konnte ich zeigen, dass sich die Chromatinzugänglichkeitsmuster des Genoms während der B-Zell Entwicklung verändern und genutzt werden können, um phylogenetische Beziehungsverhältnisse von sortierten gesunden B-Zell Untergruppen unterschiedlicher Entwicklungsstadien zu rekonstruieren. Anschließend charakterisierte ich Unterschiede in der Chromatinzugänglichkeit in Zellen von Patienten der chronischen lymphatischen Leukämie (CLL) im Vergleich zu $CD19^{+}$ B-Zellen von gesunden Spendern. Neben vielen Enhancer- und Promotorregionen betraf die Deregulation von epigenetischen Signalen auch Regionen, die durch reprimierende oder bivalente Histonmodifikationen gekennzeichnet waren. Enhancer mit einer erhöhten Chromatinzugänglichkeit in CLL zeigten eine Anreicherung von Transkriptionsfaktorbindestellen für NFAT und E2A, während Enhancer mit reduzierter Zugänglichkeit eine Überrepräsentation von Bindemotiven für EBF1 aufwiesen.

Um die Definition des physiologischen Zustands weiter zu verbessern, wurde das Reifungsstadium und damit das Chromatinzugänglichkeitsmuster der Ursprungszelle für jeden einzelnen CLL-Patienten mittels des auf Chromatinzugänglichkeit basierenden phylogenetischen Baums der B-Zell-Entwicklung ermittelt. Die Ergebnisse der *in silico* Modellierung des Reifestadiums der Tumor-initiierenden Zelle für individuelle Patienten stimmten mit dem IGHV-Mutationsstatus, dem etablierten Biomarker des Reifestatus, überein und ermöglichten darüber hinaus eine detailliertere Aufschlüsselung. Der Vergleich des Chromatinstatus in CLL mit dem der direkten Ursprungszelle ermöglichte die Unterscheidung zwischen entwicklungsbedingten und krankheitsspezifischen Änderungen der Chromatinzugänglichkeit. Regionen, die eine CLL-spezifische Änderung der Chromatinzugänglichkeit aufwiesen, zeigten während der Behandlung mit Ibrutinib, einem Inhibitor des krankheitsbedingt hyperaktiven B-Zellrezeptorsignalwegs, ein Angleichen des Chromatinzugänglichkeitsniveau an nicht maligne Zellen. Eine ergänzende Einzelzellanalyse zweier Patienten auf DNA-Mutationsebene offenbarte Veränderungen in der Subklonalstruktur durch die Behandlung. Darüber hinaus war ein später expandierender Ibrutinib-resistenter Subklon schon während der Remissionsphase detektierbar. Für eine weiterführende Analyse der Veränderlichkeit von epigenetischen Mustern durch Medikament wurden Effekte der Behandlung mit dem Histon-Deacetylase-Inhibitor Panobinostat im Multiplen Myelom untersucht. Panobinostat induzierte einen globalen Anstieg der Histon-Acetylierung, die vor allem auf die Ausbreitung der Acetylierung innerhalb der Genregion exprimierter Gene zurückzuführen war. Es wurden jedoch keine signifikanten Änderungen der Chromatinzugänglichkeit detektiert. Der Grad der transkriptionellen Antwort zwischen einzelnen Patienten und Multiplen Myelom Zelllinien variierte erheblich. Nur transkriptionell ansprechende Zelllinien wiesen einen G1-Zellzyklusarrest auf und zeigten in den auf Panobinostat ansprechenden Genen eine Anreicherung für Seneszenz-assoziierte.

Zusammenfassend trägt diese Arbeit dazu bei, neue Aspekte der Deregulation von genregulatorischen Elementen in Blutkrebs zu erforschen, welches großes Potential für die klinische Anwendung bietet. Der Ansatz krankheitsbedingte Veränderungen der Chromatinzugänglichkeit und Transkriptionsfaktorbindung unter Berücksichtigung der Profile der Tumor-initiierenden Zellen zu identifizieren, könnte dazu beitragen die epigenetische Deregulation und die Wirkung von Medikamenten auch in anderen Tumorentitäten besser zu verstehen.

List of figures

List of tables

Abbreviations

α-KG ..... α-ketoglutarate
ALL ..... Acute lymphoblastic leukemia
AML ..... Acute myeloid leukemia
a.u. ..... Arbitrary units
BAM ..... Binary Alignment Map (file format)
BED ..... Browser Extensible Data (file format)
bp ..... Base pairs
BSA ..... Bovine serum albumin
BTK ..... Bruton's tyrosine kinase
BTG1 ..... B-cell translocation gene 1 (ENSG00000133639)
CCND1 ..... Cyclin D1
cDNA ..... Complementary DNA
ChIP-seq ..... Chromatin immunoprecipitation followed by sequencing
Chr. ..... Chromosome
CI ..... Confidence interval
CLL ..... Chronic lymphocytic leukemia
CML ..... Chronic myeloid leukemia
CNV ..... Copy number variation
Conc ..... Concentration
C_t value ..... Cycle threshold value
CTLA4 ..... Cytotoxic T-lymphocyte associated protein 4
DMSO ..... Dimethyl sulfoxide
DNA ..... Deoxyribonucleic acid
EBF1 ..... EBF transcription factor 1 (ENSG00000164330)
EDTA ..... Ethylenediaminetetraacetic acid
FACS ..... Fluorescence-activated cell sorting
FBS ..... Fetal bovine serum
FCS-format ..... Flow cytometry standard file format
FDA ..... United States Food and Drug Administration
FDR ..... False discovery rate
FGFR3 ..... Fibroblast growth factor receptor 3
FPKM ..... Fragments per kilobase of transcript per million mapped reads
FRiP-score ..... Fraction of reads in peaks score
FSC ..... Forward scatter
fwd ..... Forward
GO ..... Gene ontology
HAT ..... Histone acetyl transferase
HDAC ..... Histone deacetylase
ID ..... Identifier

IGHV .....Immunoglobulin heavy-chain variable locus
IGV .....Integrative genomics viewer
kb .....Kilobase (=1,000 base pairs)
kDa.....Kilodalton
MACS.....Magnetic-activated cell sorting
mESC.....Mouse embryonic stem cells
min .....Minute(s)
miRNA .....MicroRNA
µm.....Micrometer
µM.....Micromolar
MM.....Multiple myeloma
MNase.....Micrococcal nuclease
MYC.....MYC proto-oncogene (ENSG00000136997)
NFAT .....Nuclear factor of activated T cells
nm.....Nanometer
nM.....Nanomolar
NSC .....Normalized strand cross-correlation coefficient
PAGE.....Polyacrylamide gel electrophoresis
PBMC.....Peripheral blood mononucleated cells
PBS .....Phosphate-buffered saline
PCA.....Principal component analysis
PCR .....Polymerase chain reaction
PIP2.....Phosphatidylinositol 4,5-bisphosphate
PIP3.....Phosphatidylinositol-3,4,5-triphosphate
PI3K.....Phosphatidylinositol-4,5-bisphosphate 3-kinase
PLCγ2 .....Phospholipase Cγ2
PMSF.....Phenylmethylsulfonyl fluoride
PTM .....Post-translational modification
PVDF .....Polyvinylidene fluoride
PVT1.....Pvt1 oncogene (ENSG00000249859)
qPCR .....quantitative real-time polymerase chain reaction
RAC3 .....Rac family small GTPase 3 (ENSG00000169750)
RefSeq.....Reference sequence (NCBI database)
rev.....reverse
RMSEP .....Root mean squared error of prediction
RNA .....Ribonucleic acid
RNA-seq .....RNA-sequencing
RSC .....Relative strand cross-correlation coefficient
SAM .....Sequence Alignment Map (file format)
SDS .....Sodium dodecyl sulfate
SF3B1.....Splicing Factor 3b Subunit 1 (ENSG00000115524)
SSC.....Side scatter
TAD.....Topologically associated domain
TF.....Transcription factor
TPM .....Transcripts per million
TSS.....Transcription start site

UMAP ..... Uniform manifold approximation and projection
UTR ..... Untranslated region
VAF ..... Variant allele frequency
WGBS ..... Whole-genome bisulfite sequencing
ZMYM3 ..... Zinc finger, MYM-type 3 gene
2-HG D-2-hydroxyglutarate

Introduction

1 Malignant transformation during B-cell maturation

Hematopoiesis is the process of the formation of the different blood cell types from hematopoietic stem cells, which primarily reside in the bone marrow (Rieger & Schroeder, 2012). Initiation of tumorigenesis can occur at different stages of this development and leads to distinct types of blood cancers. In particular, a variety of different malignancies can arise during the continuous differentiation of lymphoid progenitors into terminally differentiated plasma cells (Figure 1) (Cobaleda & Sanchez-Garcia, 2009; Rickert, 2013).

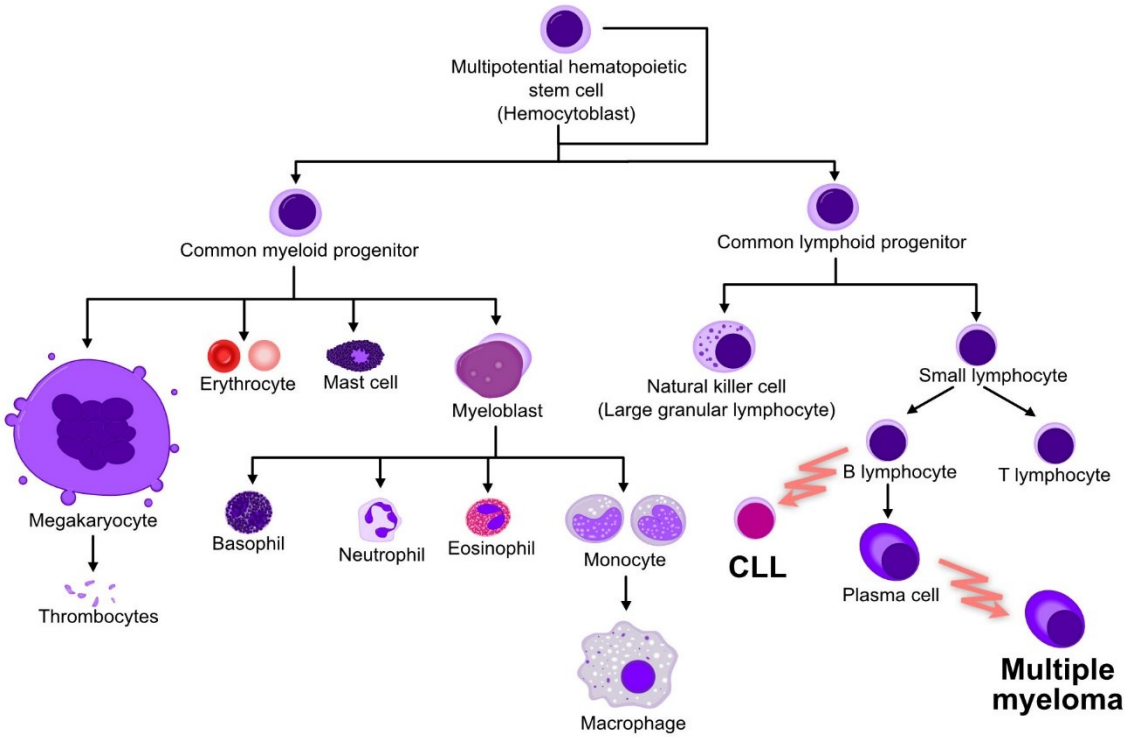


Figure 1 Malignant transformation from different maturation stages of the lymphoid lineage within hematopoiesis.
Chronic lymphocytic leukemia (CLL) and multiple myeloma (MM) originate from different B-cell maturation stages within the lymphoid lineage (Cobaleda & Sanchez-Garcia, 2009). Arrows represent malignant transformation events. Figure adapted from A. Rad and M. Häggström (https://commons.wikimedia.org/wiki/File:Hematopoiesis_simple.svg). Licensed under CC-BY-SA 3.0 (https://creativecommons.org/licenses/by-sa/3.0/).

Although in most cases the exact cell-of-origin is not well identified, the phenotypic similarity of tumor cells from individual tumor entities to normal stages of lymphoid development can be used to assign a given tumor to more or less differentiated stages. It is classically based on features like their microscopic appearance, the level of immunoglobulin gene rearrangement, expression signatures and surface antigens (Cobaleda & Sanchez-Garcia, 2009). Chronic lymphocytic leukemia (CLL) cells resemble B lymphocytes. (Cobaleda & Sanchez-Garcia, 2009; Oakes *et al.*, 2016; Seifert *et al.*, 2012). Differences in the maturation stage of the tumor initiating cell of individual patients have been observed within this single lymphoid neoplasm (Kulis *et al.*, 2015; Oakes *et al.*, 2016). Generally, acute leukemias of both the myeloid and lymphoid lineage (AML and ALL) show primarily an immature phenotype, while the chronic leukemias CML and CLL bear a resemblance to more mature cells (Cobaleda & Sanchez-Garcia, 2009; Krause, 2009). In CLL the maturity state of the malignant cells correlates with the aggressiveness of disease (Oakes *et al.*, 2016). Multiple myeloma (MM) is a disorder characterized by the clonal proliferation of a lymphocytic lineage cell at the final plasma cell differentiation stage (Cobaleda & Sanchez-Garcia, 2009; Rajkumar *et al.*, 2014). This thesis is focused on the lymphoid lineage disorders CLL and MM.

1.1 Interplay of mutations and epigenetics in the development and maintenance of lymphoid neoplasms

In recent years more and more evidence accumulated that hematopoietic malignancies have an important epigenetic component (Plass *et al.*, 2013; Xanthopoulos & Kostareli, 2019). The term “epigenetics” was coined by Conrad Waddington in 1942 (Waddington, 1942) to describe cellular differentiation processes that determine cell types in the absence of genetic changes. Since then, the definition of epigenetics has evolved to its current perception alongside the understanding of its molecular basis (Deans & Maggert, 2015). Frequently, the definition by Wu and Morris (2001) is used which describes epigenetics as “the study of changes in gene function that are mitotically and/or meiotically heritable and that do not entail a change in DNA sequence” (Wu & Morris, 2001). Epigenetic variation enables gene regulatory programs relevant for a certain developmental stage during physiological cell maturation (Smith & Meissner, 2013; Wu & Sun, 2006). However, epigenetics does not only play a role in physiological development, but also in disease pathogenesis including cancer (Baylin & Jones, 2011; Cai *et al.*, 2015). The most detailed analyses were so far conducted in hematopoietic malignancies. Epigenetic modifications are dynamic and in principle reversible contrary to changes in the DNA sequence. Thus, they are potentially interesting drug targets (Cai *et al.*, 2015).

Cancer is now recognized as disease characterized by both genetic and epigenetic alterations (Baylin & Jones, 2011). Genetic and epigenetic variation cannot be considered as completely separate events but are highly interconnected. It is emerging that the mutational events that lead to tumorigenesis are dependent on the epigenetic context (George *et al.*, 2016; Krivtsov *et al.*, 2013; Visvader, 2011). Conrad Waddington illustrated the development of the epigenetic profile of a cell during differentiation as "epigenetic landscape" (Figure 2). The illustrative model depicted the cell during development as a ball rolling down a hill into one of several valleys (cell fates) (Waddington, 1957). Although there might be exceptions to this simplified model, it is a good framework depicting the process of epigenetic regulation during differentiation (Ladewig *et al.*, 2013). Mutations can arise and finally drive cancer development at different locations in this landscape. The concept of Waddington's epigenetic landscape in the context of tumorigenesis during B-cell maturation is visualized in Figure 2.

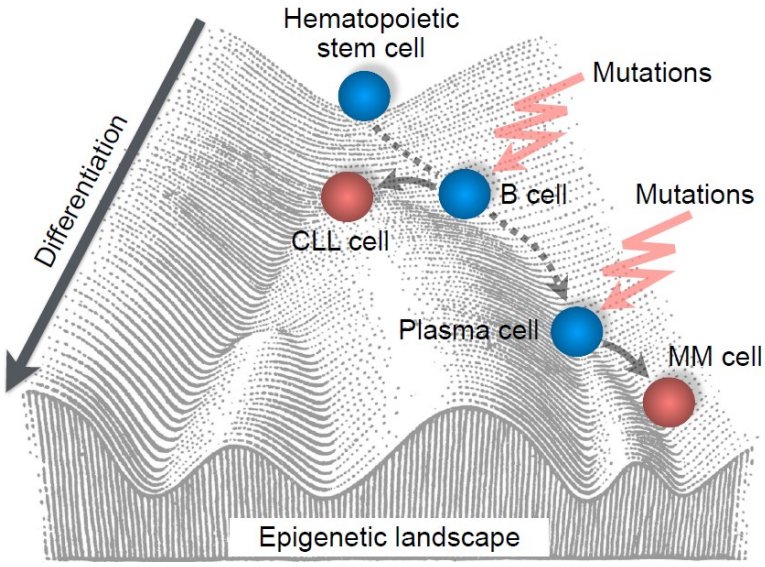


Figure 2 "Epigenetic landscape" model according to Conrad Waddington.
A cell on the normal differentiation path is illustrated as ball (blue) in a landscape. Red arrows indicate genetic events that can eventually lead to cancer development (red). Adapted from "epigenetic landscape" model of Conrad Waddington (Waddington, 1957). CLL: Chronic lymphocytic leukemia, MM: Multiple myeloma.

Genetic events can shape in turn the epigenetic landscape. Mutations can possibly change the fate of a specific cell by driving the development in a certain direction at a branching point or by even dedifferentiating back to a branching point to pursue another developmental path. This has for instance been described to occur by the impact of a mutation in certain cell fate determining transcription factors (Cobaleda *et al.*, 2007). The interplay of genetic and epigenetic alterations in the process of oncogenesis was illustrated already several years ago by Versteege and coworkers (Versteege *et al.*, 1998). They demonstrated that the genetic disturbance of epigenetic regulation here in form of SMARCB1/SNF5 mutations can act as tumorigenesis mechanism in rhabdoid tumors. Mutations in epigenetic regulators

are present in multiple tumor entities (Plass *et al.*, 2013). Epigenetic modulation excerpts a means for tumors to establish stable malignant programs with high impact on the phenotype. Epigenetic silencing of genes (e.g., *BRCA1*) has for instance been observed as alternative means for the inactivation of tumor-suppressor genes mutually exclusive of mutations in the gene (Shen & Laird, 2013).

Several findings support the view that the pathogenesis of CLL has a strong epigenetic contribution. No single driver mutation could be identified in CLL. However, epigenetic regulators are recurrently mutated in CLL patients. This includes factors such as *CHD2, ZMYM3, ASXL1, BAZ2B, IKZF3, HIST1H1B, HIST1H1E, ARID1A, SYNE1, SETD1A, SETD2* and *MLL2* (Landau *et al.*, 2015; Puente *et al.*, 2015). This points to an interplay of genetic and epigenetic factors in CLL. Furthermore, mutations were frequently detected in the non-coding part of the genome, which could be linked to the activity of transcription factors with known relevance in leukemia. These included the recurrent mutation of an enhancer region that could be linked to *PAX5* (Puente *et al.*, 2015). Additionally, an increased activity of histone deacetylases (HDAC) has been reported to correlate with a bad prognosis regarding treatment-free survival and overall survival (Van Damme *et al.*, 2014). Genome-wide epigenetic deregulation was most comprehensively studied by the analysis of DNA methylation in CLL. DNA hypomethylation has been determined to be a general characteristic of CLL (Kulis *et al.*, 2012; Landau *et al.*, 2014; Oakes *et al.*, 2014; Wahlfors *et al.*, 1992). Global DNA-hypermethylation has already been described in 1992 (Wahlfors *et al.*, 1992). DNA hypomethylation in CLL compared to normal B cells was shown to affect most frequently enhancer sites in gene bodies, thus, indicating a functional role of DNA methylation changes (Kulis *et al.*, 2012). The DNA methylation pattern was shown to be linked to the IGHV mutational status (Kulis *et al.*, 2012) and further used to define a signature for three patient subgroups with relevance for prognosis (Queiros *et al.*, 2015).

Multiple myeloma (MM) also displays epigenetic deregulation. Several chromosomal aberrations and mutations in MM are linked to epigenetic regulators. One of the most frequently translocated gene is NSD2 (also known as WHSC1 or MMSET). This gene encodes a histone methyltransferase (Barwick *et al.*, 2019). Increased NDS2 expression levels due to the translocation in a subgroup of MM patients correlated with elevated methylation levels of H3K36 and decreased methylation of H3K27(Martinez-Garcia *et al.*, 2011). Inhibition of the chromatin regulator BRD4 in MM cell lines induced the selective inhibition of known key driver genes in MM including *MYC* (Loven *et al.*, 2013). Several genes with known regulatory function in DNA methylation were shown to be mutated in MM. This includes mutations in *KDM6A*, *IDH1*, *TET2* and *DNMT3A* (Barwick *et al.*, 2019). Furthermore, MM displays local hypermethylation, while globally hypomethylation is prominent. The local hypermethylation often affects enhancers. Hypomethylation seems to increase with progression of the disease from a premalignant state to MM (Agirre *et al.*, 2015). Moreover, HDACs, important epigenetic modifiers, are overexpressed in MM compared to normal plasma cells (Mithraprabhu *et al.*, 2014).

1.2 Pathophysiology of chronic lymphocytic leukemia (CLL)

One of the most frequent leukemias with an incidence of up to 6.4 of 100,000 people per year in Western countries is chronic lymphocytic leukemia (CLL) (Smith *et al.*, 2011). CLL patients show a characteristic and progressive clonal accumulation of $CD19^+/CD5^+$ B-cells particularly in the peripheral blood, lymph nodes, other secondary lymphoid organs and the bone marrow (Zenz *et al.*, 2010). It is a disease of the elderly people with a median age at diagnosis of about 70 years. CLL is more prevalent in males with a male to female ratio of about 1.7 (Smith *et al.*, 2011). The diagnosis of CLL includes a lymphocyte count in the peripheral blood over at least three months of more than 5,000 B lymphocytes/ml. In addition, the diagnosis is determined by blood smears and immunophenotyping. Blood smears are characterized by mature small lymphocytes with little cytoplasm, a dense nucleus without apparent nuclei. Blood smears of CLL patients additionally often show ruptured so called "smudge cells". Beside CD19, CLL cells express the surface antigens CD20, CD23 and CD79b. (Hallek, 2015) and the T-cell marker CD5 (Royston *et al.*, 1980). Many patients (~80 %) are diagnosed at an early asymptomatic stage of the disease (Baliakas *et al.*, 2016). The clinical course of the disease is highly heterogeneous between patients. Some patients have indolent disease lasting over many years, while others exhibit a rather fast disease progression (Stilgenbauer, 2015). Treatment of indolent cases had no beneficial effect on overall survival (Group, 1999). Therefore, a "watch and wait" strategy is routinely applied to early asymptomatic cases (Stilgenbauer, 2015). There are two clinical classification systems to assess the stage of CLL. The Rai (Rai KR, 1975) and Binet (Binet *et al.*, 1981) systems are classically used since decades. They are widely used as they are easy to assess in clinical routine and have proven to be useful to guide treatment initiation. However, both stage systems cannot differentiate indolent and aggressive disease courses for patients diagnosed at an early stage of the disease with low tumor burden (Baliakas *et al.*, 2016). Novel treatment options and the need to decide on treatment onset fostered the identification of factors important for prognosis. Additionally to the age of a patient and its fitness, serological factors like the lymphocyte doubling time, the pattern of bone marrow infiltration, ß2-microglobulin and serum thymidine kinase levels influence prognosis (Stilgenbauer, 2015). Additionally, chromosomal aberrations, gene mutations, the expression of certain genes and especially the somatic hypermutation status of the immunoglobulin heavy-chain variable (IGHV) genes have proven prognostic potential (Baliakas *et al.*, 2016) (Stilgenbauer, 2015). Established chromosomal aberrations for risk stratification are the 17p deletion, 11q deletion, trisomy 12 and 13q deletion alone, which can be detected by fluorescence *in situ* hybridization (Baliakas *et al.*, 2016). Patients with the aforementioned aberrations in the stated order can be assigned to risk groups from high to low risk (Dohner *et al.*, 2000). CLL patients showing a 13q deletion as sole genetic aberrations even seem to have a similar life expectation as healthy individuals of the same age and gender if paired with a favorable IGHV mutation status (Baliakas *et al.*, 2016). Contrary, patients with a 17p deletion or *TP53* mutation are considered as "ultra high risk" (Stilgenbauer, 2015). Apart from TP53 mutations, gene mutations of *BIRC3* (bad prognosis) and *NOTCH1* or *SF3B1*(intermediate prognosis) were identified to

have prognostic potential (Baliakas *et al.*, 2016). Further prognostic markers are the expression levels of *ZAP70*, *CD38* and *CD49d* (Stilgenbauer, 2015). One of the markers with most promising potential and routinely used in clinical trials is the somatic hypermutation status of the immunoglobin heavy-chain variable region (Ghia *et al.*, 2007).

CLL patients can be subdivided into two groups by the presence or absence of somatic hypermutations in the immunoglobin heavy-chain variable region genes (IGHV mutational status) (Fais *et al.*, 1998), which was associated with the aggressiveness of disease (Damle *et al.*, 1999; Hamblin *et al.*, 1999). The IGHV mutational status was shown to have independent and strong prognostic capacity (Oscier *et al.*, 2002). The mutational status is determined by calculating the percentage of base homology to the germline sequence of the locus (Ghia *et al.*, 2007; Zenz *et al.*, 2010). Routinely, a 98 % cut-off is chosen to achieve a binary classification scheme and to account for potential polymorphic variations between individual patient germline sequences and the reference sequence (Ghia *et al.*, 2007). A sequence similarity to the germlines sequence of 98 % or less is thus considered as “IGHV mutated”, while a deviation of less than 2 % is considered as “IGHV unmutated” (Jain *et al.*, 2018). Interestingly, the continuous percentage of homology to germline was found to have prognostic potential (Jain *et al.*, 2018). IGHV mutated CLL cases exhibit a prolonged survival and longer time to first treatment compared to IGHV unmutated cases (Damle *et al.*, 1999).

1.3 Pathophysiology of multiple myeloma (MM)

Multiple myeloma is a plasma cell disorder, which is characterized by the clonal proliferation of these terminally differentiated lymphocytic lineage cells (Rajkumar *et al.*, 2014). Characteristic for MM is the infiltration of the bone marrow with clonal plasma cells exceeding a percentage of 10 % and symptoms of end organ damage (Kumar *et al.*, 2017). End organ damage signs include calcemia as well as renal failure, anemia, and bone lesions also called “CRAB features” (Rajkumar *et al.*, 2014). The clonal plasma cells of most patients furthermore secrete monoclonal immunoglobulin proteins (M proteins). Many clinical symptoms of the disease are associated with the M proteins and the clonal proliferation of cells or their secretion of cytokines (Kumar *et al.*, 2017). The plasma cell disorder belongs to a range of neoplasms termed monoclonal gammopathies (Kumar *et al.*, 2017). MM almost exclusively develops from premalignant stages referred to as monoclonal gammopathy of undetermined significance (MGUS) and smouldering multiple myeloma (SMM) (Kumar *et al.*, 2017; Landgren *et al.*, 2009). The asymptomatic monoclonal gammopathy MGUS is the most frequent (Kumar *et al.*, 2017) and might exist in about 3 % of people above 50 (Kyle *et al.*, 2006). However, not all MGUS patients develop MM, but the risk to progress is about 15 % over 25 years (Kumar *et al.*, 2017). MM is a very heterogeneous disease and is even sometimes described as a range of diseases with common clinical

manifestation (Kumar *et al.*, 2017). Additionally, a high intratumor heterogeneity is observed which is manifested by the development of multiple genetic subclones (Bolli *et al.*, 2014).

1.4 Genetics in lymphoid neoplasms

Although originating from the same lineage, the chromosomal aberrations and mutational landscape differs quite a bit between lymphoid malignancies. CLL for instance displays rather few rearrangements compared to other B-cell malignancies (Puente *et al.*, 2015). While CLL is rather stable, MM patients exhibit a huge variety of complex karyotypes and aberrations. Aneuploidy is a rare event in CLL (Barlogie & Gale, 1992), while it is common in MM (Kumar *et al.*, 2017). Further, the clonal evolution in MM shows a higher complexity compared to CLL. Mutational frequencies also differ. *NRAS* mutations are frequent in MM, whereas they are rarely observed in CLL (Barlogie & Gale, 1992).

1.4.1 Molecular genetics of CLL

The genome of CLL is characterized by a relatively low number of genomic aberrations and gene mutations not only compared to other cancer entities, but also compared to other B-cell malignancies (Gaidano & Rossi, 2017; Lawrence *et al.*, 2013). The average mutational frequency per megabase (Mb) was determined to be 0.87 (Puente *et al.*, 2015). Most CLL patients show only a small number of chromosomal aberrations with only about 35 % cases displaying two or more acquired genomic copy number aberrations (Malek, 2013). About 20 % of CLL patients even have a normal karyotype (Dohner *et al.*, 2000).The most frequent chromosomal aberration is the 13q14 deletion with about 55 %, followed by the 11q22-23 deletion (~18 %), trisomy 12q (~16 %) and 17p13 deletion (~7 %). In two thirds of all 13q deletion cases it was the sole chromosomal aberration (Dohner *et al.*, 2000).

Although no frequently recurring mutations exist in CLL patients, there is a variety of mutations which occur at a lower frequency across CLL patients (Gaidano & Rossi, 2017). The mutational landscape of the coding and non-coding genome has been extensively studied in recent years (Landau *et al.*, 2015; Puente *et al.*, 2015). The most frequently mutated genes with frequencies between 5-10 % according to Puente et al. are *NOTCH1* (~13 %), *ATM* (~10 %), *BIRC3* (~9 %), *SF3B1* (~9 %), *CHD2* (~6 %), *TP53* (~5 %) (Puente *et al.*, 2015). The recurrently mutated genes are not spread over all biological functions, but are enriched in certain pathways and biological functions (Gaidano & Rossi, 2017). Pathways affected by mutations included NOTCH1 signaling (*NOTCH1*, *FBXW7*), NF-κB signaling (*BIRC3*, *TRAF3*, *EGR2*, *NFKBIE*), B-cell receptor and Toll-like receptor pathways (*MYD88*, *TLR2*, *CD79A*, *CD79B*, *IRAK1*), chromatin structure (*ZMYM3*, *HIST1H1E*, *CHD2*, *SETD1*, *MLL2*), DNA damage response, cell cycle control and apoptosis (*TP53*, *ATM*, *SAMHD1*, *POT1*, *BAX*) and RNA metabolism and ribosomal processing (*XPO1*, *SF3B1*, *RPS15*) (Gaidano & Rossi, 2017; Puente *et al.*, 2015). Some

mutations are linked to other features as mutations in the B-cell and Toll-like receptor pathways only occurred in IGHV mutated CLL patients. Mutations in non-coding parts of the genome included the miRNA mir-142, the 3'UTR of *NOCH1* and regions near the 5'-end of *BACH2*, *BTG2*, *BCL6*, *CXCR4* or *TCL1A* (Puente *et al.*, 2015). The absence of frequent diver mutations and the generally low mutation frequency suggests that non-genetic determinants of CLL maintenance and progression exist (Rendeiro *et al.*, 2016).

1.4.2 Molecular genetics of MM

MM is genetically complex and characterized by various chromosomal translocations and aneuploidies. Frequently, translocations involve the immunoglobulin heavy chain (*IGH*) locus on chromosome 14. (Kumar *et al.*, 2017). The highly active *IGH* enhancer induces an aberrant increase in the expression of the corresponding translocation partner (Gonzalez *et al.*, 2007). These translocations recurrently affect the genes *CCND1*, *NSD2* (*WHSC1*/*MMSET*) and *FGFR3* and represent early events during disease development (Kumar *et al.*, 2017; Manier *et al.*, 2017). They are employed to define clinical subtypes (Manier *et al.*, 2017). The most common translocation is t(11;14), which is present in ~14 % of MM patients and affects the expression of *CCND1*. *CCND1* encodes the cell cycle regulator cyclin D1. The translocation t(4;14) occurs in ~11 % of MM patients. The effected genes are *NSD2* and frequently but not always the fibroblast growth factor receptor 3 (*FGFR3)* (Kumar *et al.*, 2017). *NSD2* encodes a histone methyltransferase which thus represents a link between genetic and epigenetic dysregulation (Barwick *et al.*, 2019). As secondary events the translocation of *MYC* is very common in MM with 15-20 % in newly diagnosed MM patients and even higher percentages in specific subgroups (Manier *et al.*, 2017). A deregulated MYC expression was further observed in a high percentage of MM patients (Chng *et al.*, 2011). Further chromosomal aberrations are 1p deletion, 1q gain, 13q deletion and 17p deletion. A single driver mutation is absent in MM, however there are several genes which are recurrently mutated (Kumar *et al.*, 2017). Recurrently mutated genes comprise *KRAS* and *NRAS* with each about 20 % occurrence. Furthermore, mutated genes like *DIS3*, *FAM46C* and *TP53* are found in about 10 % of MM patients each. Mutations are also found but less frequent in *TRASF3*, *BRAF*, *PRDM1*, *RB1*, *CYLD*, *EGR1*, *HIST1H1E* and *MAX* (Walker *et al.*, 2015).

Existing cell line models of MM picture the existing complexity of MM. The cell lines KMS-11 (Namba *et al.*, 1989), MM1.S (Greenstein *et al.*, 2003) and U266 (Nilsson *et al.*, 1970) display differing genetic aberrations and expression patterns important for subgroups of MM. KMS-11 and MM1.S overexpress *MYC* (*c-MYC*), whereas U266 is characterized by the absence of *MYC* expression. U266 cells instead show a high *MYCL* expression. KMS-11 and MM1.S both have a complex *MYC* rearrangement including the juxtaposition to the *IGH* locus which places *MYC* under the control of the *IGH* enhancer. This is combined with a *MYC* duplication and inversion (KMS-11) or duplication (MM1.S) (Dib *et al.*, 2008). KMS-11 cells further harbor a t(4;14)(p16.3;q32) translocation, which leads to the

overexpression of *NDS2* and *FGFR3* (Richelda *et al.*, 1997) and is associated with poor prognosis (Kumar *et al.*, 2017). MM1.S was derived from the MM.1 cell line. It was selected from a heterogeneous population by its sensitivity to dexamethasone. In addition to the MYC overexpression MM1.S cells have a t(14;16)(q32;q23) translocation, thus overexpress MAF. Neither CCND1 nor FGFR3 are dysregulated in MM1.S (Greenstein *et al.*, 2003). The cell line U266 is mostly diploid with several chromosomal gains and losses (Nilsson *et al.*, 1970). U266 overexpresses *CCND1*, due to the insertion of an *IGH* enhancer on chromosome 11 adjacent to *CCND1* (Gabrea *et al.*, 1999).

1.5 Targeted therapy of hematopoietic neoplasms

Currently, a paradigm shift in the therapy of hematopoietic cancers takes place. New therapies with drugs targeting specific molecular pathways with aberrant activity are replacing conventional therapy with cytotoxic agents. These new therapeutic approaches have the potential to indirectly or directly change the chromatin state (Hallek, 2013). It is important to better understand how this new generation of drugs affects and possibly reverts the deregulated epigenetic chromatin state observed in hematopoietic neoplasms.

1.5.1 Indirect chromatin modifying drugs in CLL and MM therapy

Treatment of CLL

For many years alkylating agents like chlorambucil were used in single agent therapy for CLL treatment. Although most patients well tolerate chlorambucil, which made it to a standard treatment choice, complete remissions were rarely achieved. There is range of new regiment which are approved or in clinical development for CLL. They target different pathways with aberrant activity or support the immune system. Antibodies and immunomodulatory drugs are important parts of CLL therapy. Aside from rituximab more antibodies for example against CD20 (ofatumumab, obinutuzumab) have been developed. Anti-apoptotic signaling is targeted by Bcl-2 inhibitors like venetoclax. Another new approach targets the aberrant high B-cell receptor (BCR) signaling observed in CLL. There are several new kinase inhibitors targeting this important survival pathway for CLL cells (Figure 3).

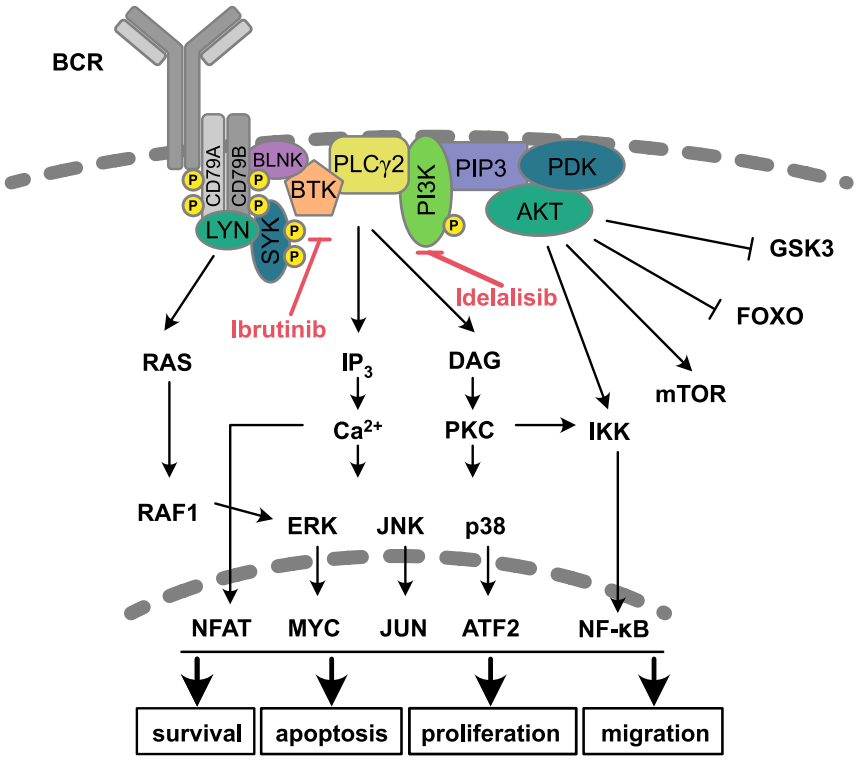


Figure 3 B-cell receptor signaling pathway and inhibitors in CLL. Antigen binding to the B-cell receptor (BCR) induces the phosphorylation of the CD79A/CD79B homodimer by LYN, which enables SYK binding. The following activation of PI3K leads to the conversion of PIP2 to PIP3. BTK, which is recruited to the plasma membrane via PIP3, is phosphorylated and activated. BTK activates its substrate PLCγ2, which leads to the activation of downstream signaling pathways involved in B-cell survival and proliferation (Singh *et al.*, 2018). Figure adapted from Stevenson *et al.* (Stevenson *et al.*, 2011).

Increased activation of NF-κB through BCR signaling promotes proliferation and survival of normal B-cells (Desai *et al.*, 2015). The small molecule inhibitors ibrutinib and idelalisib interfere with BCR signaling by inhibition of the pathway components Bruton's tyrosine kinase (BTK) or Phosphatidylinositol-4,5-bisphosphate 3-kinase (PI3K), respectively (Hallek, 2013). Ibrutinib inhibits BTK specifically and irreversibly by its covalent binding to C481(Burger & O'Brien, 2018). Inhibition of BCR signaling has been proven to be very successful therapy of CLL with normally mild side effects. Ibrutinib is administered as long-term treatment and very effective in the majority of patients irrespective of other risk factors with response rates of over 85 %. Especially combination treatments for instance with venetoclax are currently evaluated (Burger & O'Brien, 2018). However, patients with initial or developing ibrutinib resistance exist. Acquired ibrutinib resistance manifests as Richter's transformation or progressive disease (Pula *et al.*, 2019). Both mutations in components of the BCR signaling pathway like the ibrutinib binding site (C481S) and clonal shifts have been observed in ibrutinib resistant patients (Burger *et al.*, 2016).

Treatment with ibrutinib in most patients induces as direct response in the first weeks an initial increase in leukocyte blood counts (Burger & O'Brien, 2018). This process termed peripheral lymphocytosis was attributed to the involvement of BTK in the regulation of cell-surface receptors important for adhesion and chemokine reception thus affecting the migration and homing of B-cells. The disruption of the homing of B-cells to lymphoid organs leads to the relocation of those cells into the peripheral blood (de Gorter *et al.*, 2007).

Treatment options for MM

MM therapy improved in recent years (Kumar *et al.*, 2017). More categories of drugs with good efficiencies were available also due to an increase in the knowledge about molecular disease mechanisms. This includes proteasome inhibitors like bortezomib in combination with immunomodulatory drugs such as lenalidomide and thalidomide. Autologous haemopoietic stem cell transplantation (ASCT) depends on the physical condition and age of the patient and plays, if eligible, an important role in therapy and treatment decisions (Kumar *et al.*, 2017). Traditional treatment options (alkylating agents, corticosteroid) are further extended by monoclonal antibodies, direct chromatin modifying drugs (Kumar *et al.*, 2017) and T-cell engaging therapies (Mikkilineni & Kochenderfer, 2021). The variety of regiments also increases combination possibilities. Although therapeutic options improved and lead to an elongation and increased degree of response, all patients eventually relapse (Kumar *et al.*, 2017).

1.5.2 Chromatin modifying drugs

The new generation of drugs for blood cancer therapy include also agents with the potential to directly change the chromatin state by targeting epigenetic modifiers. The first approved drug with direct epigenetic activity for specific hematopoietic malignancies was an DNA methyltransferase inhibitor (5-aza-citidine) in 2004, followed by the first histone deacetylase (HDAC) inhibitor (vorinostat) in 2006 (Figure 4). Further HDAC inhibitors were approved including panobinostat for the therapy of multiple myeloma (Cai *et al.*, 2015). HDACs were shown to be overexpressed in MM cell lines and primary patient cells and high HDAC1 levels in patients were linked to poor prognosis (Mithraprabhu *et al.*, 2014). Panobinostat shows only modest activity supplied as monotherapy. Therefore, panobinostat is usually applied as combination therapy often with the proteasome inhibitor bortezomib (Tandon *et al.*, 2016). Their synergistic effect is mainly attribute to the simultaneous inhibition of the proteasome (bortezomib) and aggresome (panobinostat) pathways (Hideshima *et al.*, 2011). Panobinostat is a pan-HDAC inhibitor, which inhibits HDAC classes I, II, IV efficiently at low concentrations compared to other non-selective HDACs. Panobinostat exhibits half maximal inhibition at concentration in the low nanomolar range of maximal 13.2 nM for most HDACs except HDAC4, 7, 8 (Atadja, 2009). Despite their name histone deacetylases catalyzed the deacetylation of histones but also non-histone proteins

(Choudhary *et al.*, 2009). HDACs are the antagonists of histone acetyltransferases (HATs). Histone acetylation plays a critical role in gene regulation by controlling the accessibility of DNA and the activity of regulatory elements (Kouzarides, 2007).

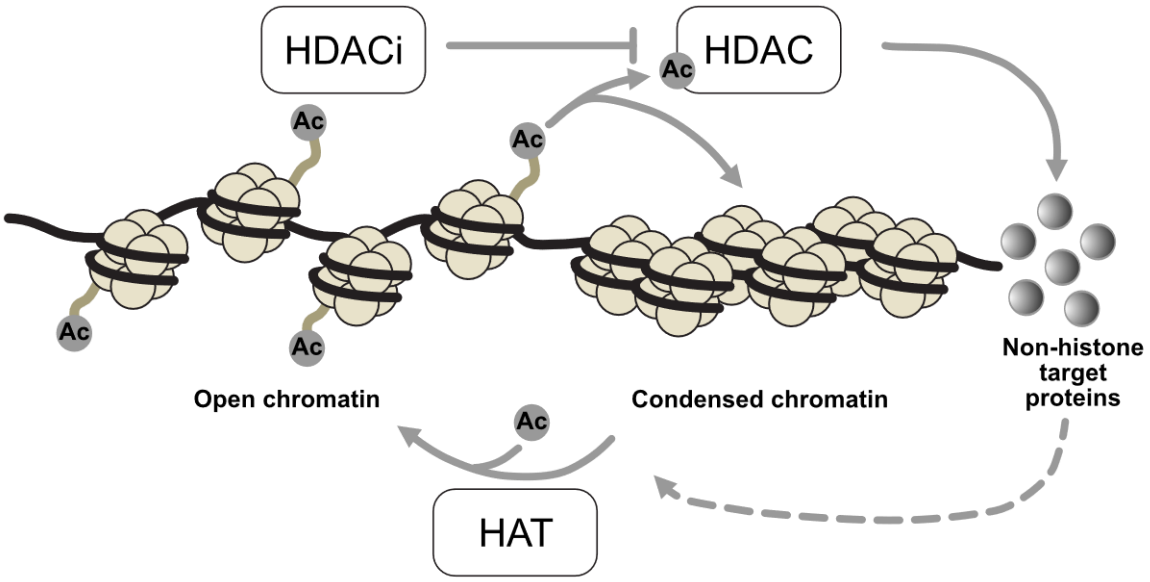


Figure 4 Mode of action of histone deacetylase inhibitors.
Histone deacetylase inhibitors inhibit histone deacetylases, which catalyze the deacetylation of histone and non-histone proteins. Histone acetyltransferases catalyze the invert reaction. Activity changes of non-histone proteins by deacetylation can have various indirect effects on chromatin (dotted arrow) (Ververis *et al.*, 2013). HDACi: Histone deacetylase inhibitor; HDAC: Histone deacetylase; HAT: Histone acetyltransferase.

2 Epigenetic regulation of transcription

2.1 Chromatin as storage for information

2.1.1 Chromatin organization

In eukaryotic cells the DNA is packaged in the nucleus in a complex of nucleic acids and proteins called chromatin. The main protein components of the chromatin complex are the core histone proteins H2A, H2B, H3 and H4, which enable to store the DNA in a compacted form in the nucleus (Figure 5). The core histones form an octamer consisting of two of each of these four histones. The DNA helix is left-handed wrapped around these histone octamers in ~1.7 turns comprising ~147 bp DNA which forms a nucleosome. Nucleosomes are connected via the intervening linker DNA of typically around 40-50 base pairs in human cells. (Luger *et al.*, 1997). The nucleosome chain is further structured on different levels and in higher order (Gorkin *et al.*, 2014).

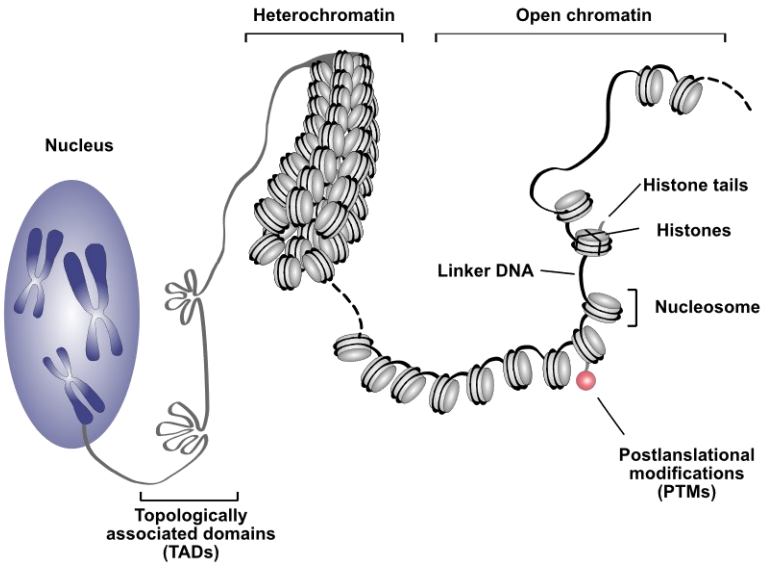


Figure 5 Chromatin organization on different layers.
Illustration shows DNA compaction and accessibility in the nucleus of eukaryotes at several levels of compaction, which is influenced by post-translational modifications of histone proteins. Figure adapted from (Baylin & Jones, 2011).

Individual parts of the nucleosome chain of a chromosome can be arranged at different degrees of compaction. The chromatin can exist in a very condensed or closed conformation, which is associated with HP1 binding (Maison & Almouzni, 2004). Simultaneously, other parts of the chromatin fiber can have a loose and open conformation. The chromatin further forms regulatory domains of interacting chromatin on the scale of about 1 mega base pair called topologically associated domains (TADs) (Dixon *et al.*, 2012). Each hierarchical layer can affect and regulate the accessibility of the genetic information. The position of nucleosomes is not static along the DNA sequence. Nucleosome positioning effects the accessibility of regulatory elements in the DNA to transcription factors and the transcriptional machinery. Nucleosome-free regions are more accessible for certain factors needed for transcription, than the DNA around nucleosomes (Bai & Morozov, 2010). Similarly, a condensed and repressed chromatin state prevents the binding of factors and the transcription of genes (Spitz & Furlong, 2012). The condensation state of chromatin is strongly impacted by post-translational modifications (PTMs) of the histone proteins. PTMs can potentially influence the compaction of chromatin in two ways. PTMs can change the charge of the nucleosome tails and thus their electrostatic interaction with the DNA and other histone proteins. At the same time, they can act as binding platform for the recruitment of non-histone proteins. Those can have enzymatic activity including chromatin remodeling abilities (Kouzarides, 2007). Aside from DNA methylation, the post transcriptional modification of histones presents a very versatile mechanism to store epigenetic information (Jenuwein & Allis, 2001).

2.1.2 Histone modifications

Histones have a globular conformation, while their N-terminus ("histone tail") is mostly unstructured. A variety of amino acid residues in the histone tails can be chemically modified. Those covalent post transcriptional modifications of histone variants comprise acetylation, methylation or phosphorylation of histone tails at different residues. (Kouzarides, 2007). The combinatorial nature of histone modifications adds another layer on the genetic code, which was also termed the "histone code" (Jenuwein & Allis, 2001; Strahl & Allis, 2000). A set of specified enzymes interacts with histones and adds, removes or reads their covalent modifications, which influences chromatin structure and gene expression. They are also called chromatin "writers", "erasers" and "readers" (Brien *et al.*, 2016; Gillette & Hill, 2015).

Technological advances enabled to map the genomic location of post translationally modified histones genome-wide. Chromatin immunoprecipitation followed by sequencing (ChIP-seq) (Barski *et al.*, 2007; Park, 2009) revealed the genome-wide distribution and functional implications of huge variety of PTMs (Bernstein *et al.*, 2010). Certain histone modifications could be associated with the activity status of regulatory elements and gene expression. Enhancers and promoters are both methylated at the histone H3 lysine residue 4 (H3K4) but can be distinguished by the number of methyl groups. While promoters are associated with trimethylation (H3K4me3), enhancers are characterized by monomethylation of

H3K4 (H3K4me1) (Heintzman *et al.*, 2007) (Figure 6). Active enhancers and promoters are characterized by the presence of H3K27ac additionally to H3K4 methylation (Heintzman *et al.*, 2009; Rada-Iglesias *et al.*, 2011). The acetylation of H3K27 at enhancers distinguishes active from primed enhancers, which are only marked by H3K4me1 (Creyghton *et al.*, 2010). In addition, a third class of enhancer states was detected, which is marked by the presence of a repressing histone modification. Those repressed enhancers harbor nucleosomes with H3K4me1 as well as H3K27me3 and are associated with inactive genes (Rada-Iglesias *et al.*, 2011). Besides, gene bodies of active genes are associated with H3K36me3 (Roadmap Epigenomics *et al.*, 2015; Wagner & Carpenter, 2012).

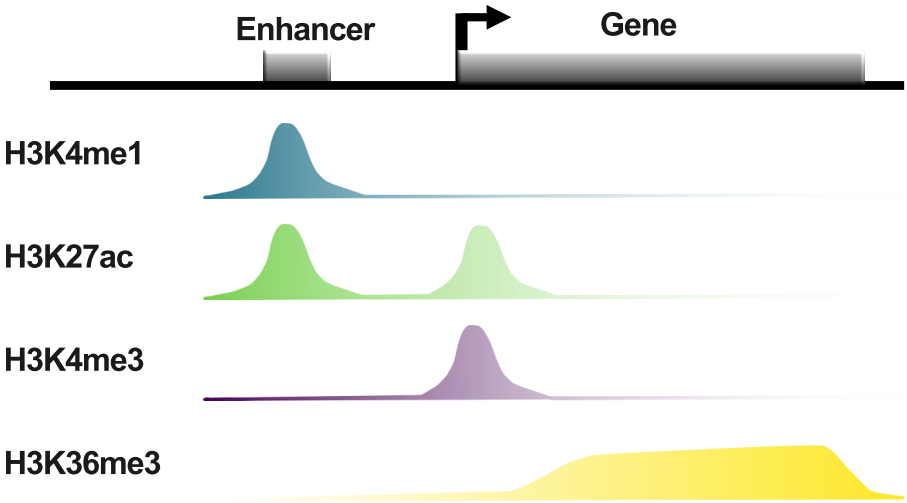


Figure 6 Scheme of genomic location of post-translational histone modifications associated with active enhancers and promoters.
Simplified illustration of the distribution of selected histone modifications at promoters and enhancers as detectable by ChIP-seq. Active enhancers are characterized by H3K27ac additionally to H3K4me1, which is also found at primed enhancers. Promoters of active genes also show H3K27ac in addition to H3K4me3. The gene body of transcribed genes is enriched for H3K36me3.

2.2 Enhancers

Complex gene expression pattern relevant for a certain developmental stage are established by the coordinated action of regulatory element, which are subjected to epigenetic regulation. Cis-regulatory elements include promoters, enhancers, insulators and silencers (Schoenfelder & Fraser, 2019). Enhancers are supposed to play a fundamental role in this process. Enhancers are important cis-regulatory elements, which can be bound by transcription factors and are thought to enhance the transcription of a target gene by the interaction with its promoter (Gasperini *et al.*, 2020) (Figure 7). The first description of such an DNA element, that enhances transcription of a gene in *cis* independent of its orientation and at different positions dates back to the year 1981 (Banerji *et al.*, 1981; Moreau *et al.*, 1981). Since then, several hundreds of thousands putative enhancer regions were identified in various

mammalian cell types (Encode Project Consortium, 2012; Schoenfelder & Fraser, 2019). Enhancers were found to be located in intergenic regions up to ~1,000 kb upstream or downstream of their target gene. Others are positioned in introns of their target gene or unrelated genes (Schoenfelder & Fraser, 2019). The mechanism by which an enhancer confers its activity on a gene is still not fully resolved and many models have been proposed (Furlong & Levine, 2018). It is now well established that physical interaction between enhancers and promoters plays an important role in this process (Schoenfelder & Fraser, 2019). Enhancers and promoters were shown to be able to come into physical contact (Carter *et al.*, 2002).

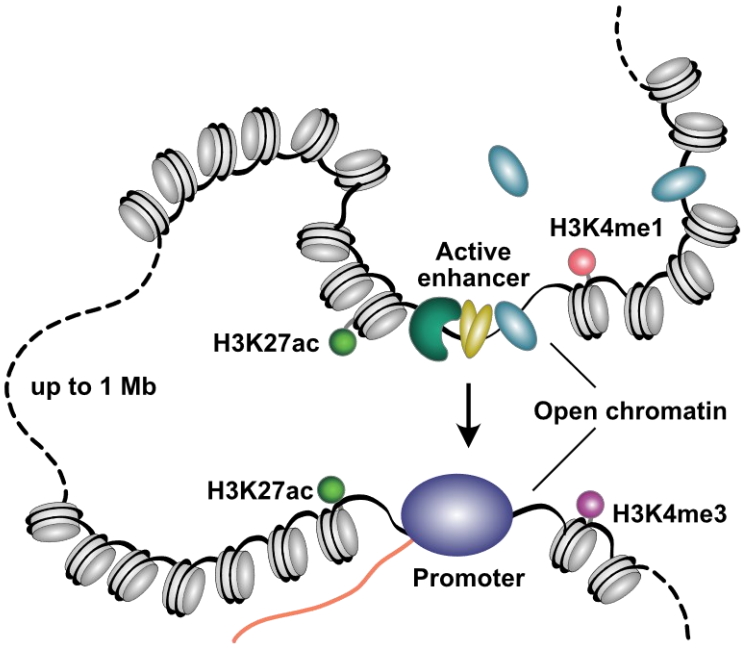


Figure 7 Scheme of enhancer and promoter interaction. Active enhancers and promoters are marked by accessible chromatin and flanked by the indicated histone modifications.

The proximity of enhancers and promoter seems to accompany transcriptional activity (Furlong & Levine, 2018; Schoenfelder & Fraser, 2019). At the ß-globin promoter it was firstly demonstrated that an artificially induced contact of an enhancer and a promoter via looping can activate gene expression (Deng *et al.*, 2012). The prevailing model currently is, that enhancer and promoters come into close proximity by forming a loop of the intervening DNA ("looping model") by the interaction of factors bound at the DNA elements and recruited cofactors (Furlong & Levine, 2018). Enhancer can further control different genes in different contexts. There is evidence that enhancer promoter interactions are preferentially established within TAD boundaries (Furlong & Levine, 2018; Schoenfelder & Fraser, 2019). A recent example shows the possibility that promoters can compete for enhancers. The promoters of *PVT1* and *MYC* are in competition for the contact to intronic enhancers within *PVT1*. Thus, the PVT1 promoter controls *MYC* expression. The disruption of the *PVT1* enhancer interaction for instance by a mutation can lead to an aberrantly high expression of *MYC* which is often observed in cancer (Cho *et al.*, 2018). Furthermore, a variety of features could be associated with enhancers and their activity.

Enhancers harbor many transcription factors binding sites and show an enrichment of cofactors such as p300 and Mediator. Active enhancers are depleted of nucleosomes. They are however flanked by nucleosomes with specific PTMs especially H3K4me1 and H3K27ac (Gasperini *et al.*, 2020; Shlyueva *et al.*, 2014) (Figure 7). More recently bidirectional transcription at enhancers was associated with active enhancer regions (Andersson *et al.*, 2014; Kim *et al.*, 2010). Further, the chromatin accessibility associated with active enhancers facilitates the binding of transcription factors, which is pivotal to their activity (Gasperini *et al.*, 2020).

2.3 Assay for transposase-accessible chromatin using sequencing (ATAC-seq)

There are several methods to map accessible chromatin. Apart from formaldehyde-assisted isolation of regulatory elements sequencing (FAIRE-seq), DNase I hypersensitive sites sequencing (DNase-seq) and micrococcal nuclease sequencing (MNase-seq; maps nucleosome positioning), the method repertoire was more recently extended by the assay for transposase-accessible chromatin using sequencing (ATAC-seq) (Buenrostro *et al.*, 2013; Giresi *et al.*, 2007; Ponts *et al.*, 2010; Song & Crawford, 2010; Tsompana & Buck, 2014). ATAC-seq is a method to measure chromatin accessibility which is suitable for small amounts of cells. It was developed in 2013 by Buenrostro et al. (Buenrostro *et al.*, 2013). ATAC-seq gained fast popularity as readout for accessible chromatin, which is shown by the increase of published ATAC-seq datasets and manuscripts (Yan *et al.*, 2020). Its popularity has risen for several reasons. Besides the simple and fast protocol, it is very versatile. Alongside open chromatin sites, it also allows for instance the identification of nucleosome positions near regions with higher accessibility. The sensitivity and specificity of ATAC-seq were determined to be similar to DNase-seq (Buenrostro *et al.*, 2013), but ATAC-seq needs far smaller amounts of starting cell numbers (Tsompana & Buck, 2014). Thus, it is perfectly suited for primary human patient samples, where cell material is inherently limited. Additionally, recent advances even allow to assess chromatin accessibility on the single cell level (Buenrostro *et al.*, 2015b; Cusanovich *et al.*, 2015).

ATAC-seq utilizes the capabilities of a hyperactive form of the Tn5 transposase (Buenrostro *et al.*, 2013). Endogenously, each homodimer of the wildtype transposase binds the transposon DNA at one of its flanking IS50 elements, which are arranged in opposite orientation, to integrate it into a stretch of DNA (Reznikoff, 2003). A hyperactive version of the 19 bp outside ends of the IS50 elements, which is also called mosaic end, is sufficient for binding and transposition (Adey *et al.*, 2010; Reznikoff, 2003). Instead of transposon DNA, the transposase dimer can be loaded *in vitro* with two separate adaptors (non-identical) for next generation sequencing, which are not connected by a stretch of DNA. To do so both adaptor sequences contain the 19 bp mosaic end sequence. The hyperactive Tn5 transposome

complex is able to integrate its sequencing adaptor cargo into DNA and thereby also fragments it (Adey *et al.*, 2010) (Figure 8). The fragmentation originates from the missing DNA connection between the two adaptors. The two transposase monomers cut the two complementary target DNA strands with a 9 bp offset to attach the adaptors to the 5'ends. This leads to a doubling of this 9 bp sequence in the final sequencing library (Reznikoff, 2008) which mark the core binding base pairs of the transposition event (Adey *et al.*, 2010). Due to steric hindrance in less accessible chromatin regions, the transposase tends to integrate the adaptors into DNA at open chromatin sites. Thus, generated sequencing reads preferentially originate from and mark accessible chromatin (Buenrostro *et al.*, 2013). Although its numerous benefits, ATAC-seq also poses some challenges. Even till now, there were only a few bioinformatic tools specifically developed for ATAC-seq (Yan *et al.*, 2020). Thus, tools developed for other sequencing readouts need to be adjusted. Besides, small contaminations of mitochondrial DNA may lead to a vast fraction of sequencing reads originating from the mitochondrial genome (Buenrostro *et al.*, 2015a), which is per se rather accessible (Bogenhagen, 2012). The mitochondrial reads contain no chromatin accessibility information and thus reduce effective sequencing depth for the analysis of open chromatin (Buenrostro *et al.*, 2013; Yan *et al.*, 2020). This issue was addressed by improved protocols, which reduce the fraction of mitochondrial reads drastically (Corces *et al.*, 2017).

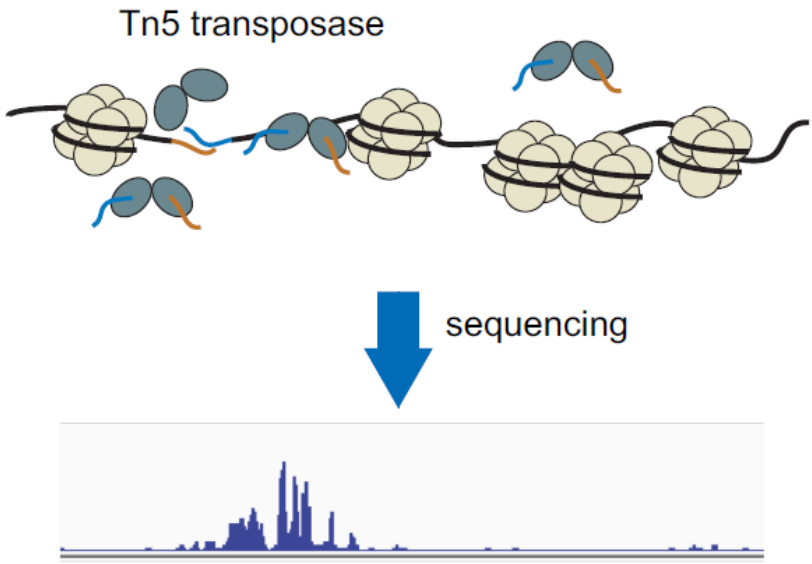


Figure 8 Scheme for identification of open chromatin region with ATAC-seq and the resulting sequencing read distribution.
The Tn5 transposase (grey) acts as dimer. Each enzyme of the dimer is loaded with a sequencing adaptor (blue, orange). The transposase complex integrates those adaptors into DNA in accessible chromatin regions (left). This leads to a fragmentation of the genome. If proximal transposition events occur amplifiable fragments for next generation sequencing are generated (bottom). Steric hindrance prevents the transposition in regions of condensed chromatin (right). Footprints in the signal can mark the binding of factors, which protected the chromatin (Buenrostro *et al.*, 2013).

ATAC-seq can be considered a condensed readout for the chromatin state, which gives information on multiple levels of gene regulation. Hence, it also allows to perform a genome-wide analysis of the

chromatin in cases where a comprehensive epigenomic analysis using various chromatin readouts is not possible due to sample or money limitations (Buenrostro *et al.*, 2013; Buenrostro *et al.*, 2015a). ATAC-seq allows to simultaneously map active regulatory elements like promoters and enhancers by their accessibility including differentially accessible regions between conditions (Smith & Sheffield, 2020) and nucleosome positioning (Buenrostro *et al.*, 2013). Nucleosome positions can be inferred by separating sequencing reads spanning one or more nucleosomes from reads of subnucleosomal length, which are all contained in ATAC-seq libraries. Besides, ATAC-seq enables to identify transcription factor binding, which is visible as a "footprint" in the accessibility signal (footprinting analysis) due to the protection of DNA by this factor. The identity of the transcription factor can be inferred by the presence of an annotated binding site with known motif (Buenrostro *et al.*, 2013). The view hundred base pair long accessible regions can contain several transcription factor binding sites, which comprise typically less than 20 bp (Heinz *et al.*, 2010). Thus, the footprinting analysis is important to identify transcription factors linked to accessible regions. The analysis of enriched transcription factor motifs in differentially accessible regions between different conditions can give further valuable information on the change in the activity of specific transcription factors (Smith & Sheffield, 2020). Finally, the normally discarded reads which map to the mitochondrial genome can be utilized for lineage tracing when performing ATAC-seq on the single cell level (Ludwig *et al.*, 2019).

2.4 Transcription factors

Epigenetic modifications like histone methylation or acetylation alone are not sufficient to induce gene expression. Their effect on gene expression is mediated by the integrated action of transcription factors (TF). Thus, gene expression is regulated by an interplay of epigenetic patterns, chromatin modifiers and transcription factor binding (Lee & Young, 2013). The basic concept of transcriptional regulation in the absence of a chromatin contribution goes back to work conducted in bacteria in 1961 (Jacob & Monod, 1961): TFs are proteins which bind DNA at regulatory elements and promote the assembly of the transcription machinery (Lee & Young, 2013; Spitz & Furlong, 2012). TFs recognize and bind short specific DNA sequences in regulatory elements, which are called transcription factor binding motifs. Those TF binding motifs are in eukaryotes typically only about 6-12 bp long (Spitz & Furlong, 2012; Wunderlich & Mirny, 2009). TFs contain DNA binding domains, which mediate the sequence-specific DNA binding. TFs are usually classified into families by their binding domains via sequence similarity (Weirauch & Hughes, 2011).

The same TF can regulate several genes simultaneously in a cell type and chromatin context dependent manner (Muckenhuber *et al.*; Spitz & Furlong, 2012). The relatively short binding motif of a certain TF is often present at numerous locations in the eukaryotic genome but not all sites are actually bound by the TF (Lambert *et al.*, 2018). Therefore, the binding motif sequence seems not to be sufficient to explain

TF binding in eukaryotes (Wunderlich & Mirny, 2009). Besides, regulatory elements often contain binding sites for different TFs (Spitz & Furlong, 2012). Cooperate binding of several TFs is often observed (Lambert *et al.*, 2018; Spitz & Furlong, 2012) and typical for eukaryotes (Wunderlich & Mirny, 2009). One example of cooperative binding is the synergy between the unrelated transcription factors NFAT and AP-1 (Chen *et al.*, 1998; Macian *et al.*, 2001). As other proteins, TFs can be post-translationally modified for example by acetylation or phosphorylation. Post-translational modifications can impact on the activity of the respective TF for instance by effecting DNA-binding, protein stability or localization (Filtz *et al.*, 2014). In this process, TF acetylation might act as both activity enhancing and activity repressing factor. While acetylation of p53 leads to an enhancement of DNA-binding, the acetylation of other transcription factors disrupts DNA binding or interferes with protein-protein interactions (Kouzarides, 2000). ChIP-seq experiments allow to determine the binding sites of a single transcription factor genome-wide (Park, 2009).

3 Disease-specific epigenetic changes in the context of the tumor cell-of-origin

3.1 Tumor cell-of-origin and heterogeneity

The tumor cell-of-origin is the normal cell from which tumorigenesis initiated. The tumor-initiating cell obtained the first oncogenic event. Further cancer-promoting events accumulate and lead to the progression of malignant transformation. Physiological development of an organ system is a hierarchical advancement from stem cells over common and committed progenitor cells to terminally differentiated cells. In principle, each cell can transform, if it possesses the ability to proliferate and acquires the necessary self-renewing and differentiation blockage capacities. The identification of the cell-of-origin and its molecular phenotype might lead to a better understanding of processes occurring during malignant transformation (Visvader, 2011). Tumorigenesis occurring from different stages of the development of an organ could also lead to the origin of intertumor heterogeneity and the classification of tumor subtypes. Differences in the transforming event which hit cells in a defined developmental stage can give rise to heterogeneity between patients. Aside from this, another explanation for heterogeneity are transforming events which hit cells in different developmental stages (Visvader, 2011). New single cell technologies further allow to assess apart from interpatient heterogeneity also existing intrapatient heterogeneity. The specification of tumor subtypes has been mostly conducted on the basis of histopathology and surface markers. However, several examples show that those markers are not necessary sufficient to identify the cell-of-origin (Visvader, 2011). This emphasizes the importance of molecular markers of cellular origin. The exact origin of CLL also could not be conclusively resolved on the basis of surface markers and is still a focus of research (Seifert *et al.*, 2012).

3.2 CLL originates from a continuous B-cell differentiation trajectory

CLL patients can be subdivided into two clinically relevant groups according to the presence of IGHV rearrangements a process included in normal B-cell maturation (Fais *et al.*, 1998). The binary classification has prognostic value and was associated with the aggressiveness of disease of individual CLL patients (Damle *et al.*, 1999; Hamblin *et al.*, 1999). Genome-wide epigenetic studies allowed to

further delineate the cellular origin of CLL cells for individual patients. Kulis and coworkers showed that the two clinical subgroups can be further separated in three subtypes by DNA methylation analysis. The combined analysis of DNA methylation changes during physiological B-cell development of distinct potential CLL precursors and CLL patients lead to further insights. The progressive DNA methylation changes occurring during normal development are reflected in CLL patients. Two epigenetically defined subgroups were largely overlapping with groups defined by the IGHV status but with an additional intermediate group (Kulis *et al.*, 2012). Queiros could further develop a classifier to discriminate the three subgroups (Queiros *et al.*, 2015). The intermediate epigenetic phenotype of some patients was confirmed by Oakes et al. in 2016 (Oakes *et al.*, 2016). Moreover, CLL does not originate from B-cells in three distinct maturation stages, but can initiate from a continuum of B-cell maturation stages showing progressive physiological DNA methylation changes (Kulis *et al.*, 2015; Oakes *et al.*, 2016). This has valuable prognostic implications as the developmental stage of the cell from which tumorigenesis initiated influences the aggressiveness of the disease of individual CLL patients (Oakes *et al.*, 2016; Queiros *et al.*, 2015).

3.3 The analysis of disease-specific epigenetic changes in CLL

In CLL the epigenetic patterns are disrupted. Genome-wide epigenetic dysregulation has been observed for DNA methylation signatures (Kulis *et al.*, 2012; Landau *et al.*, 2014; Oakes *et al.*, 2014; Wahlfors *et al.*, 1992). A tumor cell harbors molecular patterns of its cell-of-origin and additionally acquires disease specific patterns (Oakes *et al.*, 2016; Visvader, 2011; Wierzbinska *et al.*, 2020). Therefore, tumor-specific epigenetic patterns need to be evaluated in the context of epigenetic changes occurring during physiological B-cell development. The selection of an adequate healthy reference for the analysis of the epigenome in CLL is thus important but inherently difficult. Comparisons of malignant epigenetic patterns are most often carried out in comparison to carefully selected healthy reference subpopulations. CLL-induced epigenomic changes are classically identified in relation to $CD19^+$ B cells of healthy donors (Landau *et al.*, 2014; Oakes *et al.*, 2014; Pei *et al.*, 2012). However, $CD19^+$ B-cells from peripheral blood are a mixture of cells including naïve and memory B-cells (Klein *et al.*, 1998). Thus, one has to be aware that the epigenetic profile obtained from those cells is also an average. Another conservative approach defined disease-induced chromatin accessibility changes as difference to the epigenome present in any of a range of distinct B-cell subpopulations (Beekman *et al.*, 2018). *In silico* modeling of the cell-of-origin for individual CLL patients from the DNA methylome and the identification of CLL-specific methylation changes has recently been successfully conducted (Wierzbinska *et al.*, 2020). However, if it is possible to use a similar approach to evaluate chromatin accessibility changes or other chromatin features linked to enhancer activity in CLL in relation to the cell-of-origin is still an open question.

3.4 Chromatin accessibility as differentiation stage marker

Several recent studies established that chromatin accessibility can serve as marker for the differentiation stage and indicate that it can be used to identify the cell-of-origin of tumor cells. Corces and coworkers captured the whole spectrum of hematopoiesis with 13 distinct subpopulations. In their analysis chromatin accessibility measured by ATAC-seq well discriminated the cell types of the hematopoietic lineages and showed better results than the classification by RNA-seq data. Further, the chromatin accessibility at distal regulatory elements like enhancers was more informative in terms of cell identity than at promoters (Corces *et al.*, 2016). The chromatin accessibility evolution during hematopoiesis was further analyzed by Satpathy and coworkers in 2019. They determined chromatin accessibility changes over the continuous differentiation trajectory of the complete hematopoiesis. This was conducted by reconstructing the differentiation trajectory from ATAC-seq data of single cells. They further used their healthy development reference to compare it to dysregulations in basal cell carcinoma (Satpathy *et al.*, 2019). In a proof of principal study, Georg et al. demonstrated that transformed cells can retain chromatin accessibility patterns of their cell-of-origin. For this, they made use of the fact that the expression of MLL-AF9, an important fusion oncogene in AML, can provide self-renewal capacity to both hematopoietic stem cells and committed progenitors. The transformed cells with different cell-s-of-origin could be distinguished by their chromatin accessibility profile and maintained patterns of their cellular origin (George *et al.*, 2016). Furthermore, ATAC-seq was used in CLL to define epigenetic signatures of IGHV maturation subgroups in CLL (Rendeiro *et al.*, 2016). Beekman et al. found that chromatin accessibility profiles of CLL patients divided into IGHV unmutated and mutated subgroups each show a higher resemblance to their expected normal counterpart. That is to say, unmutated CLL are more similar to naïve B-cells and mutated CLL resemble germinal center B-cells, mature B-cells and plasma cells in their accessibility profile. This indicates that CLL cells retain an imprint of the chromatin accessibility of their cell-of-origin that can be used to determine its maturation stage (Beekman *et al.*, 2018). Most recently, chromatin accessibility signatures were employed to determine the developmental origin of cells of individual T-ALL patients. First, accessibility signature regions were identified to define seven distinct healthy populations of T-cell development. Then, the resemblance of chromatin accessibility patterns of T-ALL patients of different maturation subgroups were compared to the T-cell subsets based on ~3,000 accessibility signature regions by projecting them into the same PCA space. They found that even the most mature T-ALL samples still rather resemble the immature healthy T-cell subtypes of their analyzed spectrum (Erarslan-Uysal *et al.*, 2020).

4 Scope of the thesis

Dynamic epigenetic regulation of gene expression during hematopoiesis is an essential part of the physiological development and differentiation into distinct cell types (Smith & Meissner, 2013). This process is disrupted during malignant transformation as shown for DNA methylation signatures in chronic lymphocytic leukemia (CLL) (Kulis *et al.*, 2015; Oakes *et al.*, 2016). The deregulated epigenetic signaling during tumorigenesis is crucial for the establishment and maintenance of tumor specific gene expression patterns and affects treatment response (Brookes & Shi, 2014). It has been observed that tumor cells harbor epigenetic imprints of their cellular origin and further gain disease-specific patterns during neoplastic progression (Oakes *et al.*, 2016; Visvader, 2011; Wierzbinska *et al.*, 2020). As tumorigenesis can occur at various stages of hematopoiesis, it is crucial to distinguish between developmental and disease-induced changes. The selection of an adequate healthy reference is thus key to the analysis of epigenetic deregulation. Activity changes of enhancers is likely to be a major factor for the deregulation of gene expression and is reflected in their chromatin signature (Roadmap Epigenomics *et al.*, 2015). The assay for transposase-accessible chromatin using sequencing (ATAC-seq) is a condensed readout which provides valuable functional information about the chromatin state of cis-regulatory elements like enhancers and links them to transcription factor occupancy (Buenrostro *et al.*, 2013). ATAC-seq has been used to study the regulatory landscape of normal hematopoiesis (Corces *et al.*, 2016; Satpathy *et al.*, 2019) and human primary cancers including pan-cancer studies and single entities (Beekman *et al.*, 2018; Corces *et al.*, 2018; Erarslan-Uysal *et al.*, 2020; George *et al.*, 2016; Rendeiro *et al.*, 2016).

The analysis of chromatin accessibility by ATAC-seq requires appropriate data analysis procedures (Smith & Sheffield, 2020; Yan *et al.*, 2020). Therefore, the **first aim** of my thesis was to **establish the experimental and bioinformatic procedures** for ATAC-seq needed for the identification of CLL specific chromatin accessibility changes to bring drug induced epigenetic modulation into context.

The **second aim** of the thesis was to **evaluate if chromatin accessibility patterns can be used for phylogeny reconstruction of potential tumor precursor cells and hence to determine the maturation stage of the tumor cell-of-origin and its chromatin accessibility state**. To do so, I constructed a phylogenetic tree from ATAC-seq data from six distinct B-cell subsets, which resembled the phylogeny deduced from DNA methylation. The tree was used to infer a maturation score for individual B-cell subsets and model the continuous physiological accessibility dynamics of individual accessible regions. Employing a principal component regression approach, I could thus identify the maturation stage of the cell-of-origin for individual CLL patients.

The **third aim** of this thesis was to **identify disease-induced accessibility patterns and link them to transcription factor occupancy**. For this, I first employed the classical approach using $CD19^+$ B-cells, which revealed enhancer deregulation linked to aberrant transcription factor activity. The predicted chromatin accessibility pattern of the cell-of-origin of a certain CLL patient further allowed to distinguish between CLL-specific accessibility changes and physiological epigenetic programming.

The **fourth aim** of the thesis was to **assess the effect of novel targeted drugs for the therapy of hematopoietic neoplasms on this deregulated epigenome**. To this end, I studied the effect of ibrutinib on the regions with CLL-specific accessibility changes. To further evaluate the potential modulation of aberrant epigenetic patterns by drugs, the effects of a histone deacetylase inhibitor on chromatin were studied in multiple myeloma cells.

By addressing these four aims, my thesis work contributes to a better understanding of the tumorigenesis process in CLL and possibly other hematopoietic cancers by a systematic and functional analysis approach of disease-specific epigenetic deregulation. The reversible nature of epigenetic changes opens possibilities for intervention. A better understanding of epigenetic effects of blood cancer therapy, which includes agents with the potential to directly change the chromatin state by targeting epigenetic modifiers, will eventually improve the therapy of hematopoietic cancers.

Materials and Methods

1 Materials

1.1 Custom buffers

Table 1. Buffer composition.

Buffer	Composition
ATAC-seq lysis buffer	10 mM Tris-HCl (pH 7.4), 10 mM NaCl, 3 mM $MgCl_2$, 0.1 % IGEPAL CA-630 (Buenrostro *et al.*, 2013)
Elution buffer	50 mM Tris pH 8.0, 1 mM EDTA, 1 % SDS, 50 mM $NaHCO_3$
Freezing medium	90 % FBS, 10 % DMSO
Gitschier buffer (1x)	67 mM Tris-HCl pH 8.8, 16.6 mM Ammonium sulfate, 6.5 mM $MgCl_2$, 0.5 % Triton, 0.02 % SDS, 1 % ß-mercaptoethanol, 4 µl/ml Proteinase K (stock: 20mg/µl) (stored at 4 °C; ß-mercaptoethanol and Proteinase K are added shortly before use)
High salt buffer	50 mM Hepes pH 7.9, 500 mM NaCl, 1 mM EDTA, 1 % Triton X-100, 0.1 % sodium deoxycholate, 0.1 % SDS
Lithium buffer	20 mM Tris pH 8.0, 1 mM EDTA, 250 mM LiCl, 0.5 % IGEPAL CA-6301, 0.5 % sodium deoxycholate
MNase buffer	25 mM KCl, 4 mM $MgCl_2$, 1 mM $CaCl_2$, 50 mM Tris HCl pH 7.4
20× PBS	45 mM Na_2HPO_4; 181 mM NaH_2PO_4; 3 M NaCl (pH 7.4)
1× RIPA buffer	150 mM NaCl, 1 % NP40, 0.5 % sodium deoxycholate, 0.1 % SDS, 50 mM Tris pH 8.0, 1× Halt Protease Inhibitor Cocktail
1× TBS	25 mM Tris, 0.15 M NaCl, pH 7.5
1× TBS-T	25 mM Tris, 0.15 M NaCl, pH 7.5, 0,1 % Tween20
TE buffer	10 mM Tris pH 8.0, 1 mM EDTA
Tn5 wash buffer	20 mM Tris pH 8, 1 mM EDTA, 0.1 % SDS
1× transfer buffer	25 mM Tris, 192 mM glycine, 0.5 % (w/v) SDS; freshly add: methanol (final concentration: 20 %)
10× SDS-PAGE running buffer	250 mM Tris,1920 mM glycine, 1 % (w/v) SDS, pH 8.3
1× sonication buffer	10 mM Tris pH 8.0, 200 mM NaCl, 1 mM EDTA, 0.1 % Na-Deoxycholate, 0.5 % n-Lauroylsarcosine, 1× Halt Protease Inhibitor Cocktail
6× Laemmli buffer	375 mM Tris pH 6.8, 12 % SDS, 60 % glycerol, 300 mM DTT, 0.06 % bromphenol blue

1.2 Commercial kits and reagent

Table 2. Commercial kits and reagents.

Kit/ reagent	Item number	Company
Agencourt AMPure XP, 60 ml	A63881	Beckman Coulter Life Sciences, USA
Agencourt RNACLEAN XP, 40 ml	A63987	Beckman Coulter Life Sciences, USA
Agilent High Sensitivity DNA Kit	5067-4626	Agilent Technologies, USA
AllPrep DNA/RNA Micro Kit	80284	Qiagen, Netherlands
AllPrep DNA/RNA Mini Kit	80204	Qiagen, Netherlands
Applied Biosystems SYBR Green PCR Master Mix	4309155	Thermo Fisher Scientific, USA
Pierce BCA Protein Assay Kit	23227	Thermo Fisher Scientific, USA
ß-mercaptoethanol	63689	Sigma-Aldrich, Germany
BSA (solution)	B9001S	New England Biolabs, USA
BSA (10 %; FACS)	SRE0036	Sigma-Aldrich, Germany
BSA (powder; Western blot)	A3294	Sigma-Aldrich, Germany
CD19 MicroBeads, human	130-050-301	Miltenyi Biotec, Germany
CellEvent Senescence Green Flow Cytometry Assay Kit	C10840	Thermo Fisher Scientific, USA
Chromium single cell 3' reagent kit v2		10x Genomics, USA
Chromium Single Cell 3’ Library & Gel Bead Kit v2	PN-120237	10x Genomics, USA
Chromium Single Cell A Chip Kit	PN-1000009	10x Genomics, USA
Chromium i7 Multiplex Kit	PN-120262	10x Genomics, USA
Clarity Western ECL substrate	1705061	Bio-Rad, Germany
Dimethyl sulfoxide (DMSO)	D2650	Sigma-Aldrich, Germany
Pierce 16% Formaldehyde (w/v), Methanol-free	28906	Thermo Fisher Scientific, USA
Glycine (1 M)	33226	Sigma-Aldrich, Germany
Glycerol	15523	Sigma-Aldrich, Germany
GlycoBlue Coprecipitant (15 mg/ml)	AM 9515	Ambion/Thermo Fisher Scientific, USA
Halt Protease Inhibitor Cocktail (100X)	87786	Thermo Fisher Scientific, USA
High Sensitivity D1000 Reagents	5067-5585	Agilent Technologies, USA
High Sensitivity D1000 ScreenTape	5067-5584	Agilent Technologies, USA
High Sensitivity D5000 Reagents	5067-5593	Agilent Technologies, USA
High Sensitivity D5000 ScreenTape	5067-5592	Agilent Technologies, USA
IGEPAL CA-630	I8896	Sigma-Aldrich, Germany
Immun-Blot PVDF Membrane	1620177	Bio-Rad, Germany
Fast SYBR Green Master Mix (2×)	4385612	Thermo Fisher Scientific, USA
Micrococcal Nuclease (300 U/µL)	EN0181	Thermo Fisher Scientific, USA
Mini-PROTEAN TGX Stain-Free Protein Gel (4-20%)	456-8094	Bio-Rad, Germany
$MgCl_2$ (25mM)	R0971	Fermentas/Thermo Fisher Scientific, USA
NEBNext High-Fidelity 2x PCR Master Mix	M0541S	New England Biolabs, USA
NEBNext Multiplex Oligos for Illumina	E7335S	New England Biolabs, USA
NEBNext Ultra Directional RNA Library Prep Kit	E7420L	New England Biolabs, USA
NEBNext Ultra II Directional RNA Library Prep Kit	7760S/L	New England Biolabs, USA
Nextera DNA Sample Preparation Kit (24 Samples)	15028212	Illumina, USA
Tagment DNA Enzyme (Tn5 transposase)	15027865	Illumina, USA

Kit/ reagent	Item number	Company
Tagment DNA buffer (TD-buffer; 2x)	15027866	Illumina, USA
Nitrocellulose membrane	1620112	Bio-Rad, Germany
Nuclei EZ Lysis Buffer	N3408	Sigma-Aldrich, Germany
NucleoSpin RNA Plus kit	740984	Macherey-Nagel, Germany
PageRuler Plus Prestained Protein Ladder, 10 to 250 kDa	26619	Thermo Fisher Scientific, USA
Panobinostat	13280	Cayman Chemical, USA
Phenylmethylsulfonyl fluoride (PMSF)	6367	Carl Roth, Germany
Powdered milk	T145.2	Carl Roth, Germany
Powdered milk	1602	GERBU Biotechnik, Germany
Propidium Iodide (1 mg/ml solution in water)	P3566	Thermo Fisher Scientific, USA
Protease Inhibitor Cocktail (100X)	5871	Cell Signaling Technology, USA
Protein G Magnetic Beads (ChIP-Grade)	9006	Cell Signaling Technology, USA
Proteinase K solution (20mg/mL)	M3037.0005	Genaxxon bioscience, USA
Protein G Magnetic Beads (ChIP-Grade)	9006	New England Biolabs, USA or Cell Signaling Technology, USA
Qubit dsDNA HS assay Kit	Q32851	Thermo Fisher Scientific, USA
Qubit RNA HS assay Kit	Q32852	Thermo Fisher Scientific, USA
Ribo-Zero gold rRNA removal kit (human/ mouse/ rat)	MRZH12324	Illumina, USA
Ribo-Zero magnetic gold kit (Human/Mouse/Rat)	MRZG126	Epicentre Biotechnologies, USA
RiboLock RNase Inhibitor (40 U/µL)	EO0382	Thermo Fisher Scientific, USA
RNase A, DNase and protease-free (10 mg/ml)	EN0531	Thermo Fisher Scientific, USA
RNase-free rDNase	740963	Macherey-Nagel, Germany
ROTI-Free Stripping Buffer	0083.1	Carl Roth, Germany
RQ1 RNase-Free DNase	M6101	Promega, USA
SMARTer Stranded Total RNA-Seq Kit v2 - Pico Input Mammalian	634412	Takara Bio, USA
Tapestri Single-Cell DNA Kit V2	MB03-0019	Mission Bio, USA
tert-butyl methyl ether	6746.1	Carl Roth, Germany
TO-PRO-3 Iodide (642/661) - 1 mM Solution in DMSO	T3605	Thermo Fisher Scientific, USA
Triton X-100	1.08643.1000	Merck, Germany

1.3 Antibodies

Table 3. Primary antibodies for immunoblotting

Entity	Target	Antibody	Host species	Dilution	ID	Company
MM	H3K27ac	Polyclonal Anti-histone H3 (acetyl K27) antibody	Rabbit	1:1000	ab4729	Abcam, UK
	H3K9ac	Polyclonal Anti-histone H3K9ac antibody	Rabbit	1:1000	39137	Active Motif, USA
	H4ac	Anti-acetyl-Histone H4 Antibody	Rabbit	1:2000	06-866	Merck, Germany
	H3	Histone H3 (1B1B2) Mouse mAb	Mouse	1:1000	14269	Cell Signaling Technology, USA
	GAPDH	Monoclonal anti-GAPDH antibody	Mouse	1:1000	AM4300	Thermo Fisher Scientific, USA
	BTG1	Recombinant Anti-BTG1 [EPR8274(2)]	Rabbit	1:1000	ab151740	Abcam, UK
	MYBL2	Monoclonal Anti-MYBL2, (C-terminal) antibody	Mouse	1:500	SAB1404099	Sigma-Aldrich, Germany
AML	H3	Polyclonal anti-Histone H3 antibody	Rabbit	1:3000-1:5000	ab1791	Abcam, UK
	H3K4me3	Polyclonal anti-Histone H3 (tri methyl K4) antibody	Rabbit	1:250	ab8580	Abcam, UK
	H3K9me3	Polyclonal anti-Histone H3 (tri methyl K9) antibody	Rabbit	1:1000	ab8898	Abcam, UK
	H3K36me3	Polyclonal anti-Histone H3 (tri methyl K36) antibody	Rabbit	1:1000	ab9050	Abcam, UK
	H3K27me3	Polyclonal Histone H3K27me3 antibody	Rabbit	1:1000	39155	Active Motif, USA

Table 4. Antibodies for chromatin immunoprecipitation (ChIPmentation)

Target	Antibody	Host species	Amount	ID	Lot number	Company
H3K4me1	Polyclonal anti-Histone H3 (mono methyl K4) antibody	Rabbit	4 µg per 1 x 10^6 cells	ab8895	GR193555-1	Abcam, UK
H3K27ac	Polyclonal anti-Histone H3 (acetyl K27) antibody	Rabbit	4 µg per 1 x 10^6 cells	ab4729	GR184557-2	Abcam, UK
H3	Polyclonal anti-Histone H3 antibody	Rabbit	4 µg per 1 x 10^6 cells	ab1791	GR203824-1	Abcam, UK

Table 5. Secondary antibodies for immunoblotting (detection by chemiluminescence)

Target	Antibody	Host species	Dilution	ID	Company
Anti-rabbit	Anti-rabbit horseradish peroxidase (HRP)-linked antibody	Goat	1:2000-1:5000	7074	Cell Signaling Technology, USA
Anti-mouse	Anti-mouse horseradish peroxidase (HRP)-linked antibody	Horse	1:1000-1:3000	7076	Cell Signaling Technology, USA

Table 6. Secondary antibodies for immunoblotting (detection by fluorescence)

Target	Antibody	Host species	Dilution	ID	Company
Anti-rabbit	Goat anti-Rabbit IgG (H+L) Highly Cross-Adsorbed Secondary Antibody, Alexa Fluor 568	Goat	1:1000	A11036	Thermo Fisher Scientific, USA
Anti-mouse	Goat anti-Mouse IgG (H+L) Highly Cross-Adsorbed Secondary Antibody, Alexa Fluor 568	Goat	1:1000	A11031	Thermo Fisher Scientific, USA

1.4 Instruments

Table 7. Instruments.

Instrument	Company
BD FACSAria III cell sorter	BD Biosciences, USA
2100 Bioanalyzer Instrument	Agilent Technologies, USA
ChemiDoc MP Imaging System	Bio-Rad, Germany
Covaris S2	Covaris, USA
HiSeq 2000 Sequencing System	Illumina, USA
HiSeq 4000 Sequencing System	Illumina, USA
LUNA automated cell counter	Logos Biosystems, South Korea
NovaSeq 6000 System	Illumina, USA
Qubit 2.0 Fluorometer	Thermo Fisher Scientific, USA
StepOnePlus Real-Time PCR System	Thermo Fisher Scientific, USA
Mission Bio Tapestri Platform	Mission Bio, USA
T100 Thermo Cycler	Bio-Rad, Germany
TapeStation 4200	Agilent Technologies, USA

1.5 Software

Table 8. Software.

Software	Version	Source
Adobe Illustrator	15.0.2	Adobe systems, USA
Affinity Designer	1.8.2.620	Serif (Europe) Ltd., UK
BD FACSDiva Software	8.0.2	BD Biosciences, USA
BEDTools	2.25.0	(Quinlan & Hall, 2010)
FastQC	0.11.9	(Andrews, 2010)
FlowJo	10.6.1	BD Biosciences, USA
IGV tools	2.3.23	(Robinson *et al.*, 2011)
Image Lab	5.2.1-6.0.1	Bio-Rad, Germany
Integrative Genomics Viewer	2.6.2	(Robinson *et al.*, 2011)
java	1.8.0_191	(Arnold *et al.*, 2005)
macs2	2.1.2	(Zhang *et al.*, 2008)
MultiQC	1.7 (python2.7)	(Ewels *et al.*, 2016)
RStudio	1.0.153	RStudio, USA
SAMtools	1.3	(Li *et al.*, 2009)
Seurat	3.1.0	(Stuart *et al.*, 2019)
STAR	020201	(Dobin *et al.*, 2013)
Tapestri Insights	2.1	Mission Bio, USA

1.6 External data sources

Table 9. External data sources.

External data	Source	Reference
BEDTools human.hg19.genome	https://github.com/arq5x/bedtools/blob/master/genomes/human.hg19.genome	(Quinlan & Hall, 2010)
Bidirectionally transcribed loci	http://www.cancerepisys.org/data/cancerepisys_data/ (Sandra Koser, Division of Applied Bioinformatics, German Cancer Research Center, Heidelberg, Germany)	(Mallm *et al.*, 2019)
Blacklisted regions hg19	https://www.encodeproject.org/annotations/ENCSR636HFF/ (hg19/GRCh37)	(Amemiya *et al.*, 2019; Encode Project Consortium, 2012)
ChromHMM annotation for CLL/non-malignant $CD19^+$ B-cells	http://www.cancerepisys.org/data/cancerepisys_data/	(Mallm *et al.*, 2019)
ChromHMM annotation for GM12878	https://genome.ucsc.edu/cgi-bin/hgFileUi?db=hg19&g=wgEncodeAwgSegmentation	(Encode Project Consortium, 2012)
ChromHMM annotation for K562	http://hgdownload.cse.ucsc.edu/goldenPath/hg19/encodeDCC/wgEncodeAwgSegmentation/wgEncodeAwgSegmentationChromhmmK562.bed.gz	(Encode Project Consortium, 2012)
Ensembl genome annotation version 87 for GRCh37	ftp://ftp.ensembl.org/pub/grch37/current/gtf/homo_sapiens/Homo_sapiens.GRCh37.87.chr.gtf.gz; downloaded 14.11.2018	(Cunningham *et al.*, 2019)
Human reference genome hs37d5	ftp://ftp-trace.ncbi.nih.gov/1000genomes/ftp/technical/reference/phase2_reference_assembly_sequence/hs37d5.fa.gz; downloaded 14.11.2018	(The 1000 Genomes Project Consortium *et al.*, 2015)

External data	Source	Reference
Human reference genome index for genome hs37d5	ftp://ftp-trace.ncbi.nih.gov/1000genomes/ftp/technical/reference/phase2_reference_assembly_sequence/hs37d5.fa.gz.fai; downloaded 14.11.2018	(The 1000 Genomes Project Consortium *et al.*, 2015)
RefSeq TSS	https://sourceforge.net/projects/seqminer/files/Reference%20coordinate/ (refGene_hg19_TSS.bed)	(O'Leary *et al.*, 2016)
RSeQC hg19 annotation	https://sourceforge.net/projects/rseqc/files/BED/Human_Homo_sapiens/hg19_GencodeCompV19.bed.gz; downloaded 14.11.2018	(Wang *et al.*, 2012)
run_spp.R	https://github.com/crazyhottommy/phantompeakqualtools/blob/master/run_spp.R; 07.11.2017	(Kundaje, 2013)

1.7 Patient samples

Written informed consent of all here studied patients and individuals was obtained by the respective clinical partners in the context of the funding. Cells of CLL patients (see Figure 19) were kindly provided by Dr. Daniel Mertens and Prof. Stephan Stilgenbauer (Department of Internal Medicine III, University Hospital Ulm, Ulm). Preprocessing of blood draws was conducted in the University Hospital Ulm in the diagnostics department. Enrichment of $CD19^+$ CLL cells from mononucleated cells from peripheral blood of CLL patients was performed by magnetic-activated cell sorting (MACS) using CD19 MicroBeads (Miltenyi) according to the manufacturer's protocol. Samples with a leukocyte count in the blood of the patient of less than 10^{10} leukocytes per liter were enriched for $CD19^+/CD5^+$ CLL cells by fluorescence-activated cell sorting (FACS). Leukocyte count in peripheral blood of five CLL patients treated with ibrutinib was determined at every sampling time point (Figure 9).

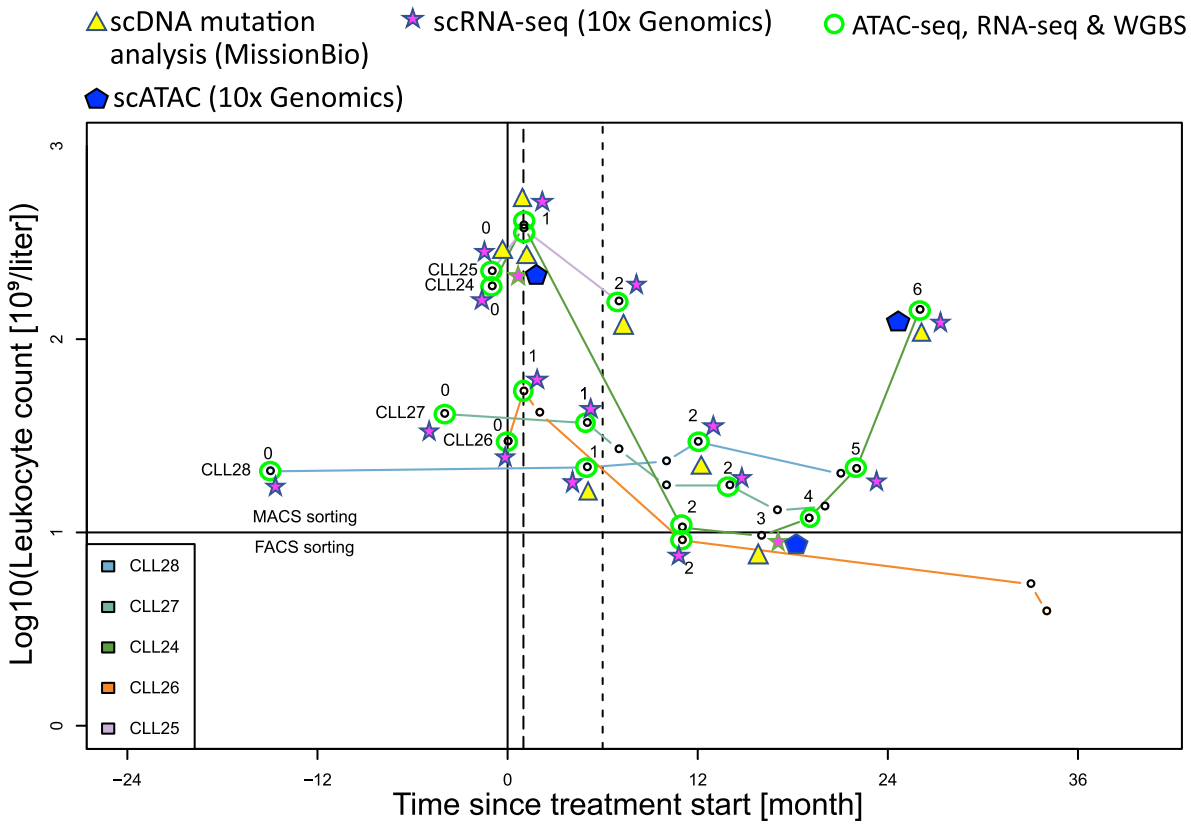


Figure 9 Tumor load and sampling timepoints of five CLL patients analyzed during ibrutinib therapy with illustration of conducted read-outs per timepoint.
Tumor load in CLL patients during ibrutinib therapy is plotted over time. Conducted read-outs are indicated at the corresponding timepoint.

Multiple myeloma patient cells were provided by the group of Prof. Marc-Steffen Raab (Research Group Translational Myeloma, German Cancer Research Center (DKFZ), Heidelberg, Germany) (Table 10; Figure 10).

Table 10. Multiple myeloma patients.

Patient ID	Sample source	Sorting
MM1	Bone marrow aspirate	$CD138^+$
MM2	Ascites	$CD138^+$
MM3	Pleural effusion	$CD138^+$
MM4	Bone marrow aspirate	$CD138^+$

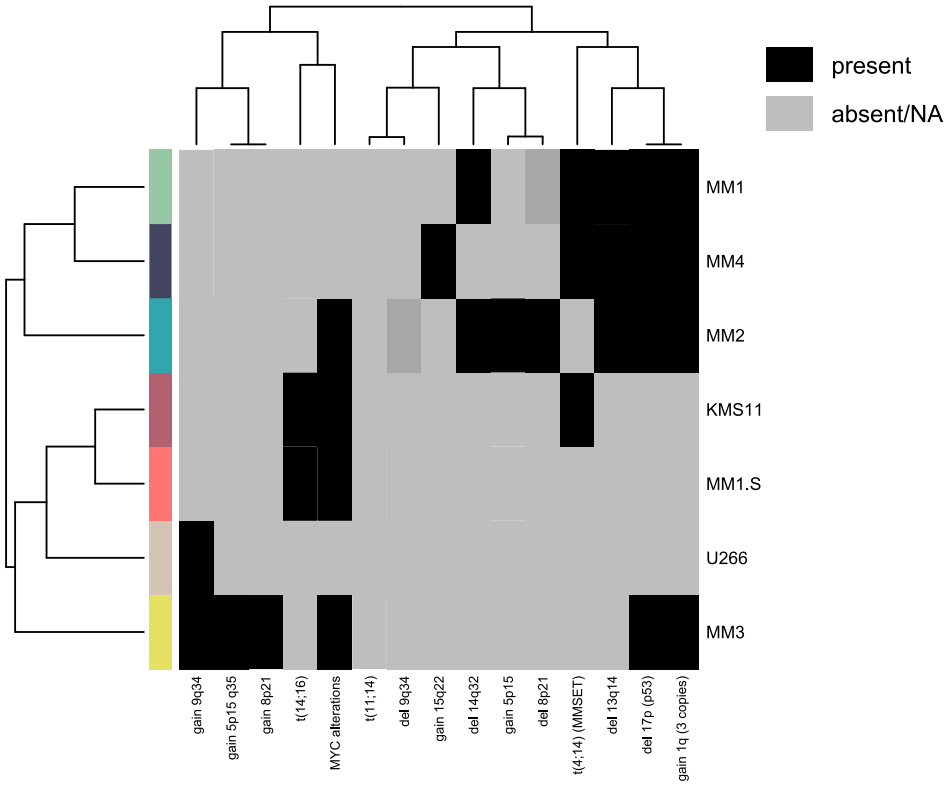


Figure 10 Illustration of genomic aberrations of the analyzed MM cell lines and MM patients.
Clustering of MM cell lines (KMS-11, MM1.S, U266) and patients (MM1-MM4) according to genomic aberrations was conducted with the function “heatmap” using Euclidean distance and complete linkage.

2 Oligonucleotide sequences

Table 11. Primers.

Oligo name	Sequence (5' → 3')	Application
Human *ND4* fwd	ACAAGCTCCATCTGCCTACG	ATAC-seq quality control qPCR mitochondrial gene
Human *ND4* rev	TTATGAGAATGACTGCGCCG	
Human *CBX3* fwd	TACTCCTCCCACCCTTGCG	ATAC-seq quality control qPCR positive control
Human *CBX3* rev	TAGTAACTAGCAGGACTCTAGCC	
Human *POU5F1* fwd	TTGGCTCTGGACCTTATCCC	ATAC-seq quality control qPCR negative control
Human *POU5F1* rev	TTCCATCACTGGCTCGTAGC	
Mouse *ND1* fwd	CTAATCGCCATAGCCTTCC	ATAC-seq quality control qPCR mitochondrial gene
Mouse *ND1* rev	TGGTTGTAAAATGCCGTATGG	
Mouse *USP5* fwd	TCCTCCTCACACAGCAGCTC	ATAC-seq quality control qPCR positive control
Mouse *USP5* rev	TCGCGCGTCTTGCCTTTAGC	
Mouse *POU5F1* fwd	TACAGCCAGCACTCTGGAGC	ATAC-seq quality control qPCR negative control
Mouse *POU5F1* rev	GACTCCTGCTACAACAATCGC	

Table 12. Sequencing library primers.

Oligo name	Sequence (5' → 3')	Application	Reference
Nextera index PCR primers - Index 1 read (custom)	CAAGCAGAAGACGGCATACGAGAT[i7]GTCTCGTGGGCTCGG	ATAC-seq/ChIPmentation - PCR	Adapted from Illumina, USA
Nextera index PCR primers - Index 2 read (custom)	AATGATACGGCGACCACCGAGATCTACAC[i5]TCGTCGGCAGCGTC	ATAC-seq/ChIPmentation - PCR	Adapted from Illumina, USA
Nextera transposase adapter read 1	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG	ATAC-seq/ChIPmentation - tagmentation	Illumina, USA
Nextera transposase adapter read 2	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG	ATAC-seq/ChIPmentation - tagmentation	Illumina, USA
TruSeq HT Index 1 (i7) Adapters (custom)	GATCGGAAGAGCACACGTCTGAACTCCAGTCAC[i7]ATCTCGTATGCCGTCTTCTGCTTG	RNA-seq	Adapted from Illumina, USA
TruSeq HT Index 2 (i5) Adapters (custom)	AATGATACGGCGACCACCGAGATCTACAC[i5]ACACTCTTTCCCTACACGACGCTCTTCCGATCT	RNA-seq	Adapted from Illumina, USA
TruSeq HT Universal (i5) Adapter (custom)	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT	RNA-seq	Adapted from Illumina, USA

3 Experimental procedures

3.1 Cell culture, treatment and sample preparation

Cell culture and panobinostat treatment was conducted by Sabrina Schumacher and Caroline Bauer (Division Chromatin Networks, DKFZ, Heidelberg).

3.1.1 Cell lines

The identity of the cell lines was regularly controlled using the multiplex human cell line authentication test service of Multiplexion (Germany).

Table 13. Multiple myeloma cell lines.

Cell line	RRID	Growth medium	Reference
KMS-11	CVCL_2989	Gibco RPMI 1640 Medium	(Namba *et al.*, 1989)
MM1.S	CVCL_8792		(Greenstein *et al.*, 2003)
U266	CVCL_0566		(Nilsson *et al.*, 1970)

3.1.2 Cell culture material

Table 14. Cell culture material.

Medium/ supplement	Company
Fetal bovine serum (FBS) Good Forte	PAN Biotech GmbH, Germany
Gibco RPMI 1640 Medium, GlutaMAX (61870010)	Thermo Fisher Scientific, USA

3.1.3 Cell culture of multiple myeloma cell lines

Cells were cultured in Gibco RPMI 1640 medium. For cultivation KMS-11 cells were seeded at a density of 2 * 10^6 in 15 ml medium in T-75 flasks. MM1.S and U266 cells were seeded with 3 * 10^6 cells in 13 ml and 10 ml medium, respectively. Cells were splitted alternating every third or fourth day. For flow cytometry experiments cells were seeded in T-25 flasks at a density of 1 x 10^6 cells in 7 ml (KMS-11), 1.5 x 10^6 cells in 7 ml (MM1.S) or 1.5 x 10^6 cells in 5 ml (U266).

3.1.4 Panobinostat treatment

Cells were always seeded 24 hours before the drug was added to the culturing media. Treatment of MM cell lines with panobinostat for sequencing analysis experiments was conducted for 24 hours. Panobinostat was freshly diluted with dimethyl sulfoxide (DMSO) in a serial dilution from a 10 mM stock solution in DMSO. Equal volumes of DMSO were added to the media of control cells. After 24-hour treatment, cells were harvested by scraping cells from the flask bottom (KMS-11) or by pipetting (MM1.S, U266). Cells were counted with a Luna cell counter, aliquoted and either directly processed or frozen in 1 ml freezing medium. For the analysis of senescence induction, cells were treated for 7 days with panobinostat. Panobinostat treatment was refreshed by medium exchange at least every 3 days to ensure a constant panobinostat concentration. If necessary, cells were splitted once during the 7-day treatment in order to avoid effects caused by a too high cell density. The potency of panobinostat in different experiments was controlled by testing its ability to induce global histone H3 hyperacetylation at residue K27 by immunoblotting.

3.1.5 Formaldehyde-fixation of cells

Cross-linking of proteins and DNA was conducted with formaldehyde. 16% (w/v) formaldehyde (Thermo Fisher Scientific) was diluted with 1 x PBS to a 1 % dilution. Cells were pelleted by centrifugation for 5 min at 300 x g. The supernatant was discarded and the cells washed once with 1 x PBS. Cell pellets were resuspended and incubated in freshly prepared 1 % formaldehyde solution for 10 min at room temperature with agitation. Fixation was stopped by incubation with glycine at a final concentration of 125 mM with agitation for 5 min at room temperature. The pellet of fixed cells was washed three times with 1 x PBS supplemented with 0.1 mM protease inhibitor phenylmethylsulfonyl fluoride (PMSF) by centrifugation for 4 min at 1,300 x g and frozen at -80 °C.

3.1.6 Freezing of viable cells

Cells were viable frozen by resuspending cell pellets of at least 10^5 cells in freezing medium consisting of 90 % FBS and 10 % DMSO. Cell suspension was carefully homogenized by pipetting and slowly cooled to -80 °C overnight using a Mr. Frosty freezing container (Thermo Fisher Scientific).

3.2 Flow cytometry

Flow cytometry experiments were conducted in collaboration with Philipp Mallm (Single-cell Open Lab, DKFZ, Heidelberg) with a BD FACSAria III cell sorter and cell preparation was done with technical support by Sabrina Schumacher (Division Chromatin Networks, DKFZ, Heidelberg).

3.2.1 Ethanol fixation of cells

Cells were pelleted at 300 x g for 4 min at 4 °C. For ethanol fixation cells were kept one ice during all processing steps. The supernatant was discarded and cells resuspended in 1 ml PBS (4 °C) by pipetting to yield a single cell suspension. Then, 3 ml 100 % ethanol (-20 °C) was added dropwise during constant vortexing, before samples were incubated for 30-60 min one ice. Cells were frozen at -20 °C till further usage.

3.2.2 DNA staining of fixed cells by TO-PRO-3 Iodide

Cells were washed twice with 1x PBS by centrifugation at 400 x g for 4 min at 4 °C. The supernatant was removed and cells resuspended and incubated for 30 min in 300 µl TO-PRO-3 Iodide staining solution consisting of 0.5 µM TO-PRO-3 Iodide (1:2000 dilution of stock, Thermo Fisher Scientific) and RNase A diluted 1:1000 in PBS.

3.2.3 Cell cycle stage determination after 24-72 hours panobinostat treatment of MM cell lines

For the determination of the distribution of cell cycle stages after 0-72 hours treatment with 8 nM panobinostat of MM cell lines, the cell lines KMS-11, MM1.S and U266 were freshly seeded and cultured for a total of 96 hours. All cells were cultivated untreated for at least 24 hours. Every 24 hours panobinostat treatment was started for a subset of cells from each cell line. DMSO treatment of cells as control was started after 24 hours thus conducted for 72 hours. Cells treated for 24, 48 or 72 hours with 8 nM panobinostat were harvested and counted with a LUNA cell counter. A total of 3 * 10^6 cells per sample were fixed with ethanol and frozen (see Materials and Methods chapter 3.2.1). One the day of the analysis by flow cytometry, DNA of cells was stained using TO-PRO-3 Iodide (see Materials and Methods chapter 3.2.2) and cells were analyzed with a BD FACSAria III cell sorter.

3.2.4 ß-galactosidase activity and DNA content measurement by flow cytometry

The ß-galactosidase activity as measure of cellular senescence was measured in MM cell line cells after treatment for 7 days with 8 nM panobinostat or DMSO as control using the CellEvent Senescence Green Flow Cytometry Assay Kit according to the manufacturer's descriptions. In short, cells were harvested and counted using a Luna cell counter. 8 x 10^5 cells per condition were washed in 1% BSA (Sigma-Aldrich, Germany) in PBS and remove supernatant by centrifugation for 5 min at 300 x g. Cells were resuspended in 100 µl fixation solution (2% formaldehyde in PBS) and incubated at room temperature for 10 min, while protected from light. Fixed cells were centrifuged for 5 min at 400 x g and washed with 200 µl 1% BSA in PBS. Then, cells were resuspended in 100 µl working solution (1:500 dilution

of CellEvent Senescence Green Probe in prewarmed CellEvent Senescence Buffer) and incubated for 1 hour at 37 °C. Cells were washed with 200 µl 1% BSA in PBS and centrifuged for 5 min at 400 x g. Then, stained cells were permeabilized by incubation for 15 min at room temperature in 100 µl 0.25% Triton X-100 in PBS containing 1% BSA. Permeabilized cells were centrifuged for 5 min at 400 x g and washed once in 150 µl 1% BSA in PBS. The supernatant was removed by centrifugation for 5 min at 500 x g. Cells were resuspended in 300 µl 1% BSA in PBS containing propidium iodide diluted 1:500 and RNase A (1:1000) for nucleic acid staining for DNA content measurement. Suspension was transferred to flow cytometry assay tubes and flow cytometry analysis was conducted using a BD FACSAria III cell sorter with a 488 nm laser and a 530 nm/30 filter.

3.2.5 Flow cytometry data analysis of cellular senescence and DNA content

The analysis of flow cytometry data was conducted with the R package flowCore version 1.52.1 (Ellis *et al.*, 2019) using R version 3.6.2 based on recommendations of Klinke et al. (Klinke & Brundage, 2009). FCS-formatted output files of the flow cytometer BD FACSDiva Software version 8.0.2 were loaded for all samples of an experiment using the function "read.flowSet" with the setting "transformation = FALSE". Apoptotic cells were gated based on forward scatter area (FSC-A) and side scatter area (SCC-A). Apoptotic cells were removed using a polygon gate by applying the function "polygonGate" with the values listed in Table 15.

Table 15. Settings for gating apoptotic cells.

Apoptotic cells (polygon gate)	Experiment1		Experiment 2 & 3	
	FSC-A	SSC-A	FSC-A	SSC-A
CORNER 1	0	25000	0	100
CORNER 2	150000	275000	400	1100
CORNER 3	150000	275000	300	1100
CORNER 4	0	275000	0	1100

Gating was visually inspected using the function "xyplot" from the R package flowViz version 1.50.0 (Sarkar *et al.*, 2008). All gates were applied using the function "Subset" of flowCore. Cells were further gated data-driven on their morphology by fitting individually for each sample a bivariate normal function to the distribution of cells in the FSC-A/SSC-A space using the function "norm2Filter". All cells within 1.2 standard deviations ("scale = 1.2") were selected. Debris was removed with a cutoff on FSC-A of 50 (experiments 2 and 3) or 25,000 (experiment 1) using the function "rectangleGate". Doublets were removed by extracting single cells with the function "gate_singlet" from the R package flowStats version 3.44.0 (Hahne *et al.*, 2020) which fits the forward scatter height (FSC-H) as function of the FSC-A with a robust linear model. Cells within the 95 % prediction band were kept ("prediction_level = 0.95"). Fluorescence values of ß-galactosidase activity were transformed linearly for smaller values (< 100) and logarithmic (log10) at higher values (> 100) using the "linlogTransform"

function described by Klinke et al. (Klinke & Brundage, 2009) using the settings “median = 0, dist = 100”. The ß-galactosidase activity threshold for the discrimination of senescent cells was determined individually for each cell line and experiment from the signal distribution of the corresponding untreated control cells. The upper peak limit was identified with the “curv1Filter”-function of the R package flowStats (Hahne *et al.*, 2020) using a bandwidth of 3 (“bwFac=3”). A gate for senescent cells was then defined and cells extracted using the flowCore function “rectangleGate”. Different gates were assigned to a dataset using the function “filter” applied using the “Subset”-function. Statistics were assessed using the function “fsApply”. Distributions of ß-galactosidase activity and DNA-content were visualized using the ggplot2 version 3.3.2 function “geom_density” and panels were arranged with the function “facet_grid”. Flow cytometry data of DNA content of cells treated 0-72 hours with 8 nM panobinostat was analyzed according to the above-mentioned steps, however by manual gating single cells using the software FlowJo.

3.3 Western blots

3.3.1 Histone extraction

Cells (10^6) were pelleted by centrifugation for 5 min at 300 x g. The supernatant was discarded and the pellet dissolved in 100 µl 0.25 M HCl by pipetting up and down. After overnight incubation at 4 °C on a rotator, cells were centrifuged for 5 min at 4 °C and 20,800 x g. The supernatant was transferred to a new tube and kept at 4 °C. The pellet was again resuspended in 0.25 M HCl (50 µl) for total extraction of all soluble proteins. Resuspension was centrifuged for 5 min at 4 °C and 20,800 x g. The supernatant was combined with the previous. When necessary, samples were stored at -20 °C overnight. Histones were precipitated similar as described in Shechter et al. (Shechter *et al.*, 2007). Trichloroacetic acid (6.1 N) was added to a final concentration of 33 % followed by an incubation on ice for 1 h. Histones were pelleted by centrifugation for 30 min at 4 °C (16,200 x g). The supernatant was discarded and the pellet washed with 700 µl tert-butyl methyl ether/ethanol (1:1). The supernatant was again removed by centrifugation for 10 min at 4 °C and 16,200 x g. The washing step was repeated with 300 µl tert-butyl methyl ether/ethanol (1:1). The pellet was air dried for 15-20 min at room temperature, dissolved in 83.3µl water and histones at the tube wall carefully resuspended. 16.6 µl Laemmli buffer (6 x) was added and the solution incubated at 95 °C for 10 min to denature the proteins. The extracted histones were stored at -20 °C till usage for immunoblotting experiments.

3.3.2 Cell lysis for whole protein extraction

Cells ($7 * 10^6$) were washed with 1x PBS, resuspended in 500 µl RIPA buffer and incubated for 1 hour on ice. The cell lysate was centrifuged for 30 min at 4 °C and 21,100 x g. The supernatant was transferred to new tubes on dry ice and frozen at -80 °C. Protein concentration was determined by using 5-10 µl lysate for a BCA assay measurement. For immunoblotting samples were mixed with 6x Laemmli buffer (final conc.: 1x) and stored at -20 °C.

3.3.3 Immunoblotting

Extracted histones were incubated for 5 min at 95 °C and size separated using SDS-PAGE on a Mini-PROTEAN TGX Stain-Free Protein Gel (Bio-Rad) for 45 min at 160 V in 1x SDS-PAGE running buffer. Proteins were transferred for 1 hour with 200 mA to a polyvinylidene fluoride (PVDF) membrane (Bio-Rad) in 1x transfer buffer. The blot was blocked with 5 % (w/v) powdered milk (Carl Roth) in 1x PBS for 1 hour, before washing once with 1x PBS. Immunoblotting with the indicated antibody was conducted overnight at 4 °C in 5 % (w/v) milk powder in PBS-T with agitation. The blot was washed three times for 5 min with 1x TBS-T with agitation. Each following blot washing step was identically conducted. The secondary antibody was applied in 5 % (w/v) milk powder in PBS for 1 hour at room temperature and access antibody removed by washing. GAPDH or H3 signal from the same blot for internal reference were obtained by stripping the blot with Roti-Free Stripping buffer (Carl Roth) for 30 min at 56 °C before blot was washed and reprobed. For chemiluminescence detection immunoblots were incubated for 5min with Clarity Western ECL Substrate (Bio-Rad) or for 1 min with SignalFire Plus ECL Reagent (Cell Signaling Technology). The signal was detected with a ChemiDoc MP Imaging System (Bio-Rad). Detection reagents were removed by washing before stripping and reprobing blots. After reprobing with anti-histone H3 antibody, immunoblots were developed by incubation for 1 min with Clarity Western ECL Substrate (Bio-Rad).

Immunoblotting experiments of MM cell lines treated with 8 nM panobinostat were conducted with whole protein lysates instead of histone extractions. Blocking was performed with 5% BSA in TBS-T. Antibody incubations were conducted using 1% BSA in TBS-T. The signal was detected by fluorescence using a secondary antibody coupled to Alexa 568.

Immunoblot quantification was done with the software Image Lab (Bio-Rad). Background signal was subtracted, before ratios of integrated signal intensities of bands were calculated between a histone modification and its corresponding internal control (H3 or GAPDH). For immunoblots of AML samples the ratios were divided by the ratio of untreated (DMSO) IDH1 wild-type cells of the corresponding blot for normalization and logarithmized (base 2). For immunoblots of MM samples ratios were normalized to the ratio of untreated KMS-11 of the same blot.

3.4 Quantitative real-time polymerase chain reaction

3.4.1 HeLa LT DNA extraction

2 x 10^6 cells were washed with 1x PBS and centrifuged at 300 x g for 5 min. After resuspension in 80 µl Gitschier buffer, cells were incubated for 1 h at 37 °C and 10 min at 100 °C. Tubes were centrifuged for 5 min at 20,800 x g at 4 °C. Then, DNA was precipitated by adding 80 µl isopropanol, 8 µl ammonium acetate and 0.3 µl GlycoBlue Coprecipitant (15 µg/µl; Thermo Fisher Scientific) and incubation over night at -20 °C. The sample was centrifuged for 30 min at 20,800 x g at 4 °C and washed for 10 min with 80 µl 75 % ethanol. The supernatant was removed after centrifugation at 4 °C for 5 min (20,800 x g). Pellet was dried and resuspended in 150 µl Tris-HCl pH 8 (stock: 10 mM). The DNA concentration was determined by a Qubit dsDNA HS assay Kit measurement to be 40.6 ng/µl.

3.4.2 Standard DNA dilution series for primer efficiency analysis

Isolated HeLa DNA was diluted in a series to 50, 12.5, 3.125, 0.78, 0.19 ng/2µl. Per replicate 2 µl HeLa DNA dilution, 5 µl Fast SYBR Green Master Mix (2 x; Thermo Fisher Scientific), 1 µl primer mix (forward and reverse primer each 5 µM in H_2O) and 2 µl H_2O were used for the qPCR. Measurement was done in technical triplicates.

3.4.3 qPCR for ATAC-seq library quality control

Libraries were diluted to 0.5 ng/µl and 2 µl (=1 ng) of the dilution was used per reaction. Each reaction was conducted in a total volume of 10 µl and contained 5 µl Fast SYBR Green Master Mix (2 x; Thermo Fisher Scientific) or Applied Biosystems SYBR Green PCR Master Mix (2 x; Thermo Fisher Scientific), 1 µl of the corresponding primer mix (forward and reverse primer each 5 µM in H_2O) and 2 µl H_2O. All measurements were performed in technical triplicates in 96-well format with a StepOnePlus Real-Time PCR System using the StepOne Software version 2.3. PCR was performed with fast ramp speed using the following program: 95 °C for 10 min and 35 cycles of [95 °C for 10 s, 62 °C for 10 s, 72 °C for 30 s]. The mean C_t value of technically replicates was used for the calculation of delta C_t values.

3.5 ATAC-seq

For the low input protocol of viable frozen cells (100,000 cells or less) cells were thawed at 37 °C in a water bath. Cells were pelleted by centrifugation for 5 min at 800 x g at 4 °C and washed once with 1x PBS. The supernatant was discarded and cells directly incubated with a mix of 5.5 µl H_2O, 10 µl TD-

buffer (Nextera DNA Sample Preparation Kit; Illumina), 2.5 µl Tn5 transposase (Nextera DNA Sample Preparation Kit; Illumina) and 2 µl 1 % IGEPAL CA-630 at 37 °C. After 30 min the transposition reaction was stopped with EDTA at a final concentration of 10 mM for 5 min on ice. The PCR was conducted with the 21 µl transposition mix adding 8.4 µl $MgCl_2$ 25mM (final conc. 3.3 mM), 1.3 µl of a 25 µM dilution of each Nextera index PCR primer (custom) and 32 µl NEBNext High-Fidelity 2x PCR Master Mix. Magnesium was added as $MgCl_2$ at equimolar concentration to the previously added EDTA to prevent the inhibition of the polymerase during PCR. Tagmented fragments were amplified with a thermo cycler program of 72 °C for 5 min, 98 °C for 30 s, 98 °C for 30 s followed by 10 to 14 cycles of 98 °C for 10 s, 63 °C for 30 s, 72 °C for 30 s and a final step of 72 °C for 1 min. The final ATAC-seq library was purified with Agencourt AMPure XP beads (1.4x) and eluted in 20 µl nuclease-free water. Final DNA concentrations were measured with a Qubit 2.0 fluorometer using the Qubit dsDNA HS assay Kit. The library size distribution was determined using a TapeStation system (Agilent Technologies, USA).

ATAC-seq of formaldehyde-fixed cells was conducted with 5 x 10^5 (AML or CLL) cells. Nuclei were isolated with 200 µl Nuclei EZ lysis buffer (4 °C) by incubating for 5 min on ice. Nuclei were centrifuged for 5 min at 800 x g at 4 °C and the supernatant discarded. Lysis was conducted by resuspending the nuclei in 180 µl ATAC-lysis buffer for 5 min on ice. After centrifugation at 4 °C for 5 min at 800 x g, the supernatant was gently removed. The transposition reaction was started by adding 5 µl H_2O, 10 µl TD-buffer and 5 µl Tn5 transposase and incubated at 37 °C for 30 min. The reaction was stopped with EDTA (final conc. 20 mM). Then reverse crosslinking was conducted by adding 22.5 µl elution buffer, 0.5 µl proteinase K and 2 µl 5 M NaCl for 2 h at 65 °C. DNA was purified with Agencourt AMPure XP beads (2x) before PCR amplification and eluted in 25 µl H_2O. The PCR and subsequent steps were conducted as described above in the low input protocol. $MgCl_2$ was however not added to the PCR reaction as EDTA had been removed by the additional magnetic bead purification.

ATAC-seq of cultured multiple myeloma cell lines was conducted directly after treatment with 1 x 10^5 viable cells. Replicates of ATAC-seq libraries were obtained by independently repeating the *ex vivo* panobinostat treatment. Nuclei were isolated and lysed as described for formaldehyde-fixed cells. Transposition was conducted with 5 µl Tn5 transposase in a final volume of 20 µl. After inactivation of the transposase by 5 µl EDTA (100 mM), 5 µl of the reaction volume was used for PCR in the presence of $MgCl_2$ in a volume of 50 µl and all following steps conducted as described above.

ATAC-seq of MM patient cells treated *ex vivo* with panobinostat was conducted with 250,000-500,000 viable frozen cells per technical replicate. Technical replicates were obtained by splitting cells of a frozen vial after thawing and by conducting ATAC-seq sequencing library preparation in parallel. Nuclei isolation and lysis was performed as for formaldehyde-fixed cells. For transposition 2.5 µl Tn5 transposase in a total of 20 µl was used and the reaction stopped with 2 µl 100 µM EDTA in 10 mM Tris HCl pH 8. PCR was conducted with the complete transposition reaction in a total volume of 52 µl

in the presence of additional $MgCl_2$ as described above. All following steps were conducted as described in the low input protocol.

The transposition reaction of freshly sorted viable cells of B-cell reference subsets was conducted by the group of PD Dr. Marc Seifert at the Institute of Cell Biology (Cancer Research), Medical School, University of Duisburg-Essen, Nordrhein-Westfalen, Essen, Germany. Cells were resuspended in a 20 µl mix containing 0.1 % NP40, 5 µl Tn5 transposase, 10 µl 2X TD buffer (Nextera DNA Sample Preparation Kit, Illumina) and incubated for 30 min at 37 °C. The reaction was stopped with 5 µl 100 mM EDTA in 10 mM Tris-HCl (pH 8). Transposed cells were frozen and shipped to Heidelberg. The PCR reaction was conducted containing 5 µl 100 mM $MgCl_2$, 32 µl NEBNext High-Fidelity 2X PCR Master Mix (New England BioLabs) and 0.5 µl custom Nextera primers (100 µM). Amplified and cleaned libraries were sequenced 50 bp paired-end on a HiSeq4000 (Illumina).

3.6 RNA-seq

RNA library preparation was conducted by Sabrina Schumacher and Caroline Bauer (Division Chromatin Networks, DKFZ, Heidelberg).

3.6.1 RNA-seq of CLL patients during ibrutinib treatment and healthy B-cell subsets

RNA from CLL patient samples with 5 * 10^5 cells or less was isolated with AllPrep DNA/RNA Micro Kit (Qiagen). Samples with more than 5 * 10^5 cells were processed with the AllPrep DNA/RNA Mini Kit (Qiagen). Frozen cell pellets were resuspended for cell lysis in 350 µl RLT Plus buffer containing freshly added ß-mercaptoethanol and were used for RNA isolation according to manufacturer's protocol. DNA was degraded using RQ1 RNase-Free DNase and RNA purified with Agencourt RNACLEAN XP magnetic beads. RNA sequencing libraries were constructed using the SMARTer Stranded Total RNA-Seq Kit v2 according to the manufacturer's protocol. Library generation was conducted with custom adapters designed based on the sequence of Illumina's TruSeq HT adaptors (see Materials and Methods chapter 2 Table 12).

3.6.2 RNA-seq of MM cell lines

After panobinostat treatment 1 * 10^6 cells per condition were lysed with 350 µl LBP buffer of the NucleoSpin RNA Plus kit (Macherey-Nagel) and frozen at -80 °C. RNA purification was conducted with the NucleoSpin RNA Plus kit. Remaining DNA was digested with RNase free rDNase (Macherey-Nagel) by adding 1/10 volume rDNase solution to 1 volume sample or with 1 U/µg RNA RQ1 RNase-Free DNase. RNA was cleaned by ethanol precipitation. For this, 3 volumes of ice-cold ethanol (100 %),

1/10 volume 2-3 M sodium acetate and 1.5-2 µl GlycoBlue were added to 1 volume sample and incubated for at least 1 hour at -20 °C. After centrifugation for 15 min at 4 °C with 16,200 x g the pellet was washed with 70 % ethanol and air dried. The RNA pellet was resuspended in 30 µl RNase-free water and 1 µl RiboLock RNase Inhibitor (Thermo Fisher Scientific). Ribosomal RNA removal was done using the Ribo-Zero Gold rRNA removal kit (5 µM panobinostat treatment experiment, Illumina) or the Ribo-Zero magnetic gold kit (8 nM panobinostat treatment experiment, Epicentre Biotechnologies) starting with 2 µg RNA. RNA was again precipitated by adding 1/10 volume 3 M sodium acetate, 2 µl glycogen (10 mg/ml) and after mixing 3 volumes of ice-cold ethanol (100%). After freezing over night at -20 °C RNA was centrifuged at 10,000 x g for 30 min. The RNA pellet was washed with ice cold ethanol (70 %). After 5 min centrifugation at 10,000 x g the supernatant was discarded and the pellet air dried. RNA was resolved in 6 µl RNase-free water supplemented with RiboLock RNase Inhibitor (final concentration 1 U/µl) and stored at -80 °C. RNA libraries were prepared with the NEBNext Ultra Directional RNA Library Prep Kit (5 µM panobinostat treatment experiment, New England Biolabs) or NEBNext Ultra II Directional RNA Library Prep Kit (8 nM panobinostat treatment experiment, New England Biolabs) according to the manufacturer's instructions.

3.7 ChIP-seq

Chromatin immunoprecipitation followed by sequencing (ChIP-seq) was conducted by Sabrina Schumacher and Lisa Dressler (Division Chromatin Networks, DKFZ, Heidelberg) according to the ChIPmentation protocol optimized for fast sequencing library preparation from small cell numbers using Tn5 transposase published by Schmidl et al. (Schmidl *et al.*, 2015). Cells were harvested and centrifuged at 600 x g for 5 min at 4 °C after treatment. Cells were washed with 1 x PBS and resuspended in 1 x PBS for counting with a LUNA automated cell counter. Cross-linking was conducted by formaldehyde-fixation (see Materials and Methods chapter 3.1.5) and fixed cells frozen as pellet. For ChIPmentation 1 x 10^6 cells per immunoprecipitation were used as starting material. Nuclei isolation and fragmentation before immunoprecipitation was conducted for cells of all immunoprecipitations of a condition together in a single vial. Frozen fixed cells were thawed and resuspended in 1 x PBS supplemented with 0.5 mM PMSF. Cells were collected at 1,300 x g at 4 °C for 5 min and the supernatant was discarded. The cells were resuspended in 1.2 ml ice cold Nuclei EZ Lysis Buffer and incubated for 5 min at 4 °C with agitation. Nuclei were pelleted by centrifugation at 500 x g and 4 °C for 5 min. After supernatant removal nuclei were resuspended in 1.2 ml ice cold Nuclei EZ Lysis Buffer and incubated for 10 min at 4 °C with agitation. Nuclei were again collected by centrifugation and washed with 100 µl 1 x PBS containing 0.5 mM PMSF. The nuclei pellet was resuspended in MNase buffer supplemented with 1 x Halt Protease Inhibitor Cocktail for chromatin fragmentation with 6 U Micrococcal Nuclease (MNase) per 3 x 10^6 cells. Fragmentation with MNase was conducted for 5 min at 37 °C with agitation.

The reaction was stopped by thorough mixing with 1/10 of final sonication volume of 10 x sonication buffer and incubation for at least 5 min on ice. Sonication was conducted with a Covaris S2 in micro tubes (microTUBE AFA Fiber Pre-Slit Snap-Cap 6x16mm, PN 520045, Covaris, USA) in a total volume of 130 µl for 10 min (duty cycle: 20 %, intensity: 8, burst: 200, mode: frequency sweeping). Samples were centrifuged at 16,200 x g for 15 min at 4 °C and the supernatant transferred to new tubes. Fragmented chromatin was frozen in liquid nitrogen and stored till further usage at -80 °C. The sample volume was divided and each aliquot filled up to 250 µl with 1 x sonication buffer containing protease inhibitor for separate immunoprecipitation (IP) using 4 µg of the antibodies directed against H3K27ac, H3K4me1 and H3 listed in Table 4. Antibody incubation was conducted for 2 hours at 4 °C with agitation. 25 µl Protein G Magnetic Beads per sample were preequilibrated by removing the storage buffer using a magnet, washing twice with 1 x sonication buffer in a fivefold volume as the storage buffer for 10 min with agitation and resuspending the beads in the original volume of 1 x sonication buffer supplemented with protease inhibitor. Preequilibrated protein G beads were added to the antibody-bound chromatin and incubated over night at 4 °C while rotating. Supernatant was removed with a magnet. Samples were washed in consecutive washing steps each conducted by incubation for 2 min on a rotator in a volume of 500 µl at room temperature. Washing was performed with once 1 x sonication buffer, once 1 x high salt buffer, once 1 x lithium buffer and twice 10 mM Tris pH 8. Tris buffer was removed from beads and the beads of each IP resuspended in 30 µl tagmentation mix containing 5 µl Tn5 transposase, 15 µl TD buffer (both included in the Nextera DNA Sample Preparation Kit, Illumina) and 10 µl nuclease-free water. Transposition was conducted for 10 min at 37 °C and the reaction stopped with 150 µl Tn5 wash buffer. Supernatant was directly removed with a magnet and beads washed once with 150 µl Tn5 wash buffer and twice with 150 µl TE buffer. For elution the supernatant was removed and the beads resuspended in 50 µl elution buffer and incubated for 15 min at 37 °C with agitation. The elution step was repeated and both eluates combined. RNA was digested by adding 10 µg RNase A and incubating for 30 min at 37 °C. Reverse cross-link was conducted by overnight incubation at 65 °C with NaCl at a final concentration of 200 mM and 20 µg proteinase K. DNA was purified using Agencourt AMPure XP beads at a 1.8-fold volume of the sample and eluted in 20 µl nuclease-free water. Amplification of DNA was performed in a thermo-cycler in a volume of 50 µl containing 19 µl DNA, 5 µl nuclease-free water, 0.5 µl of one of each Nextera index PCR primer (custom, 100 µM) and 25 µl NEBNext polymerase mix. The conducted PCR program was 72 °C for 5 min, 98 °C for 30 s, 13 cycles of [98 °C for 10 s, 63 °C for 30 s, 72 °C for 30 s] and 72 °C for 1 min. The final library was cleaned with Agencourt AMPure XP beads (1.4x) and eluted in 20 µl nuclease-free water. The DNA concentration and the fragment size distribution were determined using a Qubit 2.0 fluorometer and either a TapeStation 4200 or a 2100 Bioanalyzer Instrument, respectively.

3.8 Single cell RNA-seq

Viable frozen cells were thawed at 37 °C in a water bath. Cells were centrifuged 3 min at 300 x g at room temperature. Supernatant was removed and cells were washed once with 1x PBS. Cells were resuspended in PBS and counted with a LUNA automated cell counter. Single cell RNA-seq library preparation was conducted with the Chromium platform (10x Genomics) according to the manufacturer's instructions with the Chromium single cell 3′ reagent v2 kit (10x Genomics). Cell number input per sample was targeted at a cell recovery rate of 5,000. The concentration of cDNA after amplification and of the final scRNA-seq library was quantified with a Qubit 2.0 fluorometer using the Qubit dsDNA HS assay Kit. The fragment size distribution was controlled in a High Sensitivity DNA ScreenTape analysis (D5000) both after cDNA amplification and after final scRNA-seq library generation.

3.9 Targeted single cell DNA-seq

Targeted single-cell sequencing was conducted with the Mission Bio Tapestri Platform using the Tapestri Single-Cell DNA Kit V2 and the CLL panel provided by Mission Bio. The CLL panel covered 34 genes with 286 amplicons (amplicon length range: 175-275bp). Cells were thawed in a water bath at 37 °C. PBS (1x) was added and the cell suspension centrifuged for 5 min at 400 x g at 4 °C before the supernatant was removed. Cells were washed with 1 ml ice-cold 1x PBS and PBS was removed with a second centrifugation step for 5 min at 400 x g at 4 °C. Cell pellet was resuspended in 40-100 µl "Cell buffer" (Mission Bio) and cells disaggregated by gently pipetting up and down. Cells were counted with a LUNA automated cell counter. Cell suspension was further diluted with the "Cell buffer" to about 2,000-4,000 cells/µl of which 35 µl were loaded on the Tapestri Platform and used for library preparation. Library preparation was conducted according to the manufacturer's descriptions, unless stated otherwise. In short, cells were encapsulated, lysed and barcoded. Barcoding primers were cleaved of the barcoding beads by UV light illumination for 8 min using the UV lamp supplied for CellenONE (Cellenion, France) before amplified in a targeted PCR with 2 x 11 cycles. Libraries were cleaned before sequencing adaptors were added in a second PCR. The library quality was assessed using a TapeStation with High Sensitivity D5000 ScreenTapes (Agilent Technologies, USA). The DNA concentration was determined with a Qubit 2.0 fluorometer using the Qubit dsDNA HS assay Kit (Thermo Fisher Scientific, USA) before sequencing.

4 Sequencing

All next generation sequencing was conducted at the Genomics & Proteomics Core Facility of the DKFZ (Heidelberg, Germany), which provided demultiplexed fastq-formatted files for data analysis.

4.1 Sequencing settings

Table 16. Sequencing settings

Project	Experiment	Sequencing type	Instrument model
AML	ATAC-seq	26 bp + 74 bp paired-end	Hiseq4000 (Illumina)
Healthy reference subpopulations H18-H22	ATAC-seq	50 bp paired-end	Hiseq4000 (Illumina)
Healthy reference subpopulations H18-H22	RNA-seq	50 bp single-end	Hiseq4000 (Illumina)
Healthy reference subpopulations H18-H22	WGBS	100 bp paired-end	HiseqX (Illumina)
H6-7; H9; H11-13; H15-17	ATAC-seq	50 bp paired-end	Hiseq4000 (Illumina)
CLL patients 1-23	ATAC-seq	50 bp paired-end	Hiseq4000 (Illumina)
CLL patients 25-28	ATAC-seq	50 bp paired-end	NovaSeq6000 (Illumina)
CLL patient 25 (before treatment)	ATAC-seq	50 bp paired-end	Hiseq4000 (Illumina)
CLL patients 9,11,19; H15, 16, 17	ATAC-seq	50 bp paired-end	Hiseq4000 (Illumina)
CLL patients 25-28	RNA-seq	50 bp single-end	Hiseq4000 (Illumina)
CLL patient 24	scRNA-seq	26 bp + 74 bp paired-end	NovaSeq6000 (Illumina)
CLL patient 24 (remission)	scRNA-seq	26 bp + 98 bp paired-end	Hiseq4000 (Illumina)
CLL patients 24,25,28	scDNA-seq	150 bp paired-end	NovaSeq6000 (Illumina)
MM patients	ATAC-seq	50 bp paired-end	Hiseq4000 (Illumina)
MM cell lines	ATAC-seq	50 bp paired-end	Hiseq4000 (Illumina)
MM cell lines	ChIP-seq	50 bp single-end	Hiseq4000 (Illumina)
MM cell lines	RNA-seq	50 bp single-end	Hiseq4000 (Illumina)

4.2 Adaptor sequences for read trimming

Table 17. ATAC-seq Nextera adapter trimming sequences supplied in "NexteraPE-PE.fa" file by Trimmomatic.

Name	Sequence
PrefixNX/1	AGATGTGTATAAGAGACAG
PrefixNX/2	AGATGTGTATAAGAGACAG
Trans1	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG
Trans1_rc	CTGTCTCTTATACACATCTGACGCTGCCGACGA
Trans2	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG
Trans2_rc	CTGTCTCTTATACACATCTCCGAGCCCACGAGAC

Table 18. Adapter trimming sequences for RNA-seq libraries generated with the NEBNext Ultra directional RNA Library Prep Kit.

Name	Sequence
Universal adapter	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT
Barcode adapter	GATCGGAAGAGCACACGTCTGAACTCCAGTCAC

5 Data analysis

5.1 Genome, indices and annotations

5.1.1 Reference genome and annotations

The genome hs37d5 was downloaded from ftp://ftp-trace.ncbi.nih.gov/1000genomes/ftp/technical/reference/phase2_reference_assembly_sequence/hs37d5.fa.gz and used as reference genome for all bulk sequencing data analysis. The Ensembl genome annotation version 87 was downloaded as gtf-formatted file from Ensembl (see Materials and Methods Table 9). Gtf-files were converted to bed-format files with the commands "gtfToGenePred" and "genePredToBed" of the UCSC user apps (Kent *et al.*, 2010). The annotation file "hg19_GencodeCompV19.bed" was downloaded for RSeQC from "https://sourceforge.net/projects/rseqc/files/BED/Human_Homo_sapiens/".

5.1.2 Bowtie2, STAR indices and RSEM references

Bowtie2 indices for the alignment of sequencing reads with bowtie2 were generated for hs37d5 with the function "bowtie2-build". The alignment of RNA-seq reads was conducted with STAR for which the STAR indices for hs37d5 were generated with the function "STAR --runMode genomeGenerate --sjdbOverhang 49". The downloaded fastq-file of hs37d5 and the gtf-file of the corresponding Ensembl genome annotation version 87 with adjusted chromosome names were provided as input for the parameter "--genomeFastaFiles" and "--sjdbGTFfile", respectively. RSEM references for the calculation of normalized RNA-seq read counts were generated by the function "rsem-prepare-reference" by specifying the downloaded hs37d5 fastq-file as "reference_fasta_file" and the Ensembl version 87 genome annotation with the parameter "--gtf".

5.1.3 Functional annotations of the genome

The list of "enhancer predictive states" (Mallm *et al.*, 2019) determined from seven functionally relevant histone modifications from CLL and non-malignant $CD19^+$ B-cells including samples listed in Figure 19 A were assembled from the consensus of ChromHMM states 1, 8, 9, 11. Bases overlapping with a region +/- 1 kb around a RefSeq TSS (see Materials & methods chapter 1.6) were removed with the

function “substract” of BEDTools version 2.25.0 (Quinlan & Hall, 2010). The list of “potential active enhancer containing loci” was assembled by combining genomic regions annotated as ChromHMM states 1 and 9. Bases within +/- 1 kb of a RefSeq TSS were removed as described above.

ENCODE project CTCF biding sites were downloaded from “http://hgdownload.cse.ucsc.edu/goldenPath/hg19/encodeDCC/wgEncodeUwTfbs/” (Encode Project Consortium, 2012). The GM12878 CTCF peak files “wgEncodeUwTfbsGm12878 CtcfStdPkRep1.narrowPeak.gz” and “wgEncodeUwTfbsGm12878CtcfStdPkRep2.narrowPeak.gz” version 06.06.2010 (GEO: GSM749704, GSM749706) were merged with the BEDTools utility suite (Quinlan & Hall, 2010) before usage.

For the functional annotation of class A-D regions (Results chapter 3.2.3) the ENCODE ChromHMM genome annotation of the lymphoblastoid cell line GM12878 was used (see Materials & methods chapter 1.6).

5.2 ATAC-seq data analysis

5.2.1 ATAC-seq data preprocessing pipeline

ATAC-seq data analysis was conducted with self-written bash scripts. Automatic sequential submission of the scripts to the queueing system of a cluster was performed by a self-written script by Dr. Nick Kepper (Division of Chromatin Networks, DKFZ, Heidelberg). After automatic confirmation of correct download of raw sequencing data in fastq format (md5sums), ATAC-seq reads were trimmed to remove adaptor sequences with Trimmomatic version 0.36 (Bolger *et al.*, 2014) in paired-end mode with the settings “ILLUMINACLIP:NexteraPE-PE.fa:2:30:10:8:TRUE SLIDINGWINDOW:5:20 MINLEN:25”. The Nextera adaptor sequences provided by Trimmomatic in the file “NexteraPE-PE.fa” are listed in Table 17. Bases at the end of a read are removed if the average sequencing quality of five consecutive bases decreases below a Phred quality score of 20 and reads of less than 25 bases are removed to avoid inaccurate alignment. Trimmed sequencing reads were aligned to the human genome build hg19 containing decoy sequences (hs37d5) with the options “--very-sensitive -X 2000 –fr” by Bowtie 2 version 2.3.3.1 (Langmead & Salzberg, 2012). SAMtools version 1.3 (Li *et al.*, 2009) was used to directly convert SAM formatted output files into BAM format (“samtools view -bS”). Then, reads in BAM file were sorted according to genomic location and indexed with “samtools sort” and “samtools index”. Reads falling into blacklisted regions of hg19 by ENCODE (Table 9) were removed with “bedtools intersect -v -sorted -abam” of BEDTools version 2.14.3 (Quinlan & Hall, 2010) and output saved in BED format with “samtools view -b”. Fragment size distributions were calculated by extracting the insert sizes from the BAM files by piping the output of “samtools view” to “cut -f9”. Only

insert sizes above zero were kept to avoid using one fragment size twice as negative insert sizes originate from the second mapped mate. The distribution was plotted with the function “hist” in R using an adapted script originally written by André Holzer (former master student, Division of Chromatin Networks, DKFZ, Heidelberg). Then, reads belonging to duplicated fragments were removed by “samtools rmdup” of SAMtools version 1.3 (Li *et al.*, 2009) to account for PCR duplicates. Plotting of fragment size distribution is repeated excluding also reads mapping to the mitochondrial genome by using “awk '{ if($3!="MT") print }' ” before extracting insert sizes from BAM files. In the analysis of the samples listed in Figure 19 B, reads with a mapping quality below 20 were removed with “samtools view -q 20” of SAMtools version 1.3. BAM formatted files were converted to BED format for easy adjustment of read position and peak calling with “bamToBed” of BEDTools version 2.25.0. Reads mapping to the mitochondrial genome were removed from BED format files using the function “awk” of GNU Awk version 4.0.2.

To represent the position of the transposase binding event reads in BED format were shifted. A custom perl script was used. To find the center of the transposition event, the start position of reads on the plus strand were shifted by +4 bp. As the transposase occupies 29 bp, the starting position of this 29 bp stretch was found by subtracting 14 bp from the center. The end position was calculated as shift of +29 bp from the start of the accessible region (“start accessible region” = “start original read” + 4 - 14; “end accessible region” = “start accessible region” + 29). For reads mapping to the minus strand the center is found -4 bp from the read’s end position. The end of the accessible region is +14 bp from the center and the start -29 bp from this end position (“end accessible region” = “end original read” - 4 + 14; “start accessible region” = “end accessible region” - 29). The BED formatted file was further sorted by genomic coordinate using “sort -k1,1 -k2,2n”. If necessary for a specific bioinformatic tool, adjustments of chromosome name convention were executed by the commands “sed -e 's/chr//'” (e.g., “chr1” to “1”) or “sed -e '/^[0-9XYM]/s/^/chr/'” (e.g., “1” to “chr1”). BED formatted shifted reads were converted to BAM format with “bedToBam” of BEDTools version 2.25.0 supplying the index file of the genome “hs37d5” (Table 9) after chromosome name adjustment with “sed -e 's/chr//'“. For inspection of genome wide density of raw shifted reads in a genome browser (see Materials and Methods chapter 5.14.1), BED files were tiled using the command “java -jar igvtools.jar count --windowSize 1 b37” of IGV tools version 2.3.23 with java version 1.8.0_272.

Regions of interest with enriched accessibility were identified by calling peaks using “macs2 callpeak” version 2.1.2 (Zhang *et al.*, 2008) with python 2.7 with the ATAC-seq data adjusted parameters “--nomodel --shift -10 --extsize 28 --broad --format BED”. Peaks were called from BED format files containing the original genomic location of sequenced reads except for the evaluation of peak calling settings. Here pre-shifted reads were used where indicated. For peak calling setting evaluation, the parameter “--extsize” was varied. The shift distance was always calculated from the chosen extension

size with the formula 4 - "extension size"/2 to account for the 4 bp difference between the sequencing read start and the center of the transposase binding event.

Fraction of reads in peaks (FRiP) scores were calculated by first counting every shifted read overlapping with the called peaks of a specific sample with "intersectBed" function of the BEDTools utility suite version 2.25.0 (Quinlan & Hall, 2010). The reads falling into peak regions were then divided by the total number of all final high-quality reads (used for peak calling) to receive the FRiP score for a certain sample. For further analysis and calculation of coverage tracks, all reads were shifted to be centered on the cutting position of the transposase and extended to a total of 29 bp to represent the region occupied by the transposase (Adey *et al.*, 2010). Quality was controlled between steps with the tool MultiQC version 1.7 (Ewels *et al.*, 2016).

5.2.2 ATAC-seq sequencing depth evaluation

An ATAC-seq library containing ~92 million reads was downsampled by randomly shuffling all reads with the bash function "shuf" before the first 5, 10, 20, 40, 80 million reads were extracted. For the correlation of the subsampled data to the full data, the genome (hg19) was segmented into 50 bp bins with the BEDTools function "makewindows". The downsampled read files were counted in the 50 bp regions by "coverageBed" and the Pearson correlation coefficient of every subset to the full data set was calculated using the binned counts. The reads were preprocessed with the ATAC-seq pipeline described in Materials and methods chapter 5.2.1 with some minor changes.

5.2.3 ATAC-seq peak calling setting evaluation

Peaks with JAMM version 1.0.7.2 (Ibrahim *et al.*, 2015) were called from reads shifted by +4 bp for plus strand reads and -4 bp for minus strand reads using the script "JAMM.sh" with the parameters "-f 1" recommended by JAMM for ATAC-seq (collapses reads to 5' end) and "-d y -e 5". JAMM was run with R version 4.0.2 and the R package mclust version 5.1. Peak calling with MACS (Zhang *et al.*, 2008) was conducted using preshifted reads (-71 bp for plus strand reads and +71 for minus strand reads) as no options are available to conduct read shifting during peak calling. Peaks were called with MACS version 1.4.2 using the options "--nomodel --nolambda --bw=150 --tsize=51 --shiftsize=75 --format=BED --pvalue=1e-5". Peak calling with MACS2 was conducted using different extension sizes (--extsize) and calculating the corresponding shift distance (--shift) as 4 - half extension size. Further settings were "--nomodel" to suppress the ChIP-seq specific model building and "--format BED --broad --broad-cutoff 0.1". Tagmentase occupancy at genome wide called peaks was plotted using ngs.plot version 2.61 (Shen *et al.*, 2014) by providing the called peaks in BED-format to the option "-E " and further selecting the options "-G hg19 -R bed -FL 28 -GO none".

5.2.4 ATAC-seq footprinting analysis

For ATAC-seq footprinting analysis, genomic locations of binding motif instances of a specific transcription factor in the peak set it was found to be enriched were identified by the HOMER function "annotatePeaks.pl". For this, motif files given by the previous HOMER motif enrichment analysis and the setting "hg19 -mbed" were used. Due to the palindromic nature of transcription factor motifs this function tends to report single motifs twice. To avoid a possible bias, only one motif instance of motifs within a 3 bp window was kept. All shifted reads from all CLL replicates were pooled. The same was done for all non-malignant samples. Reads were then collapsed to 3 bp centered on the transposase binding event and BED files converted to BAM format with "bedToBam -ubam" of BEDTools version 2.25.0 using the genome index of hs37d5 and sorted and index with SAMtools. Aggregated ATAC-seq reads were plotted with ngs.plot version 2.61 (Shen *et al.*, 2014) with the settings "-G hg19 -R bed -SC global -FL 3 -RR 100 -VLN 0 -L 80". Motif positions were previously enlarged to 20 bp (one fourth of parameter "-L") centered on the motif instance to ensure proportional plotting over complete range of x-axis. Sequence logos were plotted from HOMER supplied motif files using the R package seqLogo version 1.56.0 (Bembom & Ivanek, 2020).

5.3 Transcription factor motif enrichment analysis

An assortment of the ATAC-seq consensus genomic regions (e.g., regions with gained accessibility) of a certain sample set were selected for regions with their peak center within the indicated chosen genomic annotation of interest (e.g., ChromHMM states 1, 8, 9, 11) with the function "intersectBed -f 0.5 -u" of BEDTools version 2.25.0 (Quinlan & Hall, 2010). Except for class A-D regions (Results chapter 3.2.3) which were as a subset of consensus ATAC-seq peaks directly used for motif enrichment analysis. Transcription factor motif enrichment was determined in the chosen regions with the script "findMotifsGenome.pl" of the HOMER utility suite version 4.9 (Heinz *et al.*, 2010) with the options "hg19 -size given -mask". Appropriate background files were supplied (function parameter "-bg") using the corresponding ATAC-seq consensus peak set from which regions were originally selected. If the analysis was limited to regions overlapping with a certain genome annotation, background peak regions were also restricted to the subset overlapping with the annotation. Input BED files were previously reformatted to the HOMER recommended input BED format by extracting the first three columns containing chromosome, start and end positions of the peaks. Then a fourth column containing a unique numbering of the peaks was added, a fifth empty column and a sixth column with the strand information "+". HOMER's "known motifs" were used for the analysis. "Enrichment" was calculated as ratio between the fraction of target sequences to background sequences with the motif provided by HOMER. Results were plotted using the R packages ggplot2 version 3.3.2 (Wickham, 2016) and ggrepel version 0.8.2. (Slowikowski, 2020).

5.4 RNA-seq data preprocessing

RNA-seq reads originating from ribosomal RNA were removed with the tool SortMeRNA version 2.1 (Kopylova *et al.*, 2012) using the settings "--fastx --ref silva-euk-18s-id95.fasta,silva-euk-18s-db:silva-euk-28s-id98.fasta,silva-euk-28s:rfam-5.8s-database-id98.fasta,rfam-5.8s-db". Non-ribosomal reads were aligned to the reference genome hs37d5 using STAR version 2.5.3a (Dobin *et al.*, 2013). STAR was used with the settings "--runMode alignReads --runDirPerm All_RWX --sjdbOverhang 49". The folder with the previously calculated indices (see Materials and Methods chapter 5.1.2) and the Ensembl genome annotation version 87 were provided to the function arguments "--genomeDir" and "--sjdbGTFfile", respectively. Further settings were "--outSAMtype BAM SortedByCoordinate --outWigType bedGraph --outSAMstrandField intronMotif --quantMode TranscriptomeSAM GeneCounts --outFilterMultimapNmax 20" which generated several output files. This included a raw count table ("ReadsPerGene.out.tab"), which was used for differential gene expression analysis with DESeq2. A further STAR output file containing alignments in transcriptome coordinates ("Aligned.toTranscriptome.out.bam") was used for calculating normalized gene expression values (TPM and FPKM) with RSEM version 1.3.0. The script "rsem-calculate-expression" was used specifying the bam-format of the input file ("--bam") and the setting "--forward-prob 0" as the reads generated with the employed stranded RNA sequencing library protocols equal the antisense sequence of the original RNAs. Furthermore, the STAR output file containing alignments in genomic coordinates ("Aligned.sorted.ByCoord.out.bam") was used for quality control by generation of gene body coverage plots and a table for the distribution of mapped reads to gene features. Those quality measures were generated with the RSeQC version 2.6.6 (Wang *et al.*, 2012) scripts "geneBody_coverage.py" and "read_distribution.py", respectively, using python version 2.7.12 and the previously downloaded annotation (see Material and Methods chapter 5.1.1).

5.5 ChIP-seq data preprocessing

Preprocessing of single-end sequenced ChIP-seq data was conducted with an automated analysis pipeline consisting of bash scripts written by Dr. Markus Muckenhuber and Dr. Nick Kepper (Division chromatin networks, German Cancer Research Center (DKFZ), Heidelberg, Germany). Sequencing reads were aligned to the genome using Bowtie version 1.1.2 (Langmead *et al.*, 2009) with the setting "--best --strata -v 2 -m 1 --sam" to only allow for uniquely mapped reads with a maximum of two mismatches. SAM-formatted output files were converted to the binary BAM-format using "samtools view -b -F 4" of Samtools version 1.3 before reads were sorted and indexed with "samtools sort" and "samtools index". Duplicates were removed using "samtools rmdup -s". Peak calling for H3K27ac and H3K4me1 was conducted with "macs2 callpeak" version 2.1.2 with python version 2.7 using H3 as

control and following options "--gsize hs --bw 200 --mfold 5 50 --pvalue 1e-5 --format BAM --tsize 51 -B ". BAM-formatted read files were converted to BED format using "bamToBed" of BEDTools version 2.25.0. FRiP-scores, normalized strand cross-correlation coefficient (NSC) and relative strand cross-correlation coefficient (RSC) values were calculated for quality assessment. NCS and RSC values were calculated using the script "run_spp.R" (see Materials and Methods chapter 1.6) of the phantompeakqualtools by Kundaje (Kundaje, 2013).

5.6 scRNA-seq data analysis

Preprocessing of sequencing reads was conducted with the CellRanger software version 3.1.0 (CLL) or version 2.0.2 (MM) of 10x Genomics using the human genome build hg19 version 3.0.0 (CLL) or version 1.2.0 (MM) provided by 10x Genomics. All downstream analysis was conducted with the Seurat package version 3.1.0 (Stuart *et al.*, 2019). Cells of low quality with either less than 500 detected gens, less than 1500 UMI counts or more than 10 % mitochondrial reads were removed. The resulting numbers of filtered cells are listed in Table 19. Numbers of single cells after filtering used for the scRNA-seq analysis of CLL patient samples. are listed in Table 19-Table 21.

Table 19. Numbers of single cells after filtering used for the scRNA-seq analysis of CLL patient samples.

CLL 24 (before treatment)	CLL 24 (1 month after treatment start)	CLL 24 (remission)	CLL 24 (early relapse)	CLL 24 (relapse)
4,850 cells	1,065 cells	365 cells	3,033 cells	1,938 cells

Table 20. Numbers of single cells after filtering used for the scRNA-seq analysis of MM patient.

MM 1 DMSO	MM 1 Panobinostat	MM 2 DMSO	MM 2 Panobinostat	MM 3 DMSO	MM 3 Panobinostat	MM 4 DMSO	MM 4 Panobinostat
10,871 cells	12,245 cells	5,656 cells	5,256 cells	5,135 cells	3,385 cells	7,468 cells	6,526 cells

Table 21. Numbers of single cells after filtering used for the scRNA-seq analysis of MM cell lines.

KMS-11 DMSO	KMS-11 Panobinostat	MM1.S DMSO	MM1.S Panobinostat	U266 DMSO	U266 Panobinostat
1,485 cells	1,796 cells	2,503 cells	3,166 cells	2,490 cells	2,411 cells

The count data of all timepoints was combined. Counts were normalized and the 3,000 (default) most variable genes identified using the function "SCTransform" (Hafemeister & Satija, 2019). Cell cycle bias was regressed out using the functions CellCycleScoring and the argument "vars.to.regress" of the "SCTransform" function according to the standard workflow. Dimensionality reduction was conducted using the identified most variable genes by principal component analysis. A shared nearest neighbor

graph was constructed with the function “FindNeighbors” using the first 18 (CLL) or 10 (MM) principal components as determined from elbow plots. Cells were clustered using the function “FindClusters” with default parameters and “resolution=0.08” (CLL), “resolution=0.1” (MM cell lines) or “resolution=0.14” (MM patients). Two dimensional representations were calculated using uniform manifold approximation and projection (UMAP) (Becht *et al.*, 2018). Markers of clusters were determined with the function “FindAllMarkers” using the wilcoxon rank sum test (p-value < 0.01, absolute log2 foldchange > 0.25) and the parameter “min.pct = 0.25”. The heatmap visualization of marker genes was done with the “DoHeatmap” function. Cell cycle and senescence scores were calculated using the function “AddModuleScore”. Differentially expressed genes between panobinostat treated and untreated MM cell line and patient cells were determined individually for each cell line and patient with the function “FindMarkers” using an adjusted p-value cutoff of 0.05.

5.7 Targeted scDNA-seq data analysis

Single cell DNA-sequencing data was analyzed using the Tapestri pipeline version v1.10b5 of Mission Bio. The pipeline first removed adaptor sequences from reads using Cutadapt (Martin, 2011). Reads were aligned to the hg19 reference genome with the BWA-MEM tool (Li, 2013). Cells were identified after barcode correction and genotyped using the Genome Analysis Toolkit (GATK) (McKenna *et al.*, 2010). The final loom output files were used for the analysis with the software Tapestri Insights provided by Mission Bio. Further filtering criteria for inclusion of data were genotype quality ≥ 80, read depth ≥ 10, alternate allele frequency ≥ 20, variants called in ≥ 50% of cells, cells with ≥ 50% of genotypes present, variants mutated in ≥ 0.5% of cells. This resulted in a total of 11,336 and 12,150 filtered cells and 119 and 187 filtered variants for CLL patient 24 and 25, respectively. Mutations were manually selected based on their change during the course of treatment. FISH-plot was generated using Tapestri Insights and modified using Affinity designer. Aggregated variant allele frequencies (VAF) were obtained from Tapestri Insights (“VAF by read count”), which aggregates all reads from all cells to calculate the percentage of variant reads of all reads. This is comparable to VAFs calculated from bulk sequencing data and was plotted using the R package ggplot2.

5.8 Gaussian fits

The sum of several Gaussian functions each of the following form were used to fit distributions:

a*exp(-((x-pos)/width)^2)

Parameters were estimated with the R function “nls” with setting “alg="port", ctrl = list(maxiter = 500000, tol = 1e-05, minFactor = 1/1024)”. The start, lower boundary and upper boundary values used for every application are indicated separately.

5.9 Identification of differential ATAC-seq regions in CLL patients compared to CD19$^+$ B cells

5.9.1 Differential accessibility analysis in consensus peaks

Shifted reads, which represent Tn5 transposase occupancy positions, were used for the differential accessibility analysis. Read files in BED format were converted by the BEDTools function “bedtobam -ubam” to BAM format. A consensus peak set was generated from all technical replicates (separate ATAC-seq library preparation) of 19 CLL patients and 3 healthy controls used in the classical approach which compares CLL patients to CD19+ B-cells (Figure 19 A) with the R package DiffBind (Ross-Innes *et al.*, 2012). Only peaks which were present in at least four technical replicates were selected for the consensus peak set by setting the option “minOverlap=4” in the DiffBind-function “dba.count”. Fold change values per regions were not taken from DiffBind, but calculated as follows to account for the presence of technical replicates. If technical replicates existed, counts of replicates were summed up and divided by the number of technical replicates of a specific sample. The average of all CLL patient counts per region was divided by the average count of that region of all biological replicates of the non-malignant controls. This ratio was logarithmized to base two to receive log2 fold changes.

Upper and lower log2 fold change thresholds for the determination of regions with gained or lost accessibility in CLL compared to CD19$^+$ B-cells were determined data-driven by fitting the sum of three Gaussian functions (see Materials and Methods chapter 5.8) to the trimodal log2 fold change distribution. The following start, lower boundary and upper boundary values were used (Table 22).

Table 22. Settings for Gaussian fitting to determine regions with differential accessibility in CLL compared to CD19$^+$ B cells of non-malignant donors.

	Fit 1 (unchanged)			Fit 3 (gain)			Fit 2 (loss)		
	a	width	position	a	width	position	a	width	position
INITIAL	500	0.5	0	200	1	3	150	1	-3
UPPER BOUNDARY	0	0	-1	500	3	12	150	3	-2
LOWER BOUNDARY	800	2	1	500	3	12	150	3	-2

The upper and lower thresholds were then set to the x-value of the intersection point between fit 1 and fit 2 or fit 3 with lower=-1.5 and upper=1.8 (Figure 23 B left). The threshold for the FDR value provided by DiffBind was defined as the inflexion point of the function describing the number of differentially accessible sites in dependence of the negative logarithm (base 10) of the chosen FDR threshold. Thus, the FDR threshold was determined to be ≤ 0.00759 (Figure 23 B right).

5.9.2 Correlation analysis

The correlation analysis was conducted with DiffBind normalized counts in all DiffBind consensus peak regions (n=121,227) which were averaged over technical replicates and logarithmized to base 2. Hierarchical clustering (single linkage) based on Spearman's correlation coefficient of those counts was performed with the hcluster function of the R package amap version 0.8-18 (Lucas, 2019). The visualization was conducted with the "as.dendrogram"-function and the R package dendextend (version 1.13.4) function "colored_bars" (Galili, 2015).

5.9.3 ChromHMM annotation of differential ATAC-seq peaks in CLL patients compared to CD19+ B cells

The consensus ChromHMM chromatin segmentation of the individual non-overlapping 12 functional states was taken from Mallm et al. (Mallm *et al.*, 2019). This annotation had been generated from ChIP-seq experiments of seven functionally relevant histone marks (H3K9ac, H3K4me3, H3K27ac, H3K4me1, H3K36me3, H3K27me3, H3K9me3) as a consensus of CLL patients and non-malignant CD19$^+$ B-cells controls which were selected from the same cohort as the ones used for the ATAC-seq analysis (some patients or control samples were only included in either the ChromHMM or ATAC-seq analysis due to limitations in the available amounts of required material; see Mallm et al., 2019 Appendix Figure S1). Regions with differential accessibility between CLL samples an non-malignant controls were annotated by calculating the fractional overlap of the regions (BED format) and each individual state using the BEDTools utility suite (Quinlan & Hall, 2010) function "annotateBed" version 2.25.0. A region was assigned to the ChromHMM state for which it showed the highest fractional

overlap. In case of identical fractional overlaps with two states, the state with the smaller state number was selected.

5.9.4 Annotation of active enhancers in CLL patients and non-malignant CD19$^+$ B-cells

Active enhancers defined as ATAC-seq peaks or regions showing bidirectional transcription, which are comprised or surrounded (center +/- 1 kb) by "potentially active enhancer containing genomic regions" (ChromHMM states: 1, 9), were analyzed in CLL compared to non-malignant CD19$^+$ B-cells. Thus defined active enhancer regions were first identified individually for each patient and non-malignant control. To this end, high confidence peak regions (ATAC-seq or bidirectionally transcribed loci) were selected for every individual CLL patient or control by identifying genomic regions detected as ATAC-seq peak or region with bidirectional transcription in all technical replicates. Regions overlapping any annotated RefSeq TSS (+/- 1 kb) were excluded using the BEDTools function "intersectBed -v". ATAC-seq and bidirectional transcription signal regions were extended from the region center by +/- 1 kb using "slopBed -b 1000" to capture the peak surrounding region where post-translationally modified histones linked to active enhancers (H3K27ac and K3K4me1) are expected according to Chen et al. (Chen *et al.*, 2018). Extended peaks were intersected with the ChromHMM states 1 and 9, which are predictive for intergenic and genic active enhancers ("potential active enhancer containing loci"). For this, the corresponding ChromHMM annotation determined individually for the same patient or control from their histone modification patterns was employed. Overlapping regions were merged. Then, consensus active enhancer regions were assembled separately for CLL and non-malignant B-cells based on the presence of a region in at least three out of 19 CLL patients and one (ATAC-seq peak) or two (loci with bidirectional transcription) CD19$^+$ B-cell pools of non-malignant donors. The overlap of the resulting four classes of active enhancers (active enhancers in CLL and non-malignant B-cells defined by either chromatin accessibility or bidirectional transcription) was calculated by counting the "potential enhancer containing loci" (see Material and Methods chapter 5.2.3), which overlap with either of the classes.

5.10 Phylogenetic analysis of non-malignant B-cell reference subpopulations by ATAC-seq

5.10.1 Isolation of non-malignant B-cell reference subpopulations

Non-malignant B-cell reference subpopulations were isolated from two to three healthy donors by the group of PD Dr. Marc Seifert at the Institute of Cell Biology (Cancer Research), University of Duisburg-Essen, Essen, Germany. Collected subpopulations were *naïve B cells* (IgD^+ $CD23^+$ $CD27^-$ $CD5^-$), *IgM^+ memory B cell* (IgM^+ IgD^+ $CD27^+$ $CD23^-$), *$CD5^+$ mature B cells* (IgD^+ $CD5^+$ $CD23^+$ $CD27^-$ $CD38^{low}$), *$CD5^+$ memory B cells* ($CD27^+$ $CD5^+$ $CD23^-$ $CD38^{low}$), *IgG^+ memory B cell* (IgG^+ $CD27^+$ $CD23^-$) from peripheral blood and *spleen marginal zone B cells* ($CD21^{high}$ $CD27^+$ IgM^+ $CD23^-$) from spleen tissue.

5.10.2 Identification of regions with dynamic ATAC-seq signal in non-malignant B-cell reference subpopulations

Only ATAC-seq libraries with a FRiP-score above 6.5 % were used for further analysis. For each sorted B-cell subpopulation ATAC-seq peak regions that were found in at least two individuals were kept using the “genomecov” and “merge”-functions of the BEDTools utility suite (Quinlan & Hall, 2010). A consensus peak list containing all peak regions of all subpopulations was then generated from these subpopulation specific lists using the “merge”-function of the BEDTools utility suite (Quinlan & Hall, 2010). ATAC-seq raw counts per sample were calculated for peak regions in the consensus peak list with BEDTools “annotate”-function with the parameter “-counts” (Quinlan & Hall, 2010). Raw counts were used to identify 14,377 regions with differential accessibility between the subpopulations by DESeq2 1.24.0 (Love *et al.*, 2014) using the likelihood ratio test with the design formula “design=~subpopulation” and the reduced model “reduced = ~ 1”.

5.10.3 Clustering

Normalized counts of biological replicates with FRiP-scores above 6.5 % were averaged for clustering. Counts of a certain region are divided by the mean count over all samples to account for differences in average count levels between regions. A pseudo count of 0.5 was added before counts were natural log-transformed. A distance measure based on cosine correlation was used to cluster regions according to their dynamics and implemented with the following R function:

```
cosDist = function(A){
        as.dist(1-(A%*%t(A) / (sqrt(rowSums(A^2) %*% t(rowSums(A^2))))))
}.
```

Hierarchical clustering using those distances were calculated with the R function "hclust" using complete linkage and results were plotted with the "heatmap.2" function of "gplots (Warnes *et al.*, 2020).

5.10.4 KEGG pathway analysis

Regions of interest were loaded from a bed file with the "readPeakFile"-function" of the R package ChIPseeker (Yu *et al.*, 2015) version 01.24.0 or from a comma separated (csv) file with the function "read.delim". Nearest genes were determined with the function "seq2gene" with the annotation "TxDb.Hsapiens.UCSC.hg19.knownGene" (version 3.2.2) (Carlson, 2015) and the settings "tssRegion = c(-1000, 1000), flankDistance = 3000". The enrichment of genes of certain KEGG pathways (Kanehisa & Goto, 2000) in a gene set was then determined with the function "gost" of the R package gprofiler2 version 0.1.9 with default settings (Kolberg & Raudvere, 2020).

5.10.5 Consensus peak list

A consensus ATAC-seq peak list, which contains additionally to the 14,377 developmentally dynamic accessible regions all sites which were found in at least one CLL or healthy control sample (see Figure 19), was generated using different BEDTools utility suite functions of version 2.25.0 (Quinlan & Hall, 2010). If technical replicates (library preparation from individual aliquots) existed for a sample, only genomic locations covered by a peak in both replicates were kept for that sample using the function "bedtools intersect" to find reproducible peak regions. Regions from each sample were combined ("cat"), sorted ("sort -k1,1 -k2,2n") and merged if overlapping ("bedtools merge"). Before the 14,377 differentially accessible sites between the B-cell subpopulations were added, all regions intersecting with them were removed ("intersectBed -v") to preserve their original peak location and extension. Only regions larger than 19bp were kept ("awk"). The final consensus peak set with all potentially interesting sites consists of 315,747 regions with a median size of 177 bp.

5.10.6 Normalization of ATAC-seq count matrix

Raw ATAC-seq read counts in regions of the consensus peak set were calculated with "bedtools annotate -counts" using bed-files with reads shifted to the transposase occupancy position (size: 29bp) for all CLL samples and non-malignant controls. A count matrix containing raw counts of all samples was generated. Raw reads of each region were divided by the corresponding region length in kilo base pairs. To account for differences in sequencing depth, counts of a sample were then divided by the total counts of that sample and multiplied by one million, followed by a quantile normalization over all samples with the function "normalize.quantiles" of the R package preprocessCore version 1.46.0 (Bolstad, 2020). The first part of normalization is identical to TPM (transcripts per million) calculation

for RNA-seq data. A similar approach for normalization of ATAC-seq data had been used by Corces et al. (Corces *et al.*, 2018). Quantile normalization had also been used for ATAC-seq count normalization by Rendeiro et al. (Rendeiro *et al.*, 2020). The matrix of quantile normalized ATAC-seq counts per region length in kb and per million reads was further used for the analysis of phylogenetic relations and the cell-of-origin identification and also referred to as "(quantile) normalized accessibility". An additional log-transformation (base10; pseudocount:1) of normalized counts was also evaluated and yielded very similar results.

5.10.7 Phylogenetic tree construction from ATAC-seq profiles of non-malignant B-cell reference subpopulations

A pairwise Manhattan distance matrix between all sorted B-cell subpopulation samples was calculated based on quantile normalized accessibility in differentially accessible regions. The distance matrix was used to construct a phylogenetic tree with the balanced minimum-evolution algorithm by Desper & Gascules (Desper & Gascuel, 2002) implemented as function "fastme.bal" in the R package ape version 5.4-1 (Paradis & Schliep, 2019). The quality of the tree was assessed by determining the spearman correlation between the pairwise distances on the tree computed with the function "cophenetic" of the R base package stats and the original pairwise Manhattan distances.

5.10.8 Transcriptome-based phylogenetic tree construction of non-malignant B-cell reference subpopulations

For the tree based on RNA-seq data of the same subpopulations, differentially expressed genes between all six B-cell subpopulations were determined from raw counts with DESeq2 version 1.24.0 (Love *et al.*, 2014) with the likelihood ratio test (design formula: "design=~subpopulation"; formula of reduced model "reduced = ~ 1"). Tree construction was based on pairwise Manhattan distances between all biological replicates of the six analyzed B-cell subpopulations. Distances were computed from FPKM values of the identified 7671 differentially expressed genes during B-cell maturation. Phylogenetic relations were inferred with the function "fastme.bal" (R package ape 5.3), which is based on the balanced minimum-evolution approach (Desper & Gascuel, 2002).

5.10.9 Maturation score calculation for healthy B-cell subsets and linear modeling of accessibility dynamics per region

The maturation score for healthy B-cell subsets were calculated with adjusted scripts of the MethylCOOM framework (https://github.com/justannwska/Methyl-COOM) by Wierzbinska et al. (Wierzbinska *et al.*, 2020). The "naïve B2" and "IgG memory" samples of donor H18 were used to define the two end points of the main maturation trajectory. The function "nodepath" of the R package

ape version 5.4-1 (Paradis & Schliep, 2019) was then used to extract the order of the nodes between those two samples (= main developmental axis). The two tip nodes were removed and only internal nodes were kept. A phylobase tree object was created with the function "phylo4" of the R package phylobase (version 0.8.10) (Bolker *et al.*, 2020) and converted to a "data.frame" to extract the structure of the tree. For every sample the ancestor node on the main developmental axis was extracted. The edge length from the base node of "H18 naïve B2" to the base node of every individual sample was calculated by summing up all individual edge lengths. Edge length sums were transformed to values between zero ("H18 naïve B2") and one ("H18 IgG memory") to get a developmental stage for each sample as fraction of the degree of maturation from "naïve B2" to "IgG memory" cells (=maturation score). To model the accessibility dynamics for every region a linear model was fit with the R function "lm" using the maturation score as linear predictor for the accessibility (response).

5.11 Identification of cell-of-origin and CLL-specific chromatin accessibility

5.11.1 Cell-of-origin modeling by principal component regression

The cell-of-origin of a CLL patient was define as the closest virtual cell on the main physiological maturation trajectory. Principal component regression was employed to predict the maturation stage of the cell-of-origin of an individual CLL patient from the chromatin accessibility signature of that patient taking only regions into account, which change linearly during physiological development to avoid any bias. For this, a region accessibility matrix was generated from the normalized ATAC-seq count matrix by selecting only developmentally dynamic accessible sites with linear dynamic (linear regression fit: p-value < 0.5; R2 > 0.3). The application of multiple linear regression directly on the accessibility per site was not possible due to multicollinearity in the data, originating from the coregulation of regions. Therefore, principal component regression was selected, which first conducts a principal component analysis on the region accessibility matrix to then perform multiple linear regression using the linearly independent principal components as predictors for the maturation stage. Principal component regression analysis was conducted with the R package pls version version 2.7-2 (Mevik *et al.*, 2019). The principal component regression model was constructed with the function "pcr" using the accessibility profile of the healthy B-cell subpopulations as predictors and their previously calculated corresponding maturation scores as response variable. Input data was mean centered and scaled (divided by the corresponding sample standard deviation) using the settings "scale = TRUE" and "center = TRUE". Cross-validation was employed to select the optimal number of principal components to build the model. The performance of several models differing only in the number of included n first principal components were evaluated by cross-validation using the "pls" function setting "validation = "CV". The

model based on the first seven principal components showed the smallest root mean squared error of prediction (RMSEP), which was calculated with the function "RMSEP" and was thus selected. RMSEP for all 10 analyzed models was plotted with the function "validationplot". The cross-validation prediction of the maturation scores for the selected model was plotted against the given values using the function "predplot". Model predictions for CLL patients were calculated using the function "predict" from the base R package stats. Library preparation replicates of individual patients, which were first kept separate, could be assigned to close locations on the main trajectory of the tree. Therefore, the maturation score prediction of individual cells-of-origin was repeated with combined replicates if existing. For this, normalized counts of replicates were averaged using the base R function "rowMeans".

5.11.2 Addition of cell-of-origin estimates to static tree of healthy B-cell development

The predicted maturation scores of the cells-of-origin of individual CLL patients between zero and one were back-transformed to phylogenetic distances on the tree. The cells-of-origin estimates were visualized on the static tree of healthy B-cell development by adding them individually as tips at the estimated position using the function "bind.tree" of the package "ape" version 5.4-1 (Paradis & Schliep, 2019). After the addition of the cell-of-origin estimate of the first patient, the structure of the new tree was again extracted by creating a phylobase tree object with the function "phylo4" (R package phylobase version 0.8.10) (Bolker *et al.*, 2020) and converting it to a "data.frame" to extract the node path of the main trajectory of the new tree. Then the next cell-of-origin estimate was added to the updated tree with the "bind.tree" function. This was repeated until each cell-of-origin estimate of all patients had been added. For each developmentally dynamic region the difference between the accessibility value of a patient and its cell-of-origin (inferred by linear modeling) was squared and the differences summed over all regions. For visualization the length of the tip of each CLL patient was set to 30 times of this square root of the sum of squared distances.

5.11.3 Identification of regions with aberrant accessibility in CLL

Linear modeling of accessibility dynamic of non-malignant B-cell subsets for individual regions was repeated as described in Materials and Methods chapter 5.10.9 this time not only for the dynamic regions, but also for all other regions contained in the consensus peak set. Confidence intervals (CI) of accessibility prediction for individual cells-of-origin (confidence level: 95 %) were extracted from each linear model by providing the inferred maturation scores to the function "predict" of the R base package "stats" with the setting "interval = 'prediction' ". A CLL patient was considered to have aberrant chromatin accessibility in a specific region if its accessibility was outside of the 95 % CI of prediction for the accessibility of its corresponding cell-of-origin. Further, a region was considered to have CLL-specific aberrant accessibility if at least 75 % of CLL patients showed an aberrant accessibility

signal. Regions with CLL-specific aberrant accessibility belonging to the dynamically changing regions during normal development were split into two classes with gained (class A) or lost (class B) accessibility in CLL. Likewise, all other consensus regions with stable accessibility during development were separated into two classes with gained (class C) and lost (class D) accessibility in CLL. All calculations for the identification of regions with aberrant accessibility and their classification were done in R (version 4.0.2). Linear fits for certain regions were plotted with the "stat_smooth" function of ggplot2 version 3.3.2 (Wickham, 2016) using the default argument "se=TRUE, level = 0.95" to indicate the standard error of regression line (95 % confidence level). The visualization of the differential accessibility signal of all regions of a class between samples with boxplots was conducted with accessibility counts scaled by the mean count of a region over all samples. This was done to achieve similar accessibility levels between regions. Scaled counts were generated by dividing each sample count of individual regions by the mean count of that region over all samples.

5.12 Identification of differential ATAC-seq regions in ibrutinib-treated CLL patients

Regions exhibiting a change in chromatin accessibility upon treatment with ibrutinib in CLL were determined from the raw ATAC-seq counts of shifted reads in the previously determined consensus peak list (see Material and Methods chapter 5.10.5-5.10.6). Differentially accessible regions were identified from raw counts of the first three timepoints of the five CLL patients CLL24-28 (see Figure 19) by DESeq2 version 1.28.1 (Love *et al.*, 2014). The DESeq2 analysis was conducted using the likelihood ratio test with the design formula "design= ~ patient + timepoint" for the full model and "reduced = ~ patients" for the reduced model and an adjusted p-value threshold of 0.05.

5.13 Sequencing data analysis of panobinostat-treated MM

5.13.1 Identification of differential ATAC-seq regions in panobinostat treated compared to untreated MM cell lines and patients

The cell lines KMS-11, MM1.S and U266 were treated for 24 hours with 8 nM or 5 µM panobinostat before ATAC-seq libraries were prepared. Replicates for cell lines were generated by independent repetition of treatment. Cells of four MM patients were treated *ex vivo* with 8 nM panobinostat or DMSO as control by the group of Prof. Marc-Steffen Raab (Research Group Translational Myeloma, German Cancer Research Center (DKFZ), Heidelberg, Germany). ATAC-seq libraries were prepared in technical replicates. The identification of regions with differential ATAC-seq signal in panobinostat

treated compared to untreated cells was conducted using the R package DiffBind version 2.4.5 (Ross-Innes *et al.*, 2012) with R version 3.4.3. Read counts were generated from shifted reads representing the transposase binding event with the DiffBind function "dba.count". Regions present in at least four replicates ("minOverlap=4") or two replicates ("minOverlap=2") for global changes present in all cell lines and patients or changes only present in individual cell lines, respectively, were considered for the analysis.

5.13.2 Clustering heatmap of chromatin accessibility in MM

Consensus ATAC-seq peak regions (n=195**,**368) present in at least four replicates of all cell lines and patients were employed. Normalized ATAC-seq counts in consensus peak regions, which were generated by DiffBind (scaled by size factors internally calculated with DESeq2), were extracted by the function "dba.report". The annotated heatmap of all accessible regions was generated with the R package pheatmap version 1.0.12. Normalized ATAC-seq counts were logarithmized (base 10) for clustering. Regions with similar counts were aggregated to 300 clusters using k-means clustering by setting the pheatmap-function option "kmeans_k=300". As distance measure for clustering of regions and samples "maximum" and "correlation" (Pearson correlation) was used, respectively, with complete linkage.

5.13.3 Analysis of genomic H3K27ac distribution in panobinostat-treated MM cells

5.13.3.1. Calculation of correlation functions

Autocorrelation functions were calculated for the evaluation of the genomic distribution of H3K27ac with the software MCORE (Molitor *et al.*, 2017) by executing "java -Xms64m -Xmx4096m -jar MCORE.jar" using java version 1.8.0_191. The individual H3K27ac read file was provided in BED-format as "Sample1" and "Sample2", while the corresponding H3 reads were provided as control ("Input1" and "Input2"). Correlation functions were calculated for chromosome 1 using an extension size of 100. Calculated correlations were normalized to the correlation at zero shift distance.

5.13.3.2. H3K27ac ChIP-seq read counting in gene bodies

H3K27ac reads in gene bodies defined according to the Ensembl genome annotation version 87 (in BED format) were counted using the BEDTools function "intersectBed".

5.13.3.3. Identification of expressed genes in MM cell lines

Counts of technical replicates were combined for the discrimination between expressed and not expressed genes. The sum of two Gaussian functions (see Materials and Methods chapter 5.8) was used to fit the log2 of TPM values. The following initial, upper boundary and lower boundary values were used for the two distributions (Table 23).

Table 23. Settings for Gaussian fitting to determine expressed genes in MM cell lines.

	Fit 1 (not expressed)			Fit 2 (expressed)		
	a	width	position	a	width	position
INITIAL	250	2	-4	20	1	4
UPPER BOUNDARY	1,000	7	0	1,000	7	12
LOWER BOUNDARY	0	0	-10	0	0	0

The x-value of the intersection point of the two distributions was determined as expression threshold for every condition and cell line separately. A gene was defined as being expressed in a cell line, if its expression in both treated (5 µM panobinostat) and untreated cells was above the corresponding threshold. Likewise, a gene was assigned to the group of not expressed genes, if its expression was below the corresponding threshold in both treated and untreated condition of the respective cell line.

5.13.3.4. Average read coverage over a set of genes in MM cell lines

The analysis of H3K27ac distribution over gene bodies of expressed and not expressed genes was conducted using the previously defined groups of expressed and not expressed genes (see Materials and Methods chapter 5.13.3.3). Read coverage for H3K27ac, H3 and ATAC-seq at genebodies was determined with ngs.plot version 2.61 (Shen *et al.*, 2014) with the settings "-G hg19 -R genebody -L 20000 -D refseq" and a fragment length of 50 for ChIP-seq and 3 for ATAC-seq. Combinations of files containing reads and gene names were provided as input using a configuration file. Indexed BAM-formatted reads of H3K27ac and H3 were directly used as input. For ATAC-seq, reads from all technical replicates were combined. ATAC-seq reads in BED format were collapsed to 3 bp centered on the transposase binding event. BED files were converted to BAM format using "bedToBam -ubam" of BEDTools version 2.25.0 (genome index: hs37d5) and sorted and index with SAMtools. Read distribution calculated by ngs.plot was plotted using ggplot2.

5.13.4 Differential gene expression analysis in MM cell lines

Raw read counts per gene generated by STAR (see Materials and Methods chapter 5.4) were used to determine differentially expressed genes between treated and untreated cells for each cell line separately using DESeq2 version 1.18.1 with R version 3.5.1. The effect of treatment individually for each cell

lines was analyzed by combining the cell line name and the treatment condition to a grouping factor and using it as input for the design formula ("design= ~ group"). Differentially expressed genes were extracted based on an adjusted p-value threshold of 0.05. Calculated fold changes were visualized with the function "plotMA" using shrunken fold changes calculated with the function "lfcShrink" with the option "type=normal" and used for coloring the KEGG pathway "hsa04218" (Cellular senescence).

5.13.5 Senescence score calculation from scRNA-seq data

Early senescence marker genes were taken from Zirkel et al. (Zirkel *et al.*, 2018). The overlap of early senescence marker genes was determined separately for up and downregulated genes for the three available primary human cell types umbilical vein endothelial cells (HUVECs), mesenchymal stromal cells (MSCs) and fetal lung fibroblasts (IMR90s). Only marker genes detected in all three cell types were used for senescence score calculation. ENSEMBL gene names were converted to gene symbols using the R package org.Hs.eg.db version 3.8.2. Two expression scores were calculated for every cell: one from common up and one from common downregulated genes. Scores were calculated by using the Seurat function "AddModuleScore" from the original count data of genes present in both data sets. Kmeans clustering based on the two calculated scores to identify senescent cells was performed using the base R function "kmeans" with two centroids.Senescence scores wer plotted with the function "heatscatter" of the R package LSD version 4.1-0.

5.14 Data visualization

5.14.1 Coverage tracks of sequencing-based readouts

All visualizations of coverage tracks at specific genomic loci of sequencing-based genome-wide data were generated with the Integrated genomics viewer (IGV) version 2.6.2 (Robinson *et al.*, 2011). Gene annotations are based on RefSeq. Coverage tracks of ATAC-seq data always represent actual Tn5 transposase occupancy as shifted reads were used. Reads in BAM format were sorted, indexed (SAMtools version 1.3) and converted to normalized bigwig files with python version 2.7.12 and "bam2wig.py" of the RSeQC package version 2.6.6 (Wang *et al.*, 2012) with the setting "-t 100000000", which internally calls "wigToBigWig" of the UCSC user apps (Kent *et al.*, 2010). Coverage tracks of ChIP-seq data were generated by creating bedgraph-files with "genomeCoverageBed -bga" of the BEDTools utility suite version 2.25.0 using the BEDTools provided file "human.hg19.genome" (see Materials and Methods chapter 1.6) as input for the option "-g". Bedgraph files were then tiled with IGV tools version 2.3.23 using the function "java - -jar igvtools.jar toTDF hg19" with java version 1.8.0_191.

5.14.2 Illustrations of sequencing library quality metrics

Sample overview images with quality metrics were generated with "gplots::heatmap.2" function version 3.0.4. (Warnes *et al.*, 2020) or its extension "heatmap.3" for heatmaps with more than one column color bar (https://raw.githubusercontent.com/obigriffith/biostar-tutorials/master/Heatmaps/heatmap.3.R, 18.11.2020) using the R packages scales version 1.1.1 and devtools version 2.3.2. with R version 4.0.2. Basic images were further annotated in Affinity Designer or Adobe Illustrator.

5.14.3 Colored KEGG pathways

The R package pathview (Luo & Brouwer, 2013) was used for the coloring of KEGG pathways (Kanehisa & Goto, 2000) with the species argument "hsa". Necessary gene ID conversions were conducted with the R package org.Hs.eg.db version 3.11.4 (Carlson, 2020).

5.14.4 Figures

For figure panel arrangement, the R package ggpubr version 0.4.0.999 was used. 3D figure was generated with gg3D R package (Acker, 2020). The R package viridis version 0.5.1 (Garnier, 2018) was used throughout the thesis for coloring figures.

Results

1 Development of a workflow to map genome-wide chromatin accessibility in primary human cells

Chromatin accessibility measured by the assay for transposase-accessible chromatin using sequencing (ATAC-seq) provides important information about the activity status of genomic regulatory elements and enables to relate them to transcription factor (TF) occupation (Buenrostro *et al.*, 2013; Yan *et al.*, 2020). ATAC-seq was developed in 2013 by Buenrostro *et al.* (Buenrostro *et al.*, 2013). The benefits of this condensed readout of the chromatin state are its compatibility with small cell numbers (Buenrostro *et al.*, 2013) and its suitability for primary human patient samples. The protocol was adjusted for different frozen primary patient material, which included cells frozen after formaldehyde fixation (e.g., samples Figure 19 A) (Figure 11). The cell numbers available from patients or non-malignant donors can vary. Especially during effective treatment or after sorting of small cell subpopulations very low cell numbers had to be expected. Thus, a protocol optimized for limited input material was established by reducing the number of washing steps. The library quality was assessed before sequencing with a quantitative real time polymerase chain reaction (qPCR) step (Figure 11).

ATAC-seq shares several data analysis steps with other sequencing types but also differs in some essential steps. Bioinformatic tools specifically designed for the analysis of ATAC-seq data have not been available at the start of the project and are still rather rare. Tools designed for other sequencing data types with similar tasks (e.g., peak calling for ChIP-seq data) can however be adapted for the analysis of ATAC-seq data. I designed and established an analysis pipeline by adjusting workflows existing for other sequencing-based readouts for ATAC-seq data (Figure 11). This data analysis pipeline was used for the analysis of ATAC-seq data described in this thesis and in the two independent studies by Muckenhuber *et al.* (Muckenhuber *et al.,* submitted) and Wierzbinska *et al.* (Wierzbinska *et al.*, 2020). In addition, it was applied in a project on T-cell exhaustion in mice in collaboration with the group of Dr. Martina Seiffert (CLL tumor microenvironment, German Cancer Research Center (DKFZ), Heidelberg, Germany), which shows the applicability in different contexts (Hanna *et al.* , in revision).

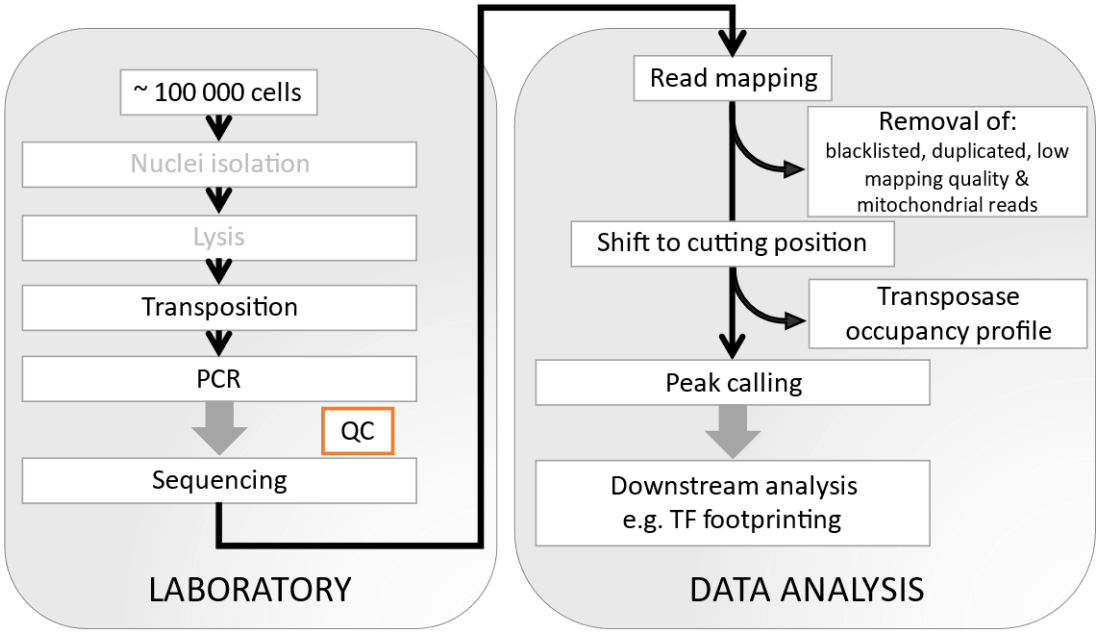


Figure 11 Illustration of ATAC-seq workflow.
The key experimental (left) and bioinformatic (right) workflow steps are illustrated. Steps only present in protocol variants are marked in grey. A quality control step (QC) by qPCR was integrated to assess the ATAC-seq sequencing library quality before sequencing. The ATAC-seq data preprocessing including all steps till peak calling were compiled into an analysis pipeline enabling further specialized downstream analyses. TF: transcription factor.

1.1 ATAC-seq sequencing library generation

1.1.1 ATAC-seq protocols for formaldehyde-fixed cells and samples with low cell numbers were established

ATAC-seq was conducted for both human primary cells and cell lines. If possible, cells were freshly processed (cell lines and sorted non-malignant B-cell subpopulations). Otherwise, cells were conserved by freezing formaldehyde-fixed or viable cells (patient cells). Nuclei from 100,000 to 500,000 cells were isolated, lysed and subjected to a transposition reaction similar as described in the original published protocol for a lymphoblastoid cell line (Buenrostro *et al.*, 2013). Due to the small amount of material available for patient samples during treatment and for sorted cell subpopulations from non-malignant donors (see Figure 19 B), washing steps were reduced ("low input protocol", only for viable frozen cells). This was accomplished by omitting the nuclei preparation step and integration of cell lysis with NP40 into the transposition reaction.

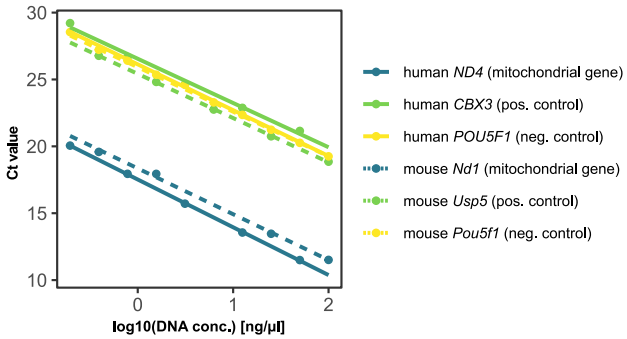
1.1.2 ATAC-seq sequencing libraries quality can be assessed by qPCR

It is important to judge the quality of the final PCR-amplified and cleaned ATAC-seq libraries before next generation sequencing. Accordingly, a quality control step was routinely used (for more detailed description see Materials and Methods chapter 5.1.3). The quality of an ATAC-seq library was determined by the enrichment of signal in accessible genomic regions compared to the genome-wide background signal. Already slight contamination with mitochondrial DNA leads to a high number of reads mapping to the mitochondrial genome due to its general high accessibility. Mitochondrial reads contain no information about local accessibility and the reads can be discarded during data analysis to receive high quality data. Consequently, high percentages of mitochondrial reads reduce the effective read depth (Yan *et al.*, 2020). To evaluate the quality of an ATAC-seq library before sequencing, we used a quantitative real time polymerase chain reaction (qPCR) step. Three primer pairs were used to assess both the signal-to-noise ratio and the number of mitochondrial reads. For human DNA, one primer pair targeted a small region near the mitochondrial genome gene *ND4* (136 bp fragment). The other two primer pairs were targeted to a small region in the *CBX3* promoter (96 bp fragment; positive control), which is generally accessible, and an intronic region in the *POU5F1* gene (98 bp fragment; negative control), where no ATAC-seq signal is expected (Figure 12 A). For ATAC-seq libraries generated from mice DNA, primer pairs targeting the *Nd1* gene (mitochondrial gene), *Usp5* promoter (positive control) and *Pou5f1* intron (negative control) were selected (Figure 12 A). I evaluated the efficiency of the primer pairs from the slope of standard curves (Figure 12 A-B). Only primers with an efficiency of at least 90 % were selected. Although the correlation between qPCR estimates and sequencing results of the signal-to-noise ratio was moderate (Figure 12 C), the method allowed to identify libraries of bad quality and to compare different protocols.

A

	Primer	DNA conc. [ng/µl]	log10(DNA conc. [ng/µl])	C_t value	Slope	Primer efficiency
Human	*ND4* (mitochondrial gene)	0,195	-0,71	20,06	-3,572	91%
		0,781	-0,11	17,94		
		3,125	0,49	15,71		
		12,500	1,10	13,56		
		50,000	1,70	11,50		
	CBX3 (pos. control)	0,195	-0,71	29,21	-3,309	101%
		0,781	-0,11	26,67		
		3,125	0,49	24,64		
		12,500	1,10	22,88		
		50,000	1,70	21,15		
	POU5F1 (neg. control)	0,195	-0,71	28,55	-3,426	96%
		0,781	-0,11	26,41		
		3,125	0,49	24,60		
		12,500	1,10	22,35		
		50,000	1,70	20,26		
Mouse	*Nd1* (mitochondrial gene)	0,391	-0,41	19,58	-3,4258	96%
		1,563	0,19	17,94		
		25,000	1,40	13,46		
		100,000	2,00	11,51		
	Usp5 (pos. Control)	0,391	-0,41	26,77	-3,3052	101%
		1,563	0,19	24,82		
		6,250	0,80	22,75		
		25,000	1,40	20,75		
		100,000	2,00	18,85		
	Pou5f1 (neg. control)	0,391	-0,41	27,22	-3,3308	100%
		1,563	0,19	25,36		
		6,250	0,80	23,28		
		25,000	1,40	21,24		
		100,000	2,00	19,25		

B



C

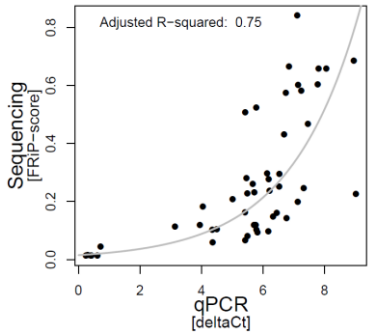


Figure 12 Establishment of pre-sequencing quality control qPCR for the accessment of signal-to-noise ratio and degree of mitochondrial DNA contamination for ATAC-seq libraries.
A Primer pair efficiency determined from the slope of **B** calibration curves generated using isolated human (HeLa cells) or mouse (mESC cells) genomic DNA. The primer pair efficiency was calculated as 10^(-1/(slope))-1. 100 % primer efficiency is equivalent to a doubling in every cycle. **C** Correlation of qPCR (difference between cycle threshold (C_t) values of positive and negative ATAC-seq control primer pairs) and sequencing (fraction of reads in peaks) results for the enrichment of signal over background using three data sets containing ATAC-seq libraries of mice and human of varying quality. Line represents fit with an exponential model. FRiP: Fraction of reads in peaks.

1.2 ATAC-seq data analysis pipeline with quality controls

1.2.1 An analysis pipeline for the preprocessing of ATAC-seq reads is set up

The sequenced reads were analyzed with a pipeline of tools adjusted for the analysis of ATAC-seq data. After each step of the ATAC-seq data analysis pipeline, the quality was controlled by summarizing the quality reports by the tool FastQC and other statistics in a single report for all samples of a project by the tool MultiQC version 1.7 (Ewels *et al.*, 2016). Reads were trimmed to remove adaptor sequences and bases of minor sequencing quality at the end of the reads with the tool Trimmomatic version 0.36 (Bolger *et al.*, 2014). Otherwise, reads cannot be correctly aligned to the genome. This is especially important for ATAC-seq data as an essential amount of library fragments are smaller than the read length of 50 - 100 bp, which leads to the problem of sequencing through the fragment into the adaptor sequence (for more detailed description see Materials and Methods chapter 5.2.1). Trimmed reads were then mapped to the genome of choice using Bowtie 2 (Langmead & Salzberg, 2012) with options emphasizing sensitivity and accuracy, and allowing a maximal fragment length of 2,000 bp. Bowtie2 output files were directly converted to the binary BAM format to reduce storage space, sorted according to genomic location and indexed (for more detailed description see Materials and Methods chapter 5.2.1). As recommended for ATAC-seq data, reads in genomic regions blacklisted by the ENCODE consortium were removed to avoid artifacts known to occur in those regions (Amemiya *et al.*, 2019; Encode Project Consortium, 2012). Then, PCR duplicates were excluded by keeping only the reads belonging to one fragment if several fragments with exact identical start and end positions exist (for more detailed description see Materials and Methods chapter 5.2.1). Reads with a mapping quality below 20 were removed. As the mitochondrial genome is accessible, contaminations with mitochondrial DNA during transposition reaction which are not completely avoidable lead to a high number of reads mapping to the mitochondrial genome (Bogenhagen, 2012; Yan *et al.*, 2020). As those reads have no information content regarding the chromatin accessibility in the nucleus, reads mapping to the mitochondrial genome were discarded (for more detailed description see Materials and Methods chapter 5.2.1). For quality assessment, the distribution of fragment sizes was inspected before and after removing PCR duplicates and mitochondrial reads (Figure 13). The fragment size distribution of ATAC-seq data showed a periodic pattern with a repeat length fitting to the stretch of DNA contained in a nucleosome, which is indicative of intact chromatin. The protection of DNA within nucleosomes from transposition leads to fragments spanning multiples of nucleosomes (Buenrostro *et al.*, 2013). Contamination with mitochondrial DNA (Figure 13 A) obscured this nucleosomal pattern visible in the typical fragment length distribution of ATAC-seq data (Figure 13 B). The distribution before read clean-up (Figure 13 A) was equivalent to the experimentally determined library size distribution. This emphasized the benefit of the pre-sequencing qPCR step especially for precious primary samples to

differentiate between contamination of mitochondrial DNA and the quality of the ATAC-seq library (signal-to-noise ratio).

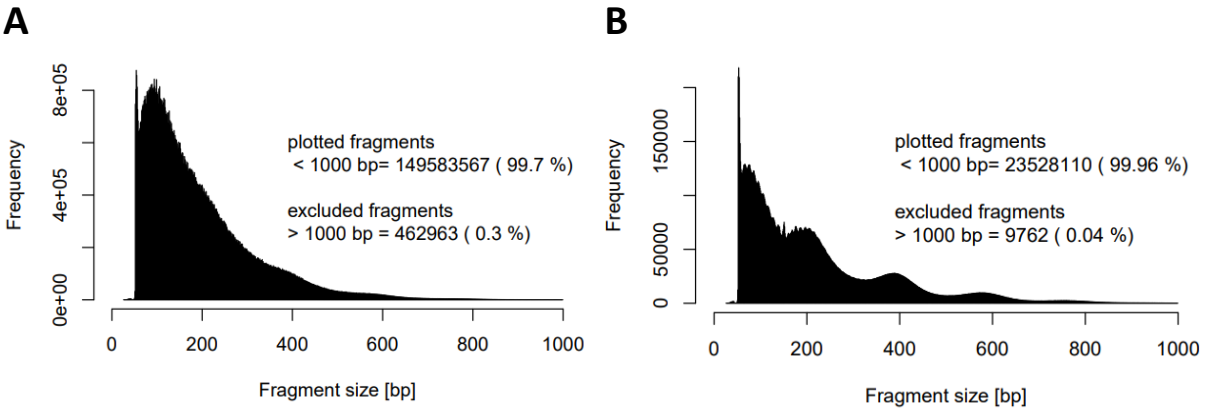


Figure 13 Fragment size distribution quality control plot.
The ATAC-seq fragment size distribution plot (**A**) before and (**B**) after removal of PCR duplicates and reads mapping to the mitochondrial genome is an output of the data analysis pipeline to control sequencing library quality. Exemplary plots were generated from an ATAC-seq library of CLL patient 24 (Figure 19) with a FRiP-score of 43 % and 19 % mitochondrial reads.

During sequencing adaptor integration, the Tn5 transposase complex cuts the two complementary DNA strands with an offset of 9 bp. The resulting 9 bp long single stranded stretches present at each of the two produced DNA fragment are filled during the first PCR step. The first sequenced base of each fragment is consequently shifted to the transposase binding event, which is centered on the middle of those 9 bp. Therefore, the actual center of the transposase binding event is on the fifth base downstream of the sequencing read start (Adey *et al.*, 2010; Buenrostro *et al.*, 2013). According to Adey *et al.* the transposase complex occupies a region of 29 bp with the 9 core base pairs plus 10 bp on either side (Adey *et al.*, 2010). Thus, a 29 bp region needs to be accessible in order that a transposase binding event can occur. Therefore, reads were shifted to be centered on the transposase binding event with an extension of 29 bp (further called ”shifted reads” or ”transposase occupancy”). These shifted reads represent the original accessible genomic position that could be occupied by the Tn5 transposase. Therefore, the shifted reads were used for density traces of specific genomic locations and all further analysis involving read counting (for more detailed description see Materials and Methods chapter 5.2.1). The shifting of the reads, used in all downstream analyses, also resulted in the desirable independence of the analysis results to the chosen sequencing read-length.

1.2.2 Optimized peak calling improves mapping of accessible regions

Regions of interest first need to be defined for ATAC-seq, similarly to ChIP-seq experiments. For the identification of genomic regions with enriched signal, different peak calling tools have been published (Ibrahim *et al.*, 2015; Zhang *et al.*, 2008). As ATAC-seq was a quite new technique at the start of my PhD project, no peak calling algorithms specifically designated to the analysis of ATAC-seq data were available and are still scarce (Yan *et al.*, 2020). Therefore, I adjusted the settings of different existing peak calling algorithms, which were originally designed for ChIP-seq data, for ATAC-seq. I compared the results from the peak calling tools JAMM (Ibrahim *et al.*, 2015), MACS (Zhang *et al.*, 2008) and its successor MACS2. In contrast to ChIP-seq, the region of interest for ATAC-seq data is not the center of the sequenced fragment, which is by default identified by these peak callers (=nucleosome position). Instead, the regions of interest are near the ends of the sequenced fragments as they mark the cutting position of the transposase enzyme and the center of the surrounding accessible bases needed for its binding. Accordingly, reads were either shifted to be centered on the transposase cutting position prior to peak calling (JAMM, MACS) or, if possible, settings were adjusted to include the read shifting in the peak calling process (MACS2). As there was no obvious choice for ATAC-seq data, the parameters effecting the extension size around the transposition event were additionally varied for MACS2, which represents a smoothing window and influenced the size of called peaks (Figure 14 B). A dataset of a CLL patient (CLL1; see Figure 19 A) was used for this evaluation. JAMM and MACS2 (shift: -10; extension size: 28) captured the borders of the accessible sites best (Figure 14 A). The flanking nucleosomes were visible in the coverage plots, while MACS2 with higher extension sizes, e.g., 100 bp fail to capture the flanking nucleosomes. Besides, higher extension sizes tended to combine accessible regions including those that are independent as judged from the genome annotation exemplary shown for a location containing two annotated TSS (Figure 14 C). Peaks called with MACS were broader and tended to include besides the core accessible region its flanking regions with slightly elevated read counts (Figure 14 A-C). In summary, MACS2 (shift: -10; extension size: 28) and JAMM performed best in determining the actual accessible region. However, JAMM is designed to find all regions with a possible enrichment and it is up to the user to subsequently decide on an appropriate significance threshold. The additional threshold choice step for JAMM is especially difficult if no replicates exist. Besides, JAMM did not allow to include the shifting of the reads to the center of the transposase binding event, therefore read shifting was required as an additional step. Thus, MACS2 with the settings "-shift -10 -extsize 28" was favored and used for all further analysis.

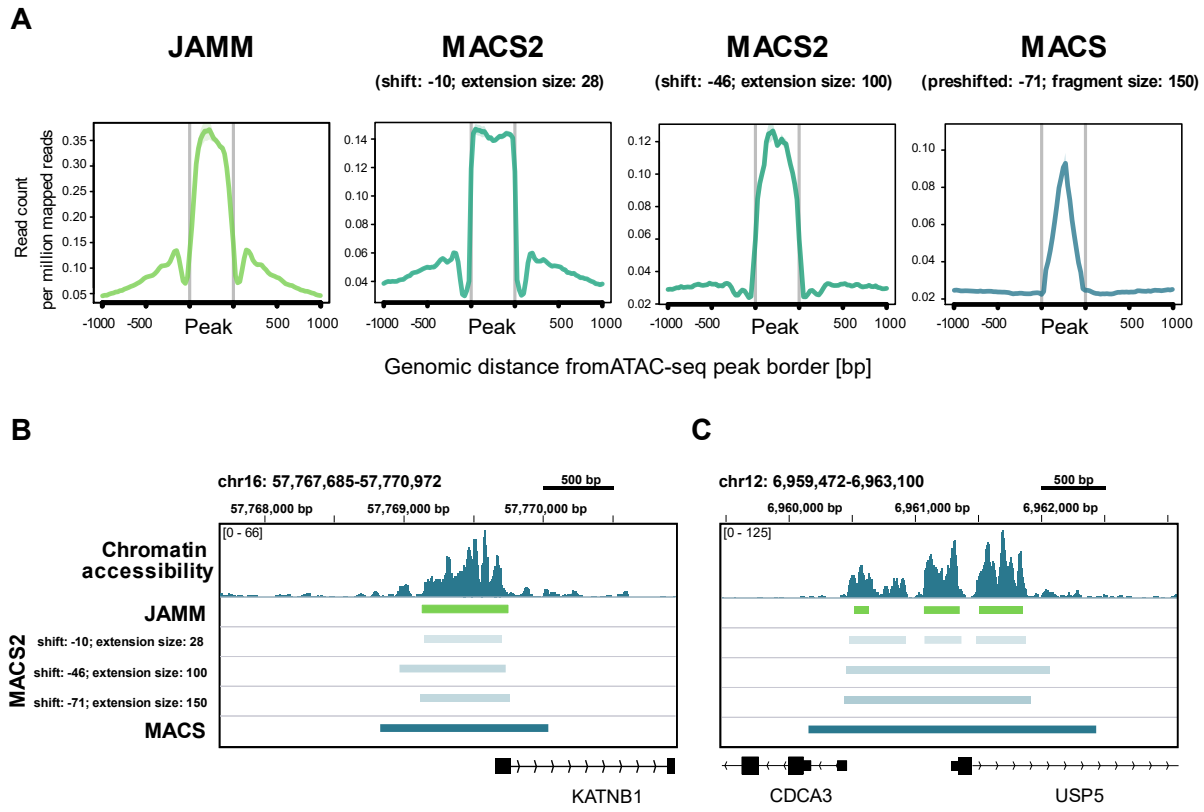


Figure 14 Comparison of different peak callers and settings for ATAC-seq data.
A Average density profile of shifted ATAC-seq reads at all called peaks and the 1 kb surrounding genomic region with the indicated method and settings for a library generated from cells of a CLL patient (CLL1). **B** Comparison of differences in the peak calling results between the indicated methods and settings at an exemplary genomic region on chromosome 16 around the *KATNB1* promoter. Density of shifted reads representing the chromatin accessibility for the transposase at a specific region is shown in the track at the top. **C** Same as B for a genomic region with two adjacent annotated TSS.

Next, I evaluated if the called peak regions determined with the selected settings (MACS2 -shift -10 -extsize 28) were located at the expected genomic positions. For this ATAC-seq was conducted for a cohort of CLL patients (see Figure 19 A) with technical assistance of Sabrina Schumacher (Division Chromatin Networks, DKFZ, Heidelberg). Accessible chromatin is mainly located at active regulatory elements like promoters and enhancer (Klemm *et al.*, 2019). Further, such active regulatory elements are associated with defined histone modifications (Roadmap Epigenomics *et al.*, 2015). Thus, chromatin state assignments based on functionally relevant histone modifications were employed. Chromatin state annotations were generated by Naveed Ishaque (Division of Theoretical Bioinformatics and Heidelberg Center for Personalized Oncology, German Cancer Research Center (DKFZ), Heidelberg) individually for the same patients (n=18) subjected to an ATAC-seq analysis (Mallm *et al.*, 2019). The detected ATAC-seq peaks in every patient were predominantly located at regions assigned to states indicative of TSS or enhancers (Figure 15). Only a small fraction of ATAC peaks could not be assigned to a chromatin state based on the histone modification data. Additionally to regulatory elements, accessible sites can also be found for example at insulators bound by CTCF (Klemm *et al.*, 2019). ATAC-seq peaks at genomic regions depleted of the analyzed histone

modifications (= "quiescent") could arise from CTCF sites. The higher percentage of ATAC-seq peaks in the quiescent state in CLL patients 13-19 coincided with a lower coverage of the ChIP-seq data. Therefore, it is attributable to the quality of the annotation not the ATAC-seq peak calling. Thus, the called ATAC-seq peaks did not only correctly capture the borders of enriched region, but were also located at genomic locations associated with accessible chromatin.

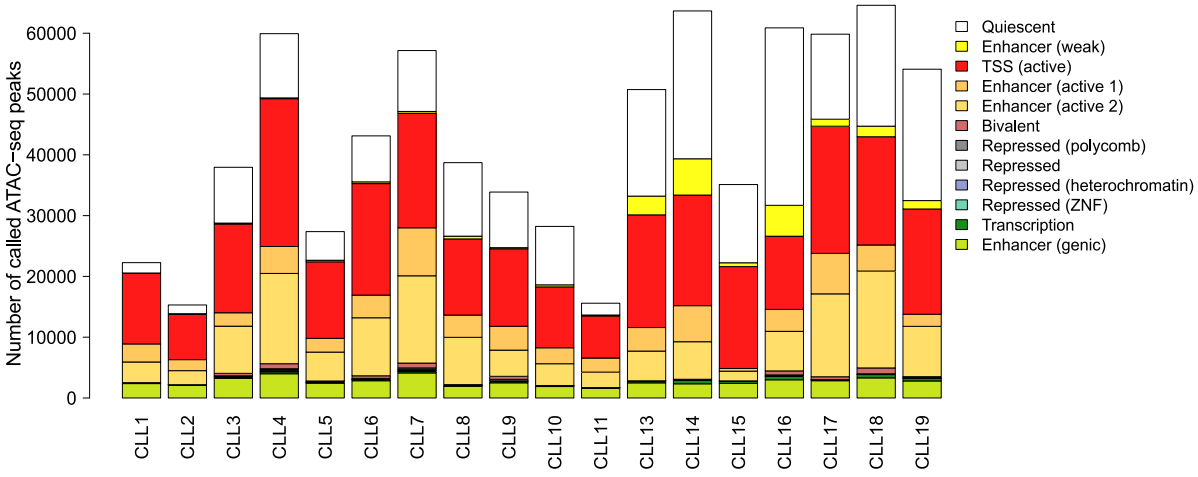


Figure 15 Functional assignment of called ATAC-seq peaks based on histone modifications. ATAC-seq peaks were assigned to functional chromatin states based on the ChromHMM annotation by Mallm *et al.* (Mallm *et al.*, 2019), which is determined from histone modification data of seven functionally relevant modifications.

1.2.3 Flexible selection options improve the applicability of the ATAC-seq analysis pipeline

The analysis pipeline was continuously further developed. In order to enable the analysis of ATAC-seq data of different species, genome builds and corresponding annotations can be added and used by selecting a single parameter. Currently, the analysis is possible for the most commonly used species human and mouse. For human data the genome builds hg19 with decoy sequences (hs37d5) or hg38 (GRCh38) and for mouse derived ATAC-seq data mm9 and mm10 can be selected (genome builds and annotations for mouse were implemented by Dr. Markus Muckenhuber (Division chromatin networks, German Cancer Research Center (DKFZ), Heidelberg, Germany). Besides, the analysis of single-end sequenced data is possible. A summary file with most important quality metrics for all samples of a project is provided including number of reads, number of called peaks and percentage of reads mapping to the mitochondrial genome. The fraction of reads in peaks (FRiP) score, which is an important measure of the signal-to-noise ratio for ATAC-seq data, was calculated and used in the following analysis to exclude low quality samples (for more detailed description see Materials and methods chapter 5.2.1).

1.2.4 A minimum sequencing depth of 25 million reads is needed for ATAC-seq

Finally, I assessed the requirements for sequencing depth for the used ATAC-seq protocol. Costs and benefits of deeper sequencing need often to be balanced as sequencing costs represent a large cost factor of an experiment. Therefore, I applied the ATAC-seq analysis pipeline to a deeply sequenced ATAC-seq library of a CLL patient (CLL1, see Figure 19 A) with 92 million mapped reads. Similar to the approach used by Jung *et al.* for ChIP-seq data (Jung *et al.*, 2014), I downsampled the full library to 5, 10, 20, 40, and 80 million reads. Then, I calculated the Pearson correlation of these downsampled libraries to the full library based on counts in 50 bp genomic bins. The correlation indicated that saturation was slowly approached for sequencing library sizes above 25 million mapped reads (Figure 16 A). I then determined how many of the peaks, which were called for the full data set, could be recaptured for the downsampled subsets. Interestingly, a library size below 25 million mapped reads captured less than 50 % of the peaks (Figure 16 B). The percentage of recaptured accessible regions in dependence of the library size was then fitted by a logarithmic function. By calculating the first derivative of this function the benefit of additional sequencing was determined. The derivative represents the percental increase in recovered peaks with every additionally sequenced 1 million reads (Figure 16 C). For library sizes above approximately 27.8 million reads the increase of regions that would have been captured with a higher sequencing depth decreased below 1 % per 1 million additional reads. Taken everything together, we aimed to generate ATAC-seq libraries of at least 25 million reads per replicate to create high quality data sets.

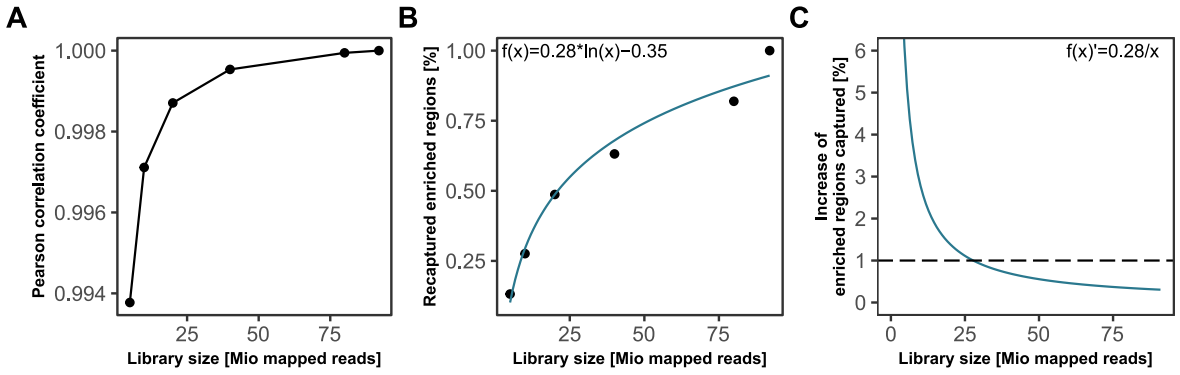


Figure 16 Impact of ATAC-seq library sequencing depth.
A Reads of a deep-sequenced ATAC-seq library (92 million reads) were downsampled to 5, 10, 20, 40, 80 million reads. Shifted ATAC-seq reads centered on the transposase binding sites were counted in 50 bp bins for every library size. The Pearson correlation coefficient of the accessibility pattern of every subsampled library and the full library is plotted against the size of the respective library. **B** The percentage of full dataset peaks which could also be detected in the downsampled datasets is shown in dependence of the corresponding library size. A logarithmic function was fitted to the data (blue). **C** First derivative of logarithmic function fitted in B with the indicated formula illustrates the increase in the percentage of captured full dataset regions with every additional million sequenced reads at a certain library size (blue). Dashed line represents arbitrary threshold for a minor benefit of greater sequencing depth of 1 %.

1.3 Application of ATAC-seq to primary human leukemia cells

The ATAC-seq workflow for identifying differences in chromatin accessibility between certain conditions was then tested in primary human leukemia cells after perturbing histone and DNA methylation. ChIP-seq experiments for several histone marks to access the location of those changes would require extensive amounts of cells, which are often not available for primary patient samples. ATAC-seq offers a condensed readout of the chromatin state with material input requirements approximately one magnitude below one ChIP-seq experiment. This enables the characterization of the epigenetic state of the chromatin for many clinical samples of patients also under treatment (Greenleaf, 2015). The application of ATAC-seq to primary cells was assessed in *IDH1* mutated AML as a system which can be targeted by the newly developed inhibitor BAY1436032 of mutated isocitrate dehydrogenase (IDH). Specific mutations in the *IDH* gene lead to a gain-of-function of the enzyme. The mutated enzyme further converts α-KG to D-2-hydroxyglutarat (2-HG). α-KG is however an important cofactor for a subset of DNA and histone demethylases, which are inhibited by the oncometabolite 2-HG (Rakheja *et al.*, 2013). Thus, this inhibitor has the potential to revert histone methylation changes induced by a well-defined neomorphic mutation in a metabolic enzyme, which makes it an interesting system with different epigenetic states to apply ATAC-seq to primary human cells.

1.3.1 *IDH1* mutated AML patient cells show partially reversible histone hypermethylation

Formation of 2-HG in *IDH1* mutated tumors is associated with a histone hypermethylation phenotype (Lu *et al.*, 2012; Sasaki *et al.*, 2012; Turcan *et al.*, 2012). Before the ATAC-seq workflow was applied to primary AML cells, this predicted change in the epigenetic state of the chromatin was validated by immunoblotting. For this, the histone methylation levels in e*x vivo* cultivated cells of AML patients carrying an IDH1 R132C or IDH1 R132H mutation were evaluated in comparison to patient-derived IDH1 wild-type cells. H3 fractions trimethylated at the functionally important residues K4, K36, K9 and K27 were assessed by immunoblotting (Figure 17). Global amounts of all four analyzed histone H3 trimethylation marks were increased in IDH1 mutated compared to IDH1 wild-type cells irrespective of the type of amino acid substitution present in the mutated enzyme. The histone hypermethylation in human primary AML cells thus affected both active and repressed chromatin associated marks. Treatment with 500 nM pan-mutant IDH1 inhibitor BAY1436032 for 14 days *ex vivo* reduced global histone H3 methylation levels in IDH1 R132C and IDH1 R132H AML patient-derived cells compared to untreated controls. In contrast, histone methylation levels in IDH1 wild-type AML patient cells did not change upon treatment which validated the specificity of the inhibitor BAY1436032 for the mutated form of the IDH1 enzyme. The aberrantly large methylated fraction of H3 in *IDH1* mutated AML was reversible with BAY1436032. However, immunoblotting cannot determine if the chromatin localization of disease-induced histone hypermethylation and drug-mediated reduction of methylation are identical.

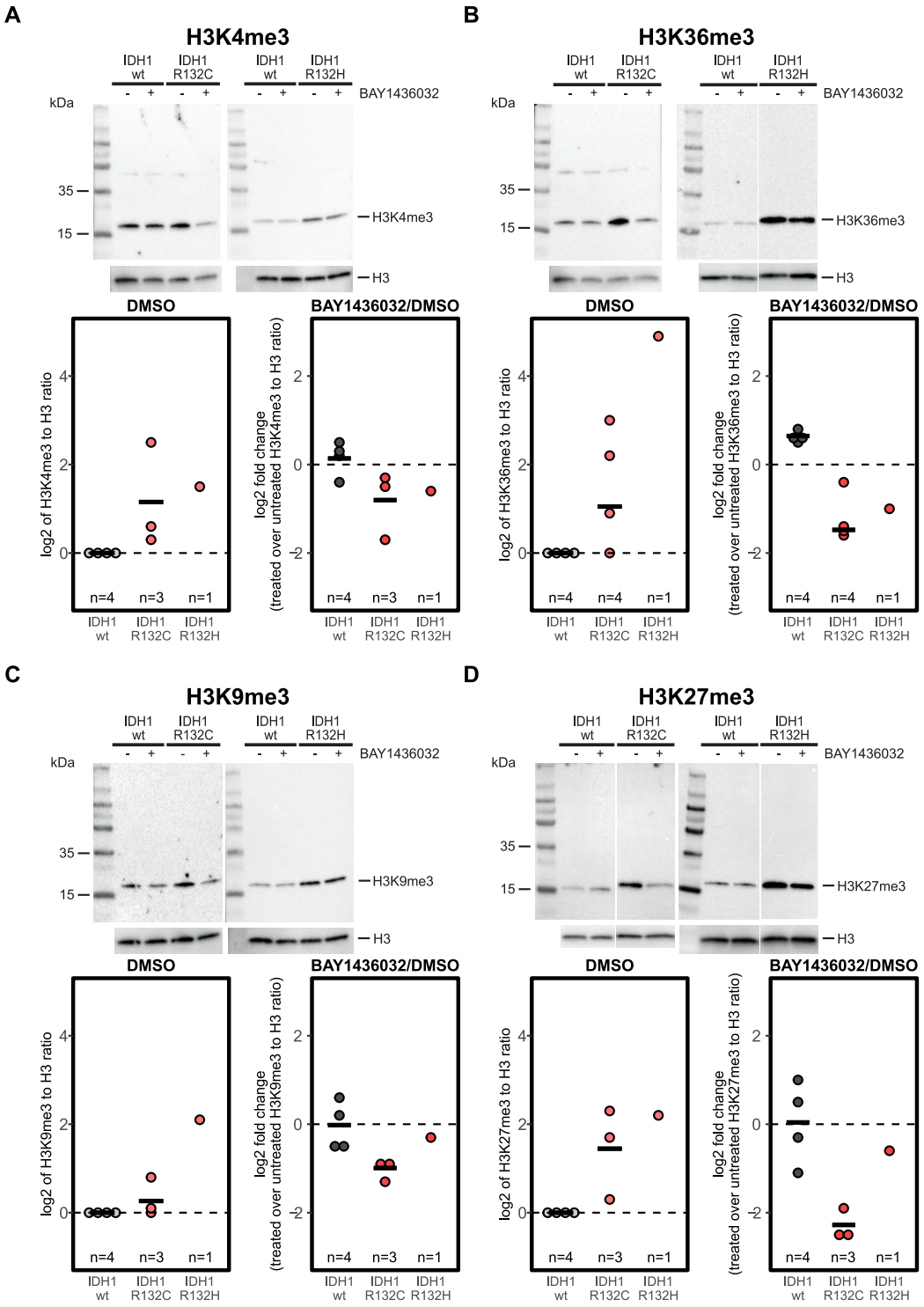


Figure 17 *Ex vivo* effect of pan-mutant IDH1 inhibitor BAY1436032 on global histone methylation levels in cells of IDH1 wild-type, IDH1 R132C or IDH1 R132H AML patients. Patient-derived cells with wild-type IDH1 or mutated IDH1 with an R132C or R132H substitution were treated *ex vivo* for 14 days with 500 nM BAY1436032 or the dissolvent dimethyl sulfoxide (DMSO). Representative blots and quantification (dot plots) of a immunoblot analysis which compared the fraction of trimethylated H3 at lysins (**A**) K4, (**B**) K36, (**C**) K9 and (**D**) K27 of inhibitor treated to untreated IDH1 wild-type and IDH1 mutant AML cells. The integrated intensity of bands from each blot separately was used to calculate logarithmized (base 2) fold changes relative to IDH1 wild-type of the fraction of the specified trimethylation to total H3 in the untreated condition (left) and the ratios logarithmized to base 2 of treated to untreated fold changes to IDH1 wild-type cells (right). For conditions with more than one measurement the mean is indicated with a line. Wt: wild-type. Ladder: PageRuler Plus Prestained Protein Ladder, 10 to 250 kDa (Thermo Fisher Scientific). Figure adapted from Chaturvedi *et al.* (Chaturvedi *et al.*, 2017).

1.3.2 Chromatin accessibility changes are linked to reversible hypermethylation

The immunoblot analysis of histone methylation levels confirmed the chromatin state changes in the analyzed IDH1 mutated AML patient samples. Next, corresponding changes in chromatin accessibility were studied by applying the ATAC-seq workflow to treated cells from an AML patient with wild-type IDH1 and an AML patient with the IDH1 R132C mutation. Cells were treated *ex vivo* for 14 days with 500 nM pan-mutant IDH1 inhibitor BAY1436032 or DMSO as negative control. The quality of the sequenced ATAC-seq libraries was good with FRiP-scores between 15 % to 23 %. Sites with high chromatin accessibility were identified in all four conditions. These sites mainly occurred at regions annotated as transcription start sites or enhancers according to a ChromHMM annotation based on the myeloid lineage cell line K562 (Lozzio & Lozzio, 1975) by ENCODE (Encode Project Consortium, 2012; Ernst & Kellis, 2012). Some regions displayed an increased accessibility in IDH1 mutated compared to wild-type cells, which was reverted upon treatment with the inhibitor of mutant-IDH1 to the wild-type level (Figure 18). More technical and biological replicates would be needed to conduct a comprehensive and reliable identification of regions with a changed accessibility attributable to the mutation in the IDH1 enzyme or the effect of the inhibitor. Nevertheless, the analysis of AML patient cells showed that the ATAC-seq workflow worked well for mapping the chromatin state of human primary cells from leukemia patient samples.

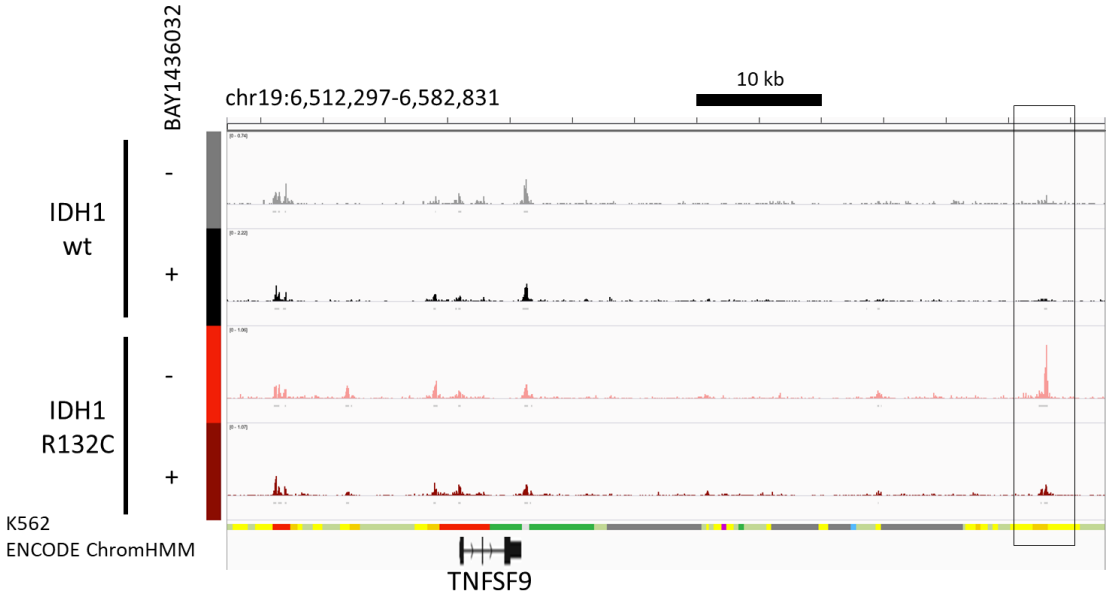


Figure 18 Chromatin accessibility of human primary cells of AML patients.
Chromatin accessibility differences at a locus on chromosome 19 of two AML patients with either wild-type IDH1 or mutated IDH1. Rectangle marks region with changed chromatin accessibility in IDH1 mutated cells upon treatment with the inhibitor BAY1436032 at an enhancer (orange) according to the ChromHMM chromatin state annotation of K562 cells by ENCODE (Encode Project Consortium, 2012; Ernst & Kellis, 2012).

2 Modeling chromatin accessibility dynamics during non-malignant CLL precursor maturation

The experimental and data analysis ATAC-seq workflow described above was applied to study deregulated epigenetic patterns during physiological development, disease onset and drug treatment. A tumor cell harbors epigenetic signatures of the non-malignant tumor initiating cell, which is also referred to as the tumor cell-of-origin. In addition, tumors acquire disease-specific epigenetic patterns that originate from aberrant cellular processes that occur during tumorigenesis (Oakes *et al.*, 2016; Visvader, 2011; Wierzbinska *et al.*, 2020). To study the deregulated epigenetic patterns in disease it is therefore crucial to use the non-malignant tumor cell-of-origin as reference. However, even within one well-defined hematopoietic neoplasm like chronic lymphocytic leukemia (CLL) differences in the maturation status of the tumor cell-of-origin are existing, which in turn affects the epigenetic landscape (Kulis *et al.*, 2015; Oakes *et al.*, 2016). Therefore, I dissected the chromatin accessibility dynamics of potential tumor initiating cells for CLL during normal B-cell maturation.

2.1 Determination of the healthy reference state for CLL patients

2.1.1 Characterization of non-malignant B-cell references and patient samples

The selection of a suitable healthy reference is particularly difficult in CLL and was previously addressed by different approaches (Beekman *et al.*, 2018; Landau *et al.*, 2014). CLL patients can be divided in IGHV mutated and unmutated with prognostic potential for the patients. The basis of this binary classification is the maturation stage of the tumor cell-of-origin of each patient. Classically, $CD19^+$ B-cells isolated from peripheral blood of healthy donors are often used as reference (Landau *et al.*, 2014; Oakes *et al.*, 2014; Pei *et al.*, 2012). This comprises however a mixture of B-cells from several maturation stages, which are present in the blood (Klein *et al.*, 1998). Alternatively, a more accurate description of the non-malignant reference state would also account for the specific maturation stage of the direct tumor cell-of-origin as described below. However, the required detailed characterization of the sample is frequently not possible in a clinical setting. Thus, the conventional approach to compare CLL cells to $CD19^+$ B-cells was first conducted in this thesis (chapter 3.1) with an ATAC-seq dataset of $CD19^+$ cells in three pools of healthy donors and identically processed samples of 19 CLL patients (Figure 19 A). The CLL patient cohort contained both IGHV hypermutation subtypes. With about 60 %

IGHV mutated and 40 % IHGV unmutated cases the cohort was representative for newly diagnosed CLL patients (Gaidano & Rossi, 2017). To address the more complex task to compare CLL cells to their actual cell-of-origin (chapter 3.2), we conducted ATAC-seq of six distinct normal B-cell subpopulations from the spectrum of potential CLL precursors of three healthy donors each (Figure 19 B). Five of eight CLL patients, which were identically processed with the low input protocol, were additionally monitored over treatment with ibrutinib for several month. First, I examined the accessibility pattern during progressive maturation of potential healthy B-cell precursors of CLL cells.

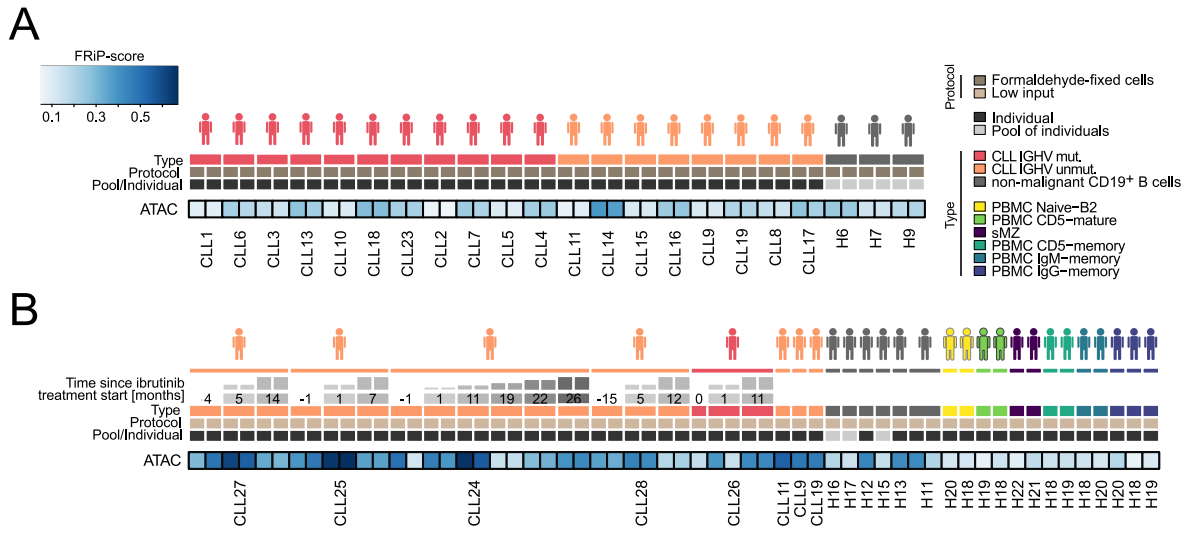


Figure 19 ATAC-seq data quality assessment of all healthy references and CLL patients. Fraction of reads in peaks (FRiP) scores for individual technical replicates (separate library preparation) of samples which were used **(A)** for the conventional approach comparing CLL patient samples to CD19⁺ B-cells from peripheral blood of healthy donors or **(B)** for the tumor cell-of-origin approach, which compares the epigenetic patterns of individual CLL patient to their direct cell-of-origin. Five patients were additionally monitored during treatment with ibrutinib ("Time since ibrutinib treatment start").

2.1.2 Chromatin accessibility changes during physiological B-cell development

As previously shown by the analysis of DNA methylation data, the state of the chromatin is changing considerably during physiological B-cell maturation of potential CLL precursors (Kulis *et al.*, 2015; Oakes *et al.*, 2016). This epigenetic plasticity enables transcriptional programs necessary for a certain maturation stage (Wu & Sun, 2006). As chromatin accessibility offers the possibility to link epigenetic alteration to transcription factor networks, I analyzed the changes in chromatin accessibility of B-cell maturation stages from which CLL might originate. For this, I used the ATAC-seq data of six defined B-cell subpopulations, which represent developmental timepoints of B-cell differentiation. The cell subpopulations were sorted from peripheral blood or spleen tissue of at least two healthy individuals (Figure 20 A). A consensus peak set of 80,226 accessible genomic regions present in a minimum of two individuals in either of the six distinct subsets were identified.

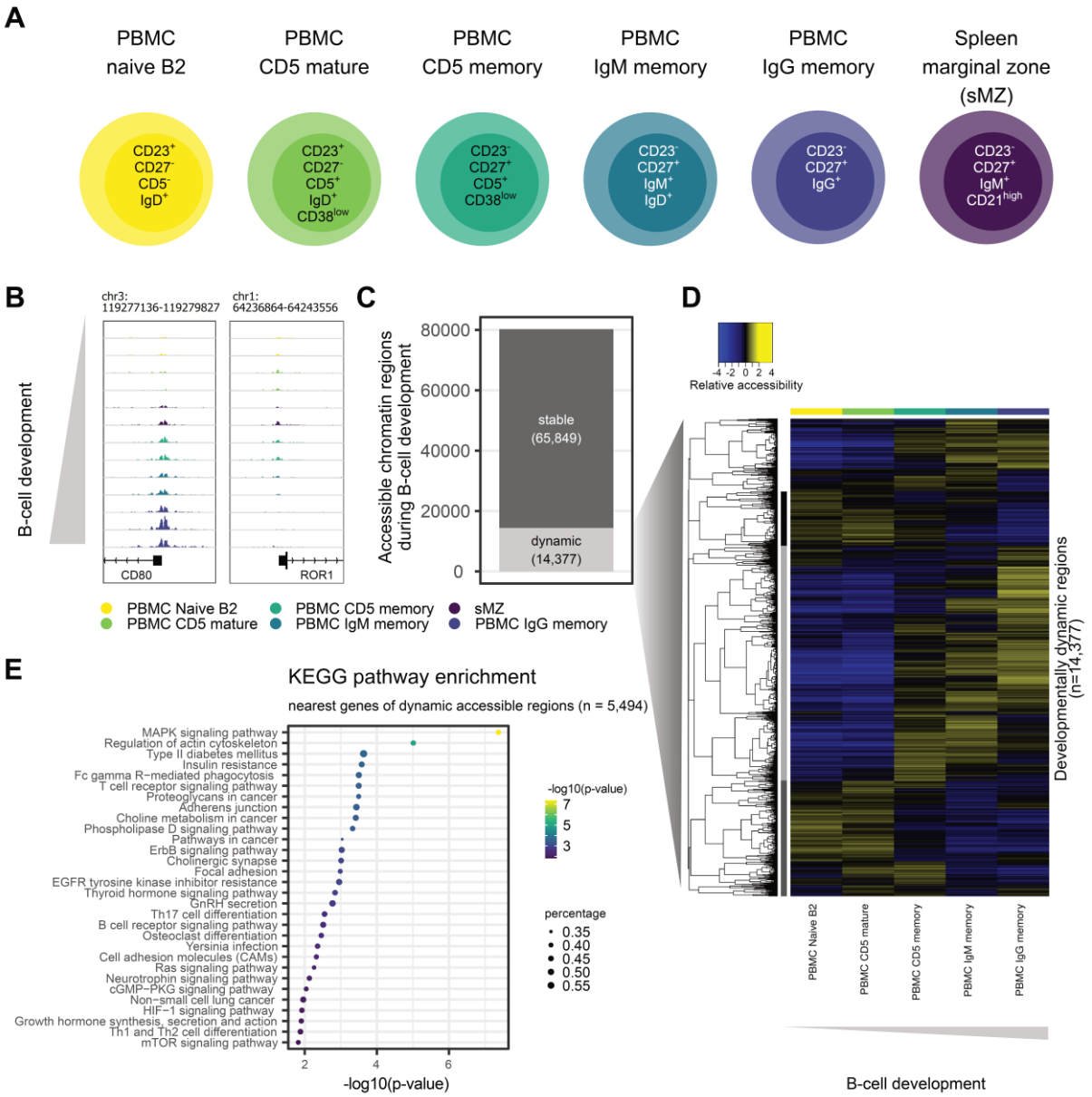


Figure 20 Identification of regions with dynamic chromatin accessibility during B-cell maturation.
A Six B-cell subpopulations previously isolated from peripheral blood and spleen tissue of two to three healthy individuals with the indicated markers represent distinct B-cell maturation stages and were used for the analysis. **B** ATAC-seq tracks of two exemplary regions with changed chromatin accessibility between the B-cell subpopulations. **C** Fraction of sites with developmentally dynamic and stable chromatin accessibility in the analyzes B-cell maturation stages. **D** Clustering of accessibility patterns at the 14,377 dynamic regions during B-cell development from naïve to memory B-cells. For normalization accessibility counts of a region for a subpopulation was averaged over all biological replicates which passed the quality control and divided by the average count over all samples for that region. Plotted is the natural logarithm of those normalized counts. Complete linkage hierarchical clustering was used based on cosine correlations to cluster the regions according to their dynamics from naïve B2, CD5 mature, CD5 memory, IgM memory to IgG memory. **E** Pathway enrichment analysis of nearest genes of 14,377 regions with dynamic chromatin accessibility.

From those consensus regions, developmentally dynamic regions were defined as sites with significantly differential ATAC-seq counts between the subpopulations using DESeq2 (Love *et al.*, 2014). 14,377 (~18 %) peak regions of the consensus set were dynamic during development in the analyzed subsets (Figure 20 B, C). The dynamic regions could be roughly divided into two subtypes. The first group of regions showed a high accessibility in naïve B2 and CD5 mature cells which was lost during the differentiation to memory B-cells (Figure 20 D). The second group had a closed chromatin conformation in naïve cells which opened up in more mature cells. The dynamic regions were assigned to their nearest gene for an KEGG pathway (Kanehisa & Goto, 2000) enrichment analysis to confirm the identification of biologically relevant sites. The nearest gene approach, as proximation for the effect of regulatory elements on gene expression, yielded among others an enrichment of genes in pathways related to B-cell development like T- and B-cell signaling and cell differentiation (Figure 20 E). In summary, the chromatin accessibility changed during physiological development of potential CLL precursors which raised the possibility to calculate phylogenetic distances between B-cell subpopulations based on differences in accessibility patterns.

2.2 Phylogenetic tree construction of B-cell development

2.2.1 A phylogenetic tree of B-cell development can be constructed from ATAC-seq

The development from naïve to memory B-cells is a continuous process with a gradual change of the epigenome as shown for DNA methylation patterns (Oakes *et al.*, 2016). To model the chromatin accessibility dynamics during development the six distinct B-cell subpopulations needed to be phylogenetically linked based on their accessibility profiles. While conventional phylogeny analysis relates taxa based on sequence information, inference of phylogenetic relationships is also possible for differentiation processes. Phylogenetic relations of B-cell subpopulations were already successfully inferred with the minimum-evolution algorithm (Desper & Gascuel, 2002) from DNA methylation data (Oakes *et al.*, 2016). Therefore, I reconstructed the phylogeny of the analyzed B-cell subsets based on chromatin accessibility patterns. To build a phylogenetic tree the balanced minimum-evolution approach by Desper & Gascuel (Desper & Gascuel, 2002) was applied on pairwise Manhattan distances between B-cell subpopulations, which were calculated from normalized ATAC-seq counts in the 14,377 developmentally changing regions. The accessibility-based phylogenetic tree showed a main trajectory of phylogeny which started with naïve B2 cells and CD5 mature cells, proceeded with CD5 memory cells followed by IgM memory cells and ended with IgG memory cells (Figure 21 A). The accessibility profile of splenic marginal zone B-cells differed greatly from peripheral blood derived subpopulations. Biological replicates were located close together on the developmental axis and results were in concordance with previously published phylogenetic relations determined from DNA methylation

between naïve, IgM memory and IgG memory B-cells (Oakes *et al.*, 2016). It is, thus, possible to infer phylogenetic relationships of distinct cell populations from ATAC-seq data.

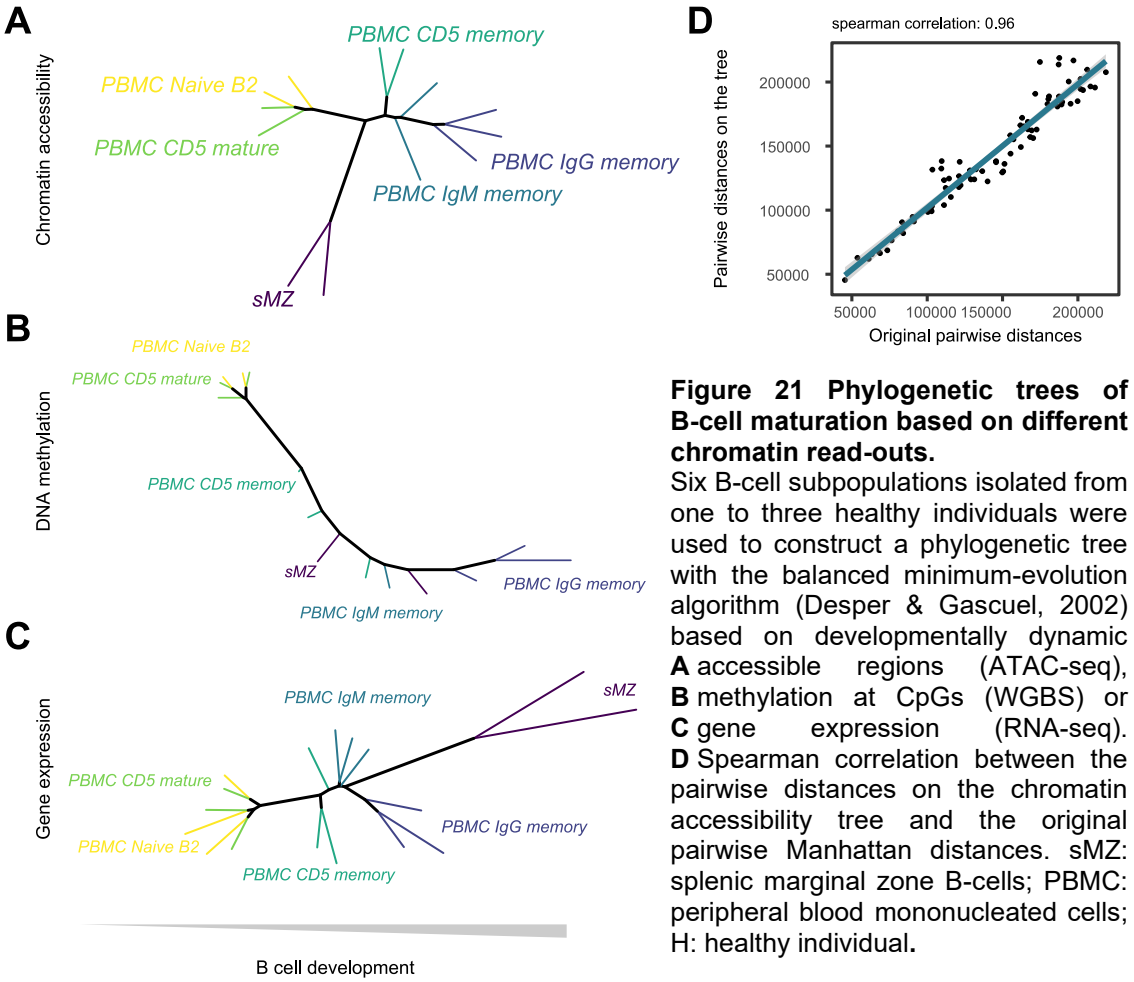


Figure 21 Phylogenetic trees of B-cell maturation based on different chromatin read-outs.
Six B-cell subpopulations isolated from one to three healthy individuals were used to construct a phylogenetic tree with the balanced minimum-evolution algorithm (Desper & Gascuel, 2002) based on developmentally dynamic **A** accessible regions (ATAC-seq), **B** methylation at CpGs (WGBS) or **C** gene expression (RNA-seq). **D** Spearman correlation between the pairwise distances on the chromatin accessibility tree and the original pairwise Manhattan distances. sMZ: splenic marginal zone B-cells; PBMC: peripheral blood mononucleated cells; H: healthy individual.

2.2.2 The chromatin accessibility-based phylogenetic tree of B-cell development is similar to trees constructed from transcriptome or DNA methylation data

I next examined how the phylogenetic relationships based on chromatin accessibility compare to the relations inferred from transcriptome (RNA-seq) and another epigenetic feature, DNA methylation (whole-genome bisulfide sequencing). In addition to ATAC-seq, the data set of normal B-cell subpopulations comprised RNA-seq and whole-genome bisulfide sequencing (WGBS) data of the same samples. Separate phylogenetic trees were generated from Euclidean distances of DNA methylation patterns by Roma Kurilov (Division of Applied Bioinformatics, DKFZ, Heidelberg) (Figure 21 B) and Manhattan distances of normalized gene expression profiles (FPKM) of 7,671 differentially expressed genes between the sorted B-cell subsets (Figure 21 C). The tree based on ATAC-seq showed a similar phylogeny than the DNA methylation-based tree and the gene expression-based tree (Figure 21). Splenic marginal zone (sMZ) B-cells were the only observed exception. Their accessibility and transcriptome profiles were farer apart from the major developmental trajectory than their DNA methylation patterns.

In addition, the two $CD5^+$ subpopulations, which were not included in previously published phylogeny analyses, were successfully assigned to the phylogenetic trajectory. CD5 mature cells were remarkable similar to naïve B-cells in their epigenetic and transcriptional profile. CD5 memory cells were consistently located between naïve cell populations and their phylogenetically closest subset, the IgM memory cells. In summary, the phylogenetic analysis can be used to study relationships of isolated small subpopulations. Besides DNA methylation patterns and transcriptional profiles, phylogenetic relations were inferred from ATAC-seq profiles and yielded comparable results.

2.2.3 Linear modeling infers continuous physiological accessibility dynamics

I next aimed to model the chromatin accessibility dynamics at individual genomic regions during physiological B-cell development as performed for DNA methylation in the methylome-based cell-of-origin modeling (Methyl-COOM) approach by Wierzbinska *et al.* (Wierzbinska *et al.*, 2020). For this I first deduced a quantitative measure for the maturation stage of the healthy B-cell subpopulation from the phylogenetic tree. To quantify the degree of maturation, the fraction of developmental progression from naïve to memory B-cells was defined as maturation score between zero and one (Figure 22). The maturation score was set to zero for the naïve B2 replicate which marks the starting point of the developmental trajectory and to one for the IgG memory replicate which marks the end point of the main trajectory. The maturation score for every biological replicate of the six B-cell subpopulations was then calculated from their position on the main phylogenetic trajectory (black). The assignment of the B-cell subpopulations to a phylogenetic stage between zero and one enabled to model the accessibility at a certain region during B-cell development by linear regression. This allowed to infer the accessibility signal continuously for a specific region at every developmental stage ranging from naïve to IgG memory B-cells.

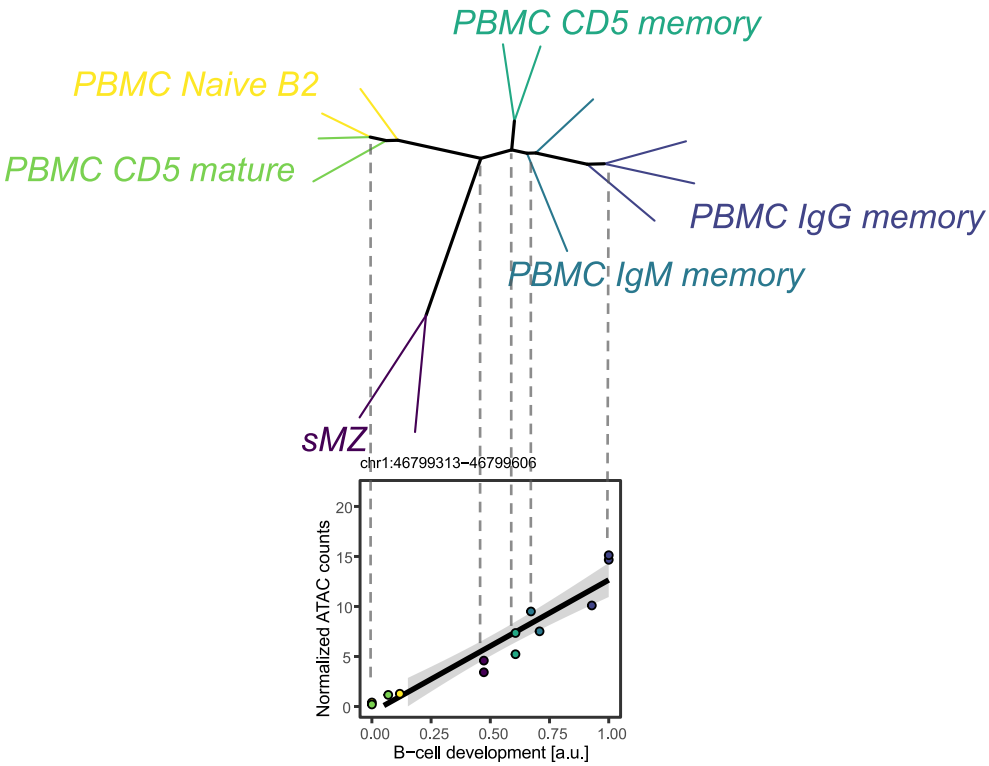


Figure 22 Illustration of phylogenetic stage assignment and linear modeling of physiological chromatin accessibility dynamics.
The two end points of the main phylogenetic trajectory were assigned to maturation scores of zero(naïve) and one (mature). Phylogenetic distances between naïve B2 cells and the branching points of individual B-cell subpopulations were transformed into evolved fractions of the main maturation path from naïve B2 to IgG memory cells. Resulting maturation stages of B-cell subpopulations were used to model the chromatin accessibility dynamics during B-cell development at individual genomic regions by linear regression (line) exemplary shown for a region on chromosome 15.

3 Multi-omics analysis of disease- and drug-induced changes of chromatin in CLL

Among the emerging targeted cancer therapies is a promising treatment approach for CLL which interferes with the aberrantly high activation of B-cell receptor signaling. Belonging to this category of approved drugs for CLL, ibrutinib restricts B-cell receptor signaling important for CLL cell survival by inhibition of the signal pathway component Bruton's tyrosine kinase (BTK) (Hallek, 2013). In the absence of many recurrent mutations, CLL is known for a deregulated transcriptome (Burger & Chiorazzi, 2013) with increasing evidence found for the presence and importance of disrupted epigenetic regulation (Chen *et al.*, 2009; Kulis *et al.*, 2015). Drugs might have an effect on various levels like the genome, epigenome and transcriptome. Before studying how drugs act on epigenetic patterns, it is important to first identify and characterize disease-induced changes in comparison to the physiological state.

3.1 Deregulation of enhancers in CLL in comparison to CD19$^+$ B-cells from healthy donors

After the analysis of chromatin accessibility patterns in non-malignant B-cells, I next aimed to study their potential disruption in the CLL disease-state. The characterization of genome-wide epigenetic deregulation in disease by ATAC-seq allows to assess activity changes of regulatory elements like enhancers and changes in TF binding in the CLL disease state (Buenrostro *et al.*, 2013). I first used the conventional approach to determine disease-induced changes in CLL cells. This approach uses a mixture of all CD19$^+$ B-cells extracted from peripheral blood mononucleated cells of non-malignant donors to assess the physiological state of CLL from the blood compartment.

3.1.1 Enhancers display extensive changes of chromatin accessibility

To identify chromatin accessibility differences between CLL cells and non-malignant CD19$^+$ B-cells ATAC-seq was conducted for CD19$^+$ cells from peripheral blood of 19 CLL patients and three non-malignant controls (pools of cells from up to five individuals) (Figure 19 A). A consensus peak set was generated from all samples consisting of 108,673 peaks (median peak size: 300 bp). Unsupervised hierarchical clustering based on Spearman correlation of the ATAC-seq counts in the consensus peaks

separated CLL patients from non-malignant controls (Figure 23 A). The R package DiffBind (Ross-Innes *et al.*, 2012) was used to identify regions with differential accessibility between CLL cells and non-malignant CD19$^+$ B-cells. The logarithmized (base 2) fold change thresholds for sites to be classified as differentially accessible in CLL compared to non-malignant cells were determined to be ≤ -1.5 and ≥ 1.8. The thresholds were identified by fitting the sum of three gaussian functions to the distribution of fold changes representing the three classes of regions with either lost, unchanged or gained accessibility in CLL (Figure 23 B). The false discovery rate threshold (FDR) was also determined from the data by the inflexion point of the number of differential sites as function of the FDR threshold. Thus, 6,111 regions with gained accessibility in CLL compared to CD19$^+$ B-cells could be identified while 8,305 regions showed a loss of accessibility signal. The cytotoxic T lymphocyte associated protein 4 (*CTLA4*) and EBF transcription factor 1 (*EBF1*) loci in Figure 23 C and D illustrate exemplary regions with gained and lost accessibility in CLL, respectively.

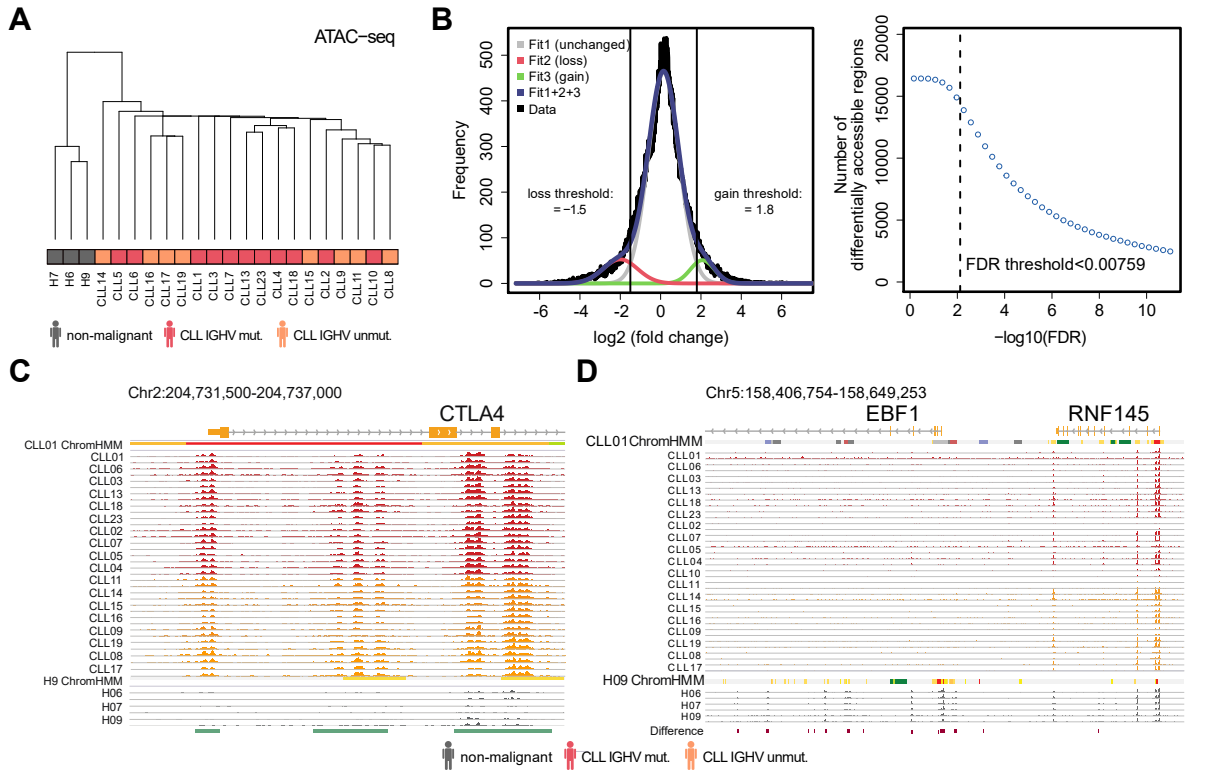


Figure 23 Chromatin accessibility differences between CLL patients and CD19$^+$ B-cells of non-malignant controls.
A Hierarchical cluster (single linkage) based on Spearman correlation between samples calculated from the accessibility signal in consensus peak set of all samples. **B** Determination of fold change (left) and FDR (right) thresholds from the data for the differential accessibility analysis by DiffBind (Ross-Innes *et al.*, 2012). The distribution of calculated log2 fold changes between CLL patients and non-malignant control was fitted by the sum of three Gaussian functions representing regions with gained, lost and unchanged accessibility between the two conditions. The indicated upper and lower log2 fold thresholds were then determined as the x-value of the intersection points between the gained/lost distributions and the distribution representing the unchanged regions. **C** Accessibility distribution at a genomic location around the *CTLA4* gene containing regions, which gained accessibility in CLL compared to non-malignant CD19$^+$ B-cells. **D** Same as C for the *EBF1* locus, which harbors sites with lost accessibility in CLL.

To gain insight into the function of the identified sites I next aimed to functionally annotate the differentially accessible regions. For this I used an annotation of the genome, which assigns each genomic location to one of twelve functional chromatin states based on ChIP-seq data of seven histone modifications (Mallm *et al.*, 2019). This annotation was generated by Naveed Ishaque (Division of Theoretical Bioinformatics and Heidelberg Center for Personalized Oncology, German Cancer Research Center (DKFZ), Heidelberg, Germany) with the Hidden Markov Model based tool ChromHMM (Ernst & Kellis, 2012) from ChIP-seq datasets of the same patients and healthy controls as used for the ATAC-seq analysis. According to the ChromHMM annotation ~12,000 regions were assigned to a state linked to transcription/TSS, enhancers or repressed (Figure 24 A). The state with the largest fraction of regions assigned to were enhancers (primed) with 41 %. If all enhancer predictive states were included, e.g., genic enhancers, this fraction even increased to 58 % (state 1, 8, 9, 11), while 8 % of regions were annotated as active TSS or transcribed (state 2, 10).

I then determined how many ATAC-seq peaks located at potential active enhancer loci were present in at least three CLL patients or one pool of healthy individuals. For the definition of potential active enhancers, I used regions annotated by ChromHMM as state 9 (H3K4me1 and H3K27ac) or state 1 (H3K4me1, H3K27ac and H3K36me1) (Mallm *et al.*, 2019), which are thus prone to contain active enhancers. About 5,300 enhancer regions, which were not present in any healthy control pool, *de novo* gained accessibility in at least three CLL patients (Figure 24 B). Around 2,500 predicted enhancers, which were present in at least one pool of healthy individuals, were not identified in the majority (>89 %) of the 19 CLL patients. Bidirectional transcription is another feature associated with active enhancers (Andersson *et al.*, 2014; Kim *et al.*, 2010). Notably, ChromHMM-predicted active enhancer containing sites which were bidirectionally transcribed (another mark associated with active enhancers; labeled as “Bidi”) in at least three CLL patients or two healthy controls seemed to be a subgroup of enhancers with ATAC-seq signal. In summary, these excessive chromatin accessibility changes in enhancer regions of CLL patients compared to non-malignant B-cells points to a deregulated enhancer landscape in CLL.

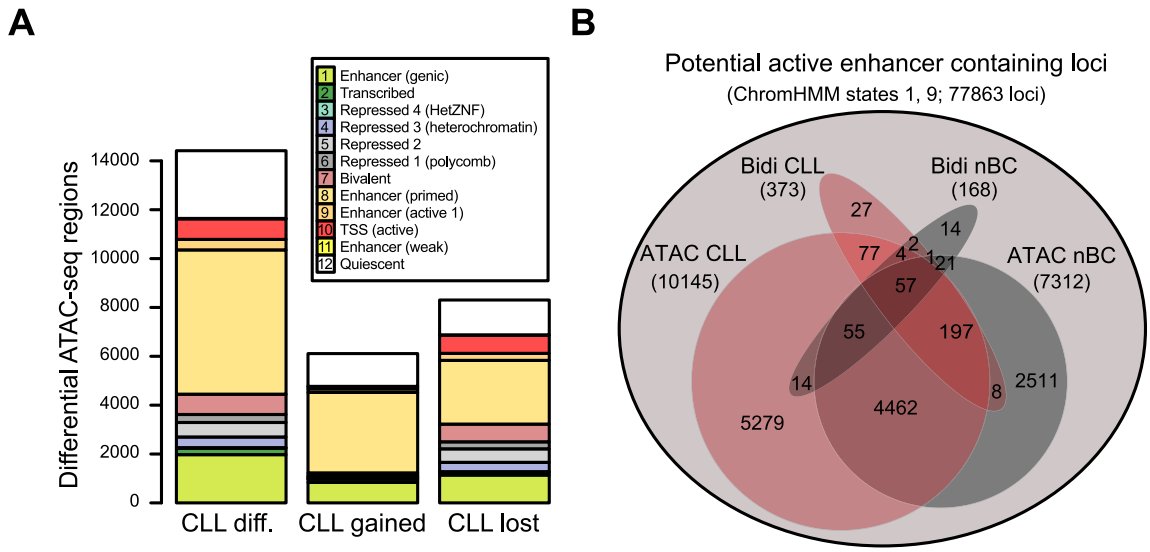


Figure 24 Enhancer deregulation in CLL compared to non-malignant CD19⁺ B-cells.
A Distribution of chromatin state annotation is shown for all consensus ATAC-seq peak regions showing a significant change of accessibility ("CLL diff") in CLL patient cells compared to CD19⁺ B-cells of non-malignant donors and split into regions which significantly gained ("CLL gained") or lost ("CLL lost") accessibility in CLL. **B** Number of ATAC-seq peaks or regions exhibiting bidirectional transcription ("Bidi") within potential active enhancer elements containing genomic regions (locations assigned to histone ChIP-seq based ChromHMM annotation states 1 and 9) found in at least three out of 19 patients ("CLL") and one (ATAC-seq) or two (Bidi) pools of non-malignant CD19⁺ B-cells from healthy donors ("nBC"). Areas are not proportional to numbers of regions. Bidirectionally transcribed loci ("Bidi") were identified by Sandra Koser (Division of Applied Bioinformatics, German Cancer Research Center (DKFZ), Heidelberg, Germany). Figure adapted from Mallm *et al.* (Mallm *et al.*, 2019).

3.1.2 Motif enrichment analyses in differentially accessible regions identify aberrant transcription factor binding in CLL

To further elucidate the molecular mechanism driving the enhancer deregulation in CLL, I analyzed the enrichment of transcription factor binding motifs in the differentially accessible regions between CLL and normal CD19⁺ B-cells. The motif enrichment analysis of ATAC-seq peak regions with significant gain or loss of accessibility in CLL compared to the physiological state was done separately and restricted to the ChromHMM annotation defined regions predictive for enhancers. ATAC-seq peaks with their peak center within a region +/- 1 kb around a transcription start site (RefSeq) (O'Leary *et al.*, 2016) were excluded from the analysis. Peaks in putative enhancer regions which gained accessibility in CLL compared to non-malignant CD19⁺ B-cells showed an enrichment of binding motifs for E2A (E protein family) and NFAT (Figure 25 A). The highest enriched motif in sites with lost accessibility in CLL was the EBF/EBF1 motif. Additionally, the binding sites for NF-κB and ETS factors were enriched (Figure 25 B). I next validated the difference in the ATAC-seq signal directly at the binding motifs of selected factors. For this I focused on the ATAC-seq signal over all transcription factor motif sites of an identified factor in the peak set it was found to be enriched. The combined accessibility signal of all CLL patients in comparison to all non-malignant CD19⁺ cells was shown for

the E2A and EBF1 motifs (Figure 25 C). While the detailed pattern within the binding motif reflects the binding-preference of the transposase , the motif-flanking ATAC-seq signal confirmed the aberrant accessibility in CLL directly around the binding motifs. In addition, the occupancy by the individual transcription factor only in CLL patients (E2A) or non-malignant B-cells (EBF1) was visible by the depth of the footprint in the accessibility signal. To validate my observation of EBF1-binding loss in CLL based on ATAC-seq data, Dr. Jan-Philipp Mallm and Sabrina Schumacher conducted a ChIP-seq experiment of the transcription factor EBF1. Indeed, 826 regions lost EBF1 binding, while only 173 regions with a gain of binding were detected (Mallm *et al.*, 2019). An EBF1 binding site in an enhancer region in the *SNX22* gene was for instance detected to have lost accessibility as well as lost EBF1 binding (ChIP-seq) in CLL, while the enhancer connected gene lost its expression (Figure 25 D).

CTCF was implicated to play a role in the formation of topologically associated domain (TAD) structures and thus enhancer-promoter interactions (Dekker & Mirny, 2016). ATAC-seq is able to detect CTCF-binding changes (Buenrostro *et al.*, 2013). Therefore, the presence of CTCF binding sites in the differentially accessible regions between CLL and normal B-cells was analyzed. 1,525 (18 %) sites with lost accessibility in CLL contained a CTCF site according to the ENCODE annotation based on ChIP-seq peaks of the lymphoblastoid cell line GM12878 (Encode Project Consortium, 2012), while this was only true for 330 (5 %) sites with gained accessibility. This indicated a loss of CTCF binding, which could also be confirmed by a footprinting analysis (Figure 25 E). A subsequent CTCF ChIP-seq experiment also found 5,964 gained CTCF binding sites, while only 441 were lost (Mallm *et al.*, 2019). In summary, the detected enhancer regions with differential accessibility between CLL cells and non-malignant $CD19^+$ B-cells could be linked to deregulated transcription factors which in turn influenced gene expression patterns.

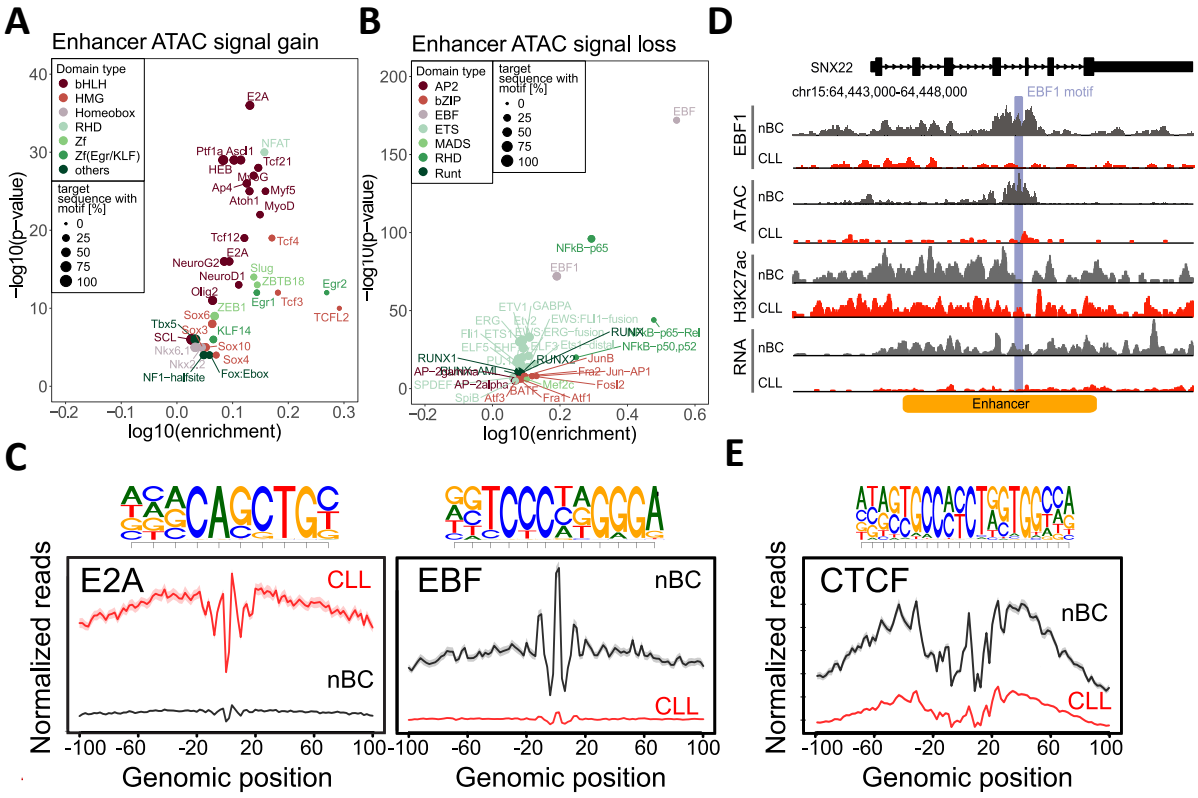


Figure 25 Transcription factors linked to chromatin accessibility changes in patient-derived CLL cells compared to CD19⁺ B-cells of non-malignant donors.
Enrichment of transcription factor binding motifs determined by the HOMER package from known motifs (Heinz *et al.*, 2010) in ATAC-seq peaks with **(A)** gained and **(B)** lost accessibility in CLL, which overlap with any region annotated as putative enhancer containing (states 1, 8, 9, 11) by a ChromHMM genome annotation (Mallm *et al.*, 2019) based on histone ChIP-seq of the same patients. Color represents HOMER supplied "motif type". **C** The aggregated normalized accessibility signal of all CLL patients or non-malignant CD19⁺ B-cells (nBC) at binding site instances of the indicated factor in regions they were found to be enriched (left: gained in CLL; right: lost in CLL) reveals a "footprint" in the signal. **D** Normalized accessibility and EBF1-binding (ChIP-seq) signal decreased in a CLL patient in comparison to a non-malignant individual at an EBF1 motif instance at an enhancer region in the *SNX22* gene, which lost its expression in CLL. Figure adapted from Mallm et al. (Mallm *et al.*, 2019). **D** Same as C for CTCF motif instances.

3.2 Tumor cell-of-origin mapping to distinguish CLL-specific accessibility patterns from physiological changes

The epigenome is not stable during B-cell maturation as elucidated in chapter 2. Furthermore, CLL has been shown to originate from a continuum of B-cell maturation stages (Kulis *et al.*, 2015; Oakes *et al.*, 2016). Consequently, the possibly best healthy reference would be the normal B-cell precursor from which tumorigenesis was initiated. Therefore, I set out to identify the tumor cell-of-origin for each individual CLL patient to use its chromatin accessibility state as healthy reference. This allows to further discriminate between disease-induced and physiologically-occurring changes, which is not achievable with the conventional comparison of CLL cells to all CD19⁺ B-cells of non-malignant donors.

3.2.1 Cells-of-origin are identified for individual CLL patients

The previously calculated B-cell maturation scores for potential CLL precursors (see chapter 2.2.3) were used to identify the tumor cell-of-origin for a certain CLL patient as originally proposed by Wierzbinska et al. for DNA methylation patterns (Wierzbinska *et al.*, 2020). The tumor cell-of-origin was defined as the closest virtual relative on the developmental trajectory. I then identified the tumor cell-of-origin by principal component regression for individual CLL patients from their ATAC-seq profiles (Figure 26 A). To ensure accurate inference of the cell-of-origin, only regions which changed linearly during development (n=6,762); linear regression fit with p-value < 0.05; $R^2 > 0.3$) were included in the regression model. The optimal number of principal components to build the model was determined by cross-validation. The model based on the first seven principal components was selected as it had the smallest root mean squared error of prediction (Figure 26 B, C). In fact, the maturation stage prediction for the cell-of-origin of a certain CLL patient was quite stable with varying numbers of included principal components (Figure 26 D). The final model was able to explain 91.7 % of variance.

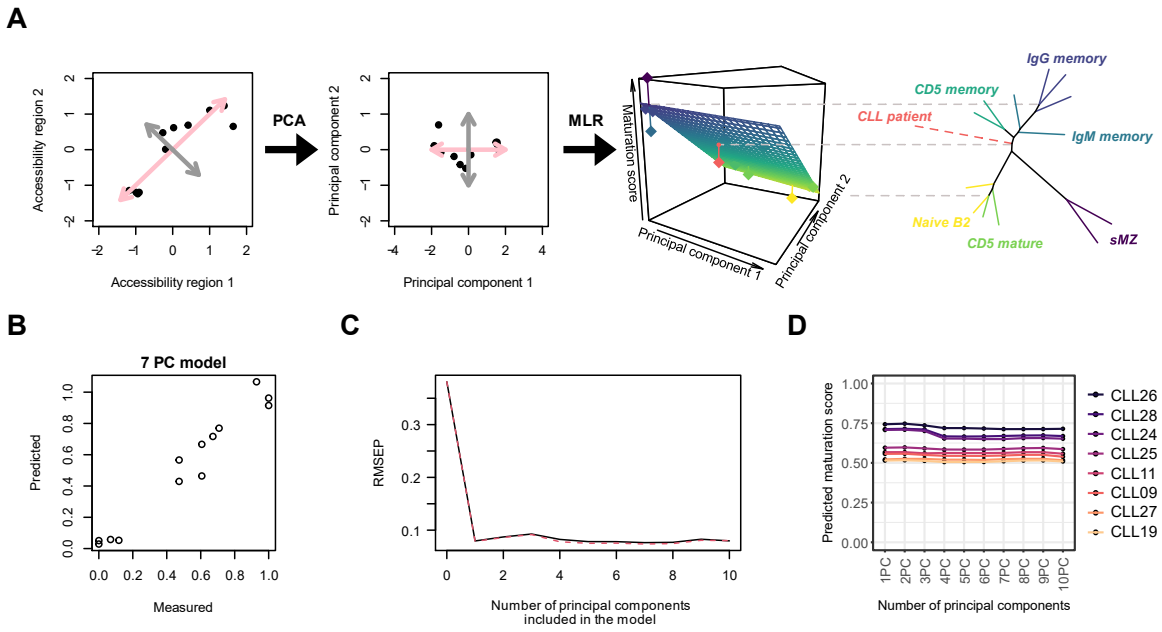


Figure 26 Principal component regression on developmentally dynamic accessible regions to determine the cell-of-origin maturation stage of individual CLL patients.
A Illustration of cell-of-origin modeling by principal component regression. Developmentally dynamic accessible sites which could be fitted by linear regression with R2 > 0.3 were used for principal component regression. Principal component regression first conducts a principal component analysis (PCA) before using the first n principal components (PCs) for multiple linear regression (MLR) to predict the maturation stage. Maturation stage prediction were visualized by labeling the cell-of-origin of a certain CLL patient on the original phylogenetic tree of B-cell development. **B** Assessment of model quality by cross-validation. The cross-validation prediction of the seven principal component regression model was compared to the known value (measured). **C** Root mean squared error of prediction (RMSEP) of cross-validation for models generated from different numbers of principal components (black line) and bias corrected RMSEP (red dashed line). **D** Dependency of maturation score prediction on number of principal components used for model generation. The number of PCs used to build the model are plotted on the x-axis, while y-axis represents the stage prediction for the cell-of-origin for the indicated patient of that model.

The final principal component regression model was applied to predict the B-cell maturation stage of the cell-of-origin of eight CLL patients based on their chromatin accessibility signatures before treatment. The predictions were in agreement with the reported IGHV mutational status of those patients, which is a clinically-relevant binary classification scheme for CLL patients based on the maturation stage of their tumor initiating cell. IGHV unmutated patients originated from a cell-of-origin in a more immature stage than the tumor initiating cell of patient 26 classified as IGHV mutated (Figure 27). Moreover, the cell-of-origin modelling allowed to further differentiate between detailed variations in developmental stages of IGHV unmutated patients.

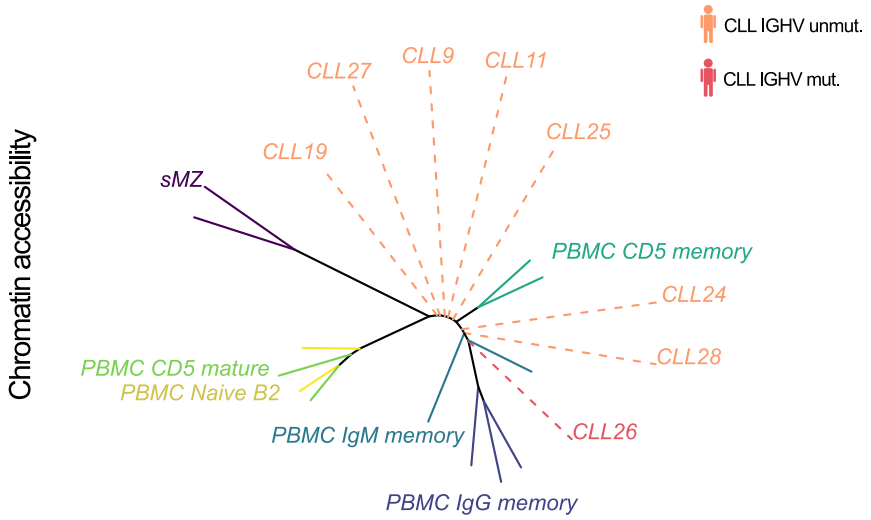


Figure 27 Assignment of eight CLL patients to their cell-of-origin on the B-cell maturation trajectory.
The base node of the dashed line which links a CLL patient to the tree denotes the maturation stage of the predicted virtual cell-of-origin. The length of the connecting line represents the square root of the sum of least squares of a patient's accessibility to its cell-of-origin inferred by linear modeling scaled by 30 for visualization.

3.2.2 Sites with CLL-specific aberrant-accessibility are identified by the comparison to the individual cell-of-origin

In addition to the epigenetic signature present in the tumor initiating cell, a tumor cell gains disease-specific epigenetic changes. The identified patient-specific cell-of-origin maturation stages could now be employed to determine these disease-specific changes. The maturation stage information allowed to compare the epigenetic signatures present in a CLL patient to the patterns which existed in its cell-of-origin, the possibly best healthy reference. In order not to miss any *de novo* accessible regions only present in CLL cells, I used a consensus peak set (n=315,747) which included in addition to the 14,377 developmentally dynamic regions all accessible regions that were identified in at least one sample. To find regions with CLL-specific aberrant accessibility, I first determined the accessibility levels of the cell-of-origin of individual CLL patients for every of those consensus regions. Accessibility levels of individual tumor cell-of-origins for a certain region were predicted from its inferred maturation stage by using the continuous modelling of accessibility for every maturation stage from naïve to memory B-cells. Linear modelling per region was repeated for all consensus regions also including those with stable accessibility levels during B-cell maturation. A region was specified as having aberrant accessibility in a CLL patient when the patient's accessibility level of that region was outside the 95 % confidence interval of the prediction for its corresponding cell-of-origin (Figure 28). Further, a

CLL-specific up or downregulation was defined as having aberrant accessibility levels in at least six out of eight CLL patients (75 %).

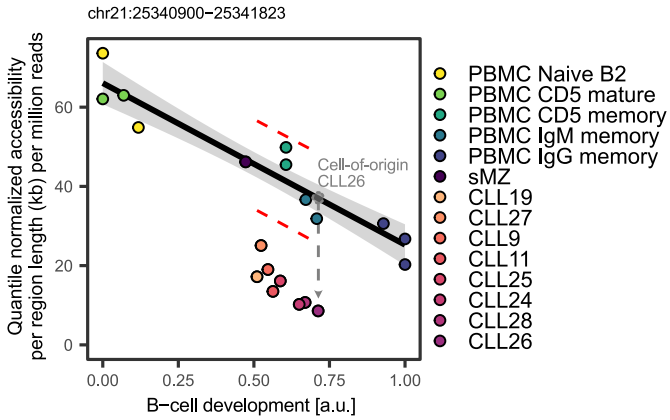
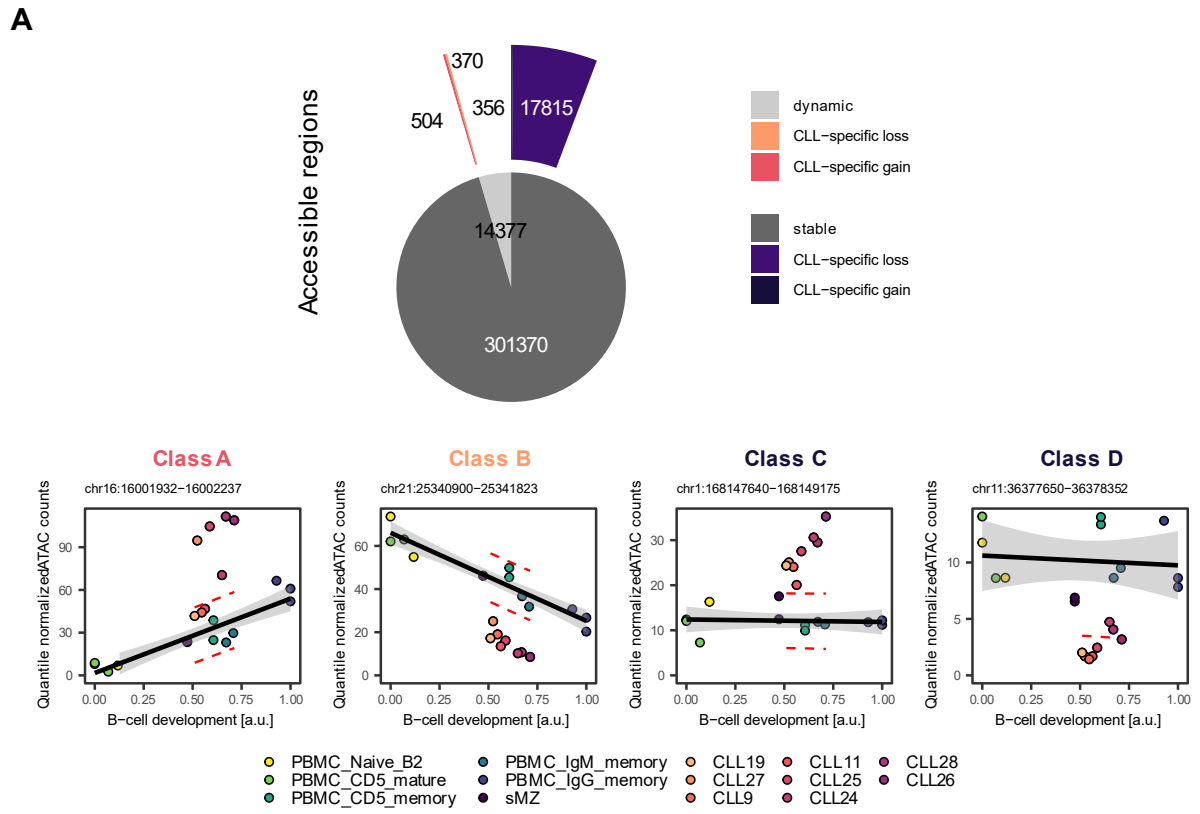
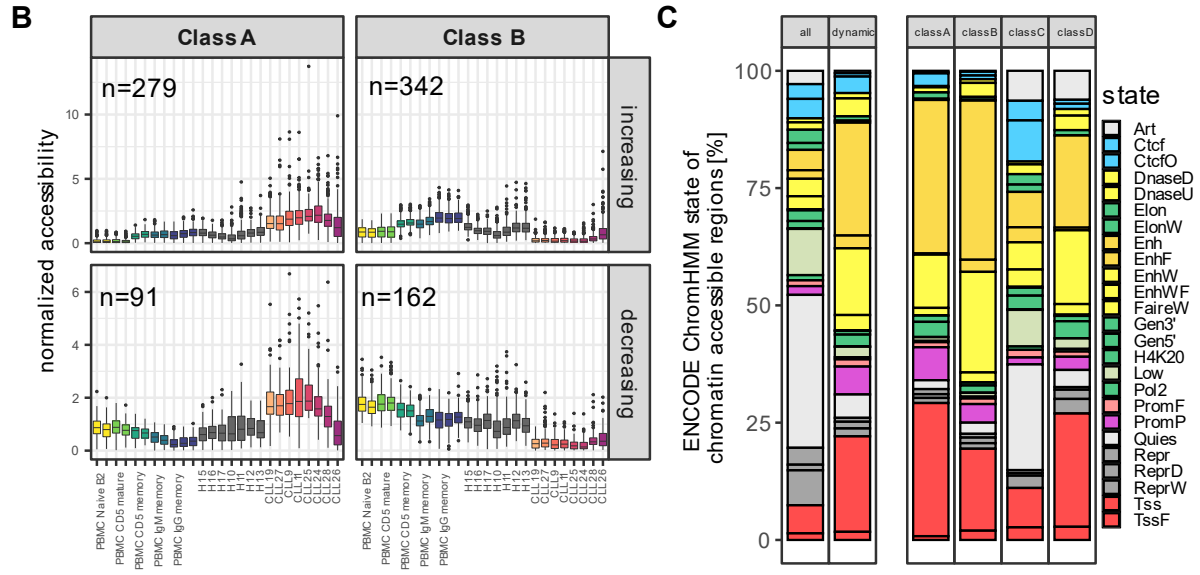
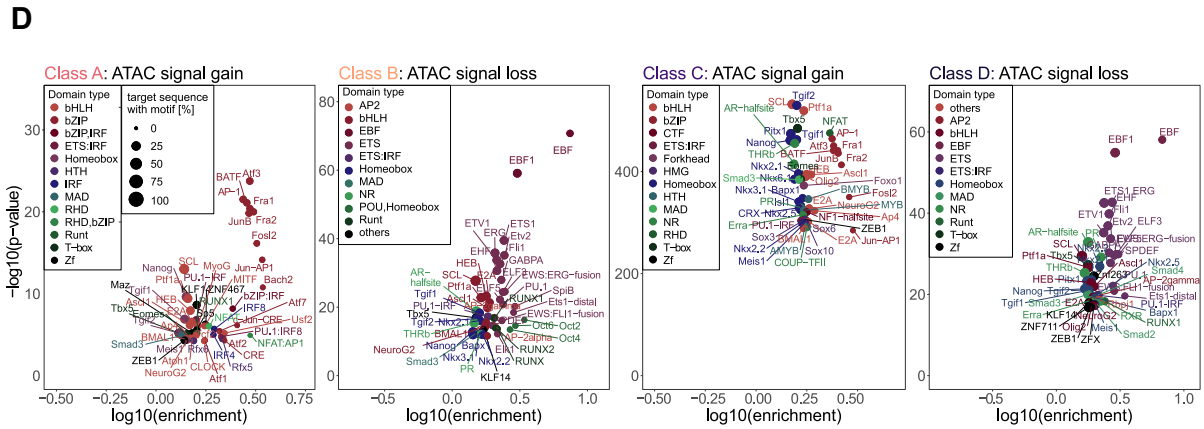


Figure 28 Illustration of definition of a region with CLL-specific aberrant accessibility.
Example for a region at which the accessibility levels of all CLL patients were outside of the 95 % confidence interval (CI) of the prediction for their corresponding cell-of-origin. Normalized accessibility signal of healthy B-cell subpopulations and CLL patients for this region on chromosome 21 is plotted against their B-cell developmental status calculated from phylogenetic tree construction and principal component regression. Red dashed lines indicate 95 % CI for accessibility prediction of the cell-of-origin for this region by linear regression (black line). The predicted cell-of-origin of patient CLL26 is exemplary illustrated. Grey area marks the standard error of regression line (95 % confidence level).

Sites with disease-induced gain or loss of chromatin accessibility in comparison to the epigenetic pattern present in the tumor initiating cell were identified from developmentally dynamic and stable regions. The 22,219 regions with aberrant chromatin accessibility in CLL were categorized into four classes according to their behavior during physiological developmental (dynamic or stable) and the direction of the disease-induced change (gain or loss) (Figure 29 A). The two classes of developmentally dynamic regions with either a CLL-specific gain (class A) or a CLL-specific loss (class B) of accessibility compared to the CLL patient's cell-of-origin contained between 350 and 500 sites each (Figure 29 A). Sites with stable accessibility during development and an increase in accessibility (class C) were the most frequent with 17,815 sites. 356 developmentally stable sites show a lower accessibility in CLL (class D) (Figure 29 A). Class A and class B regions were further subdivided into developmentally increasing or decreasing (Figure 29 B). The identified regions indeed showed an increased (class A) or decreased (class B) accessibility compared to phylogenetically closest distinct B-cell subsets (CD5 memory; IgM memory). The identified regions included TSS of genes with known association to disease biology. The TSS (RefSeq) of *GAB1*, which is implicated in B-cell receptor signaling (Mraz *et al.*, 2014), is for instance contained in a class A region (data not shown). In summary, the modelling of a patient's cell-of-origin was employed to identify regions with CLL-specific aberrant accessibility excluding confounding effects of physiological changes, which occur during normal B-cell maturation.


A
Accessible regions
370
356
17815
504
14377
301370
dynamic
CLL−specific loss
CLL−specific gain
stable
CLL−specific loss
CLL−specific gain
Class A
Class B
Class C
Class D
chr16:16001932−16002237
chr21:25340900−25341823
chr1:168147640−168149175
chr11:36377650−36378352
Quantile normalized ATAC counts
B−cell development [a.u.]
PBMC_Naive_B2
PBMC_CD5_mature
PBMC_CD5_memory
PBMC_IgM_memory
PBMC_IgG_memory
sMZ
CLL19
CLL27
CLL9
CLL11
CLL25
CLL24
CLL28
CLL26


B
Class A
Class B
n=279
n=342
n=91
n=162
increasing
decreasing
normalized accessibility
C
all
dynamic
classA
classB
classC
classD
ENCODE ChromHMM state of chromatin accessible regions [%]
state
Art
Ctcf
CtcfO
DnaseD
DnaseU
Elon
ElonW
Enh
EnhF
EnhW
EnhWF
FaireW
Gen3'
Gen5'
H4K20
Low
Pol2
PromF
PromP
Quies
Repr
ReprD
ReprW
Tss
TssF


D
Class A: ATAC signal gain
Class B: ATAC signal loss
Class C: ATAC signal gain
Class D: ATAC signal loss
−log10(p−value)
log10(enrichment)
Domain type
target sequence with motif [%]

◀ **Figure 29 Characterization of sites with CLL-specific aberrant accessibility.**
A Frequencies and classification of regions with CLL-specific aberrant chromatin accessibility. Top: Vennpie diagram shows proportion of consensus accessible sites (accessible in at least one CLL patient or healthy subpopulation) with dynamic and stable chromatin accessibility during physiological B-cell maturation and CLL-specific gain or loss of accessibility. Bottom: Normalized accessibility during maturation at exemplary regions of every class with linear modeling of physiological accessibility (black line). Grey area marks the standard error of regression line (95 % confidence level). **B** Boxplots of chromatin accessibility per sample of all class A and class B regions splitted according to their developmental increase (top) or decrease (bottom). Normalized accessibility was divided by the region's mean count over all samples to adjust for differences in the accessibility levels between regions. **C** Annotation of different region subsets from the consensus region set ("all") based on a ChromHMM genome segmentation by ENCODE for the B-cell derived cell line GM12878. F: flanking; W: weak; P: poised; Art: Heterochromatin-Repetitive; Ctcf/CtcfO: Distal CTCF-Candidate Insulator; DnaseD/ DnaseU: Candidate Weak enhancer-DNase; Elon: Transcription associated; Enh: Candidate Strong enhancer; EnhW/FaireW: Candidate Weak enhancer-DNase; Gen3/Gen5/H4K20/Pol2: Transcription associated; Low: Low activity proximal to active states; PromF: Promoter Flanking; PromP; Inactive Promoter; Quies: Quiescent; Repr: Polycob repressed; Tss: Active Promoter. **D** Top hits of transcription factor motif enrichment analysis in regions of the indicated class. Transcription factors are colored according their transcription factor family. Dot size represents percentage of regions with the motif.

3.2.3 CLL-specific aberrant open sites are enriched in promoters and enhancers and can be linked to transcription factor binding

I next aimed to characterize the genomic position of the previously identified regions and to identify transcription factors potentially involved in this epigenetic deregulation in CLL. To evaluate the regulatory activity and functional relevance of these genomic regions in normal B-cells, I used a ChromHMM state genome annotation derived from the lymphoblastoid cell line GM12878 from ENCODE (Encode Project Consortium, 2012). Regions with dynamic accessibility during B-cell development were enriched for enhancers (orange/yellow) and TSS (red) compared to all consensus regions (Figure 29 C). Class A and B regions, which are subsets of the dynamic regions, had a similar distribution to regulatory annotations. Class D regions, which were identified from developmentally stable regions, showed a state distribution nearly identical to class A and B, while class C regions were often annotated at "quiescent". This is reasonable for class C sites (defined as regions with CLL-specific gain in accessibility) as the annotation is based on the physiological epigenetic state. It indicates that a group of regions exists which gains accessibility *de novo* in CLL.

The observed accessibility changes were further linked to the binding of specific transcription factors. To achieve this, a transcription factor motif enrichment analysis was conducted separately for regions of every class (Figure 29 D). The most enriched transcription factor motif in the two classes with a CLL-specific loss of accessibility (class B, class D) was EBF1. EBF1 is a factor also involved in physiological B-cell development (Satpathy *et al.*, 2019). The EBF1 motif enrichment in regions with CLL-specific aberrant accessibility (developmental changes excluded as confounding factor) confirmed the downregulated activity of EBF1 in CLL. Furthermore, the motif of ETS transcription factors was also enriched in regions with CLL-specific loss of accessibility. Regions with gained signal (class A,

class C) showed an enrichment of JUN/AP1 factors, NFAT and E2A. Besides, a BACH2 motif enrichment in class A regions was identified. BACH2 is known to inhibit plasma cell differentiation and was reported to be upregulated in CLL (Seifert *et al.*, 2012). Although sharing the core motif with JUN factors, BACH2 motif is distinct by three additionally conserved base pairs (Heinz *et al.*, 2010). In summary, regions with CLL-specific aberrant accessibility were often located at annotated enhancers and promoters. The deregulation of enhancers and promoters was linked to involved transcription factors like EBF1 and BACH2 with importance for physiological B-cell development.

3.3 Multi-omics analysis of CLL patient response to ibrutinib

Deregulated epigenetic signaling can potentially be modulated by drug treatment. Within the new target therapy options, the treatment with the Bruton's tyrosine kinase (BTK) inhibitor ibrutinib holds great potential for CLL patients including those harboring unfavorable prognostic factors. However, resistance to ibrutinib might develop. Genetic intratumor heterogeneity is frequently observed in cancers and has an important impact on tumor progression and the response to treatment. Although targeted, the molecular intervention of drugs can affect various molecular layers including the transcriptome, the epigenome and the genetic clonal composition. Therefore, the effect of ibrutinib in CLL patients during ibrutinib treatment and in a case of acquired resistance was analyzed on multi-omics levels (Figure 30). I evaluated the impact of ibrutinib on the observed CLL-specific deregulation of chromatin accessibility patterns with regard to their potential reversibility in five CLL patients (see Figure 19). Further, I assessed shifts in clonal composition by targeted single cell (sc)DNA-sequencing and the induced effects on gene expression on the single cell level during treatment and relapse.

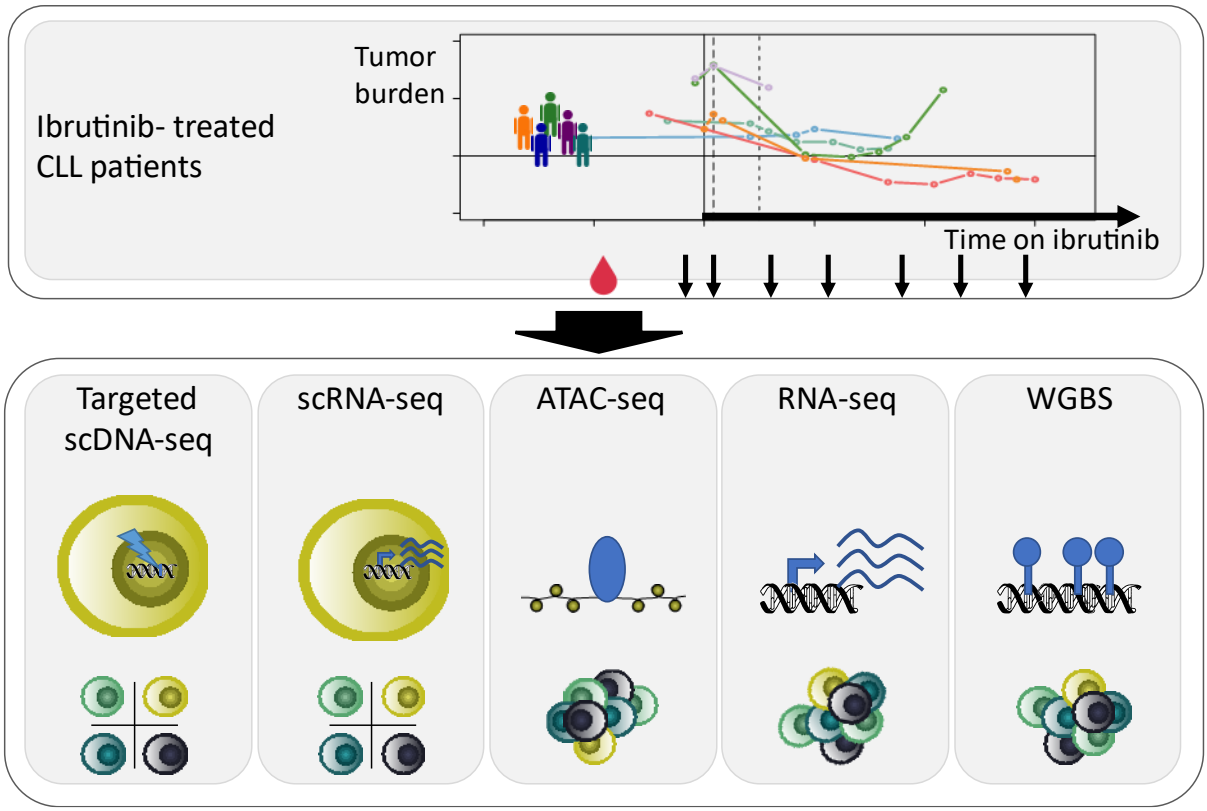


Figure 30 Overview of multi-omics analysis of CLL patient's response to ibrutinib therapy. CLL cells isolated from peripheral blood draws of five CLL patients at several timepoints during ibrutinib monotherapy were subjected to a multi-omics analysis. Each read-out was conducted for a selection of timepoints (see Material and Methods chapter 1.7). Tumor burden was monitored in leukocyte counts per milliliter at every timepoint (top). Whole-genome bisulfite sequencing (WGBS) experiments and data analysis was not conducted by myself and is not covered in this thesis.

3.3.1 Ibrutinib partially reverses the disease-induced chromatin state deregulation

The effect of ibrutinib on chromatin is especially interesting at regions which are deregulated in disease. Therefore, I examined the effect of ibrutinib treatment on the CLL-specific aberrant accessible regions. Class A and C regions exhibited a tendency to lose accessibility signal with duration of treatment, while class B and D regions rather tended to gain accessibility in individual patients (Figure 31). This suggested a potential reversibility of disease induced patterns by ibrutinib. To quantify this observation, I determined the genome-wide accessibility changes induced by ibrutinib (=ibrutinib-responsive accessible regions). 4,253 of all possible accessible genomic regions showed a consistent gain of accessibility in all five analyzed ibrutinib-treated patients, while 7,281 regions showed an ibrutinib-induced loss of accessibility. The fraction of sites with a disease-specific deregulation that can be reversed by ibrutinib was assessed by calculating the overlap of those ibrutinib-responsive accessible regions with regions of all four classes of CLL-specific deregulated regions. About 10 % of regions with CLL-specific changes approached physiological levels with ibrutinib treatment. For regions of class A (CLL-specific accessibility gain at developmentally dynamic sites) a higher fraction of 20 % of sites show a significant (adjusted p-value: 0.05) reversion of the accessibility signal with ibrutinib. In contrast the fraction of all class A-D regions with an enhancement of the disease phenotype with ibrutinib was small with ~1 %. In conclusion, the accessibility in a fraction of regions with CLL-specific aberrant accessibility approximated physiological levels during ibrutinib treatment. This showed that disease-induced epigenetic changes were partially reversible. Ibrutinib only indirectly influences epigenetic signatures. Direct chromatin modifying drugs might be employed to more specifically target aberrant epigenetic patterns in disease.

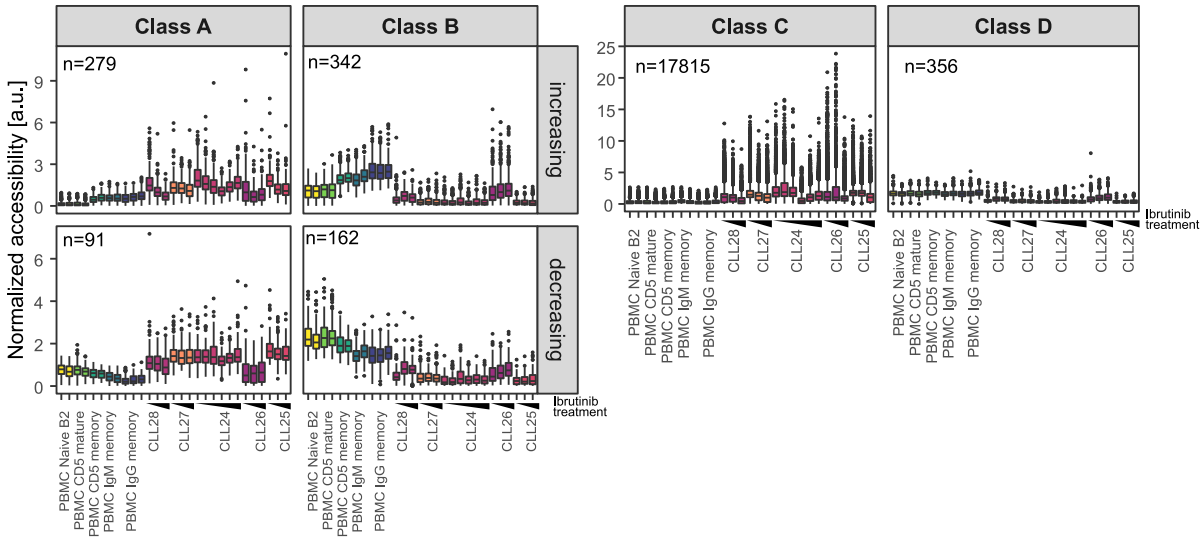


Figure 31 Effect of ibrutinib treatment on regions with CLL-specific aberrant accessibility. Boxplots show distribution of chromatin accessibility in a sample separated by treatment timepoint for all regions of the indicated class. Normalized accessibility of a regions was divided by its mean accessibility over all samples. Black triangles indicate duration of treatment.

3.3.2 Tumor evolution in ibrutinib-treated CLL patients is resolved by targeted single cell DNA sequencing

A malignant cell population is often not homogeneous and consists of several subclones. The composition and frequency of those subclones can change, especially under treatment and during resistance development (Gutierrez & Wu, 2019). Therefore, I determined the evolution of the clonal composition of CLL cells in the peripheral blood of three CLL patients during ibrutinib treatment using targeted single-cell DNA (scDNA)-sequencing (Tapestri, Mission Bio). Patient CLL24 was reported to have acquired a *BTK* mutation, which leads to the resistance to the BTK inhibitor ibrutinib. After relapse, the presence of a BTK mutation (C481S c>g) had been verified by Sanger sequencing 25 months after treatment start (Department of Internal Medicine III, University Hospital Ulm, Ulm). Using targeted scDNA-sequencing, the resistant clone with mutated *BTK* was already detectable in the sample taken before relapse of the patient in 4.6 % of cells (Figure 32 A). At treatment start none of the analyzed cells showed a *BTK* mutation. The probably during therapy newly developed resistant subclone originated from a clone with both an *ATM* (chr11:108,183,167:A/G) and a *SF3B1* (chr2:198,266,834:T/C) mutation, but no *BRAF* mutation at position chr7:140453134:T/C. The *BTK* mutated clone expanded till it accounted for more than 90 % of all analyzed CLL cells from the blood of patient 24 (Figure 32 B, D), which was accompanied by a vast increase in tumor burden (Figure 32 A). A small subclone of the *BTK* mutated clone additionally acquired an *NRAS* mutation (~ 4 %). However, the resistant subclone with the co-occurring *NRAS* mutation seemed not to have an additional advantage over the *NRAS* wildtype clone. Mutational co-occurrence was also an important factor in the clonal dynamics of a second CLL patient. The variant allele frequency of the *TP53* mutation at position chr17:7,579,472:G/C in patient CLL25 was nearly constant over the observed treatment period. The same held true for the only slightly varying prevalence of a *ZMYM3* mutation (chrX:70468672:G/GCA) of the same patient (Figure 32 E). The detection of the co-occurrence of several mutations in individual single cells by scDNA-sequencing revealed, however, a clonal dynamic in patient CLL25. As of one month after treatment start a *TP53*/*ZMYM3* double mutant subclone was the most prevalent clone (>55 %) in the blood of patient CLL25 (Figure 32 C). This clone was already present before treatment start, but with a far lower prevalence (~7 % of cells). Ibrutinib is known to induce a flush of CLL cells from lymph nodes into peripheral blood, which results in an initial increase in tumor burden in the blood (Burger & O'Brien, 2018; de Gorter *et al.*, 2007). This was also observed for patient CLL25 (data not shown). The increase in tumor burden (1 month after treatment start) coincided with the change in clonal frequencies. Instead of a selective effect of ibrutinib on single subclones, the increase of *TP53*/*ZMYM3* double mutant subclone prevalence could therefore also originate from differences in the clonal composition of CLL cells between blood and lymph nodes. The mutant *TP53*/*ZMYM3* clone might have been predominantly located in lymph nodes. The third analyzed patient CLL28 showed no major clonal dynamics between the two analyzed timepoints (5 and 12 month) after ibrutinib treatment during which tumor burden was also constant (data not shown).

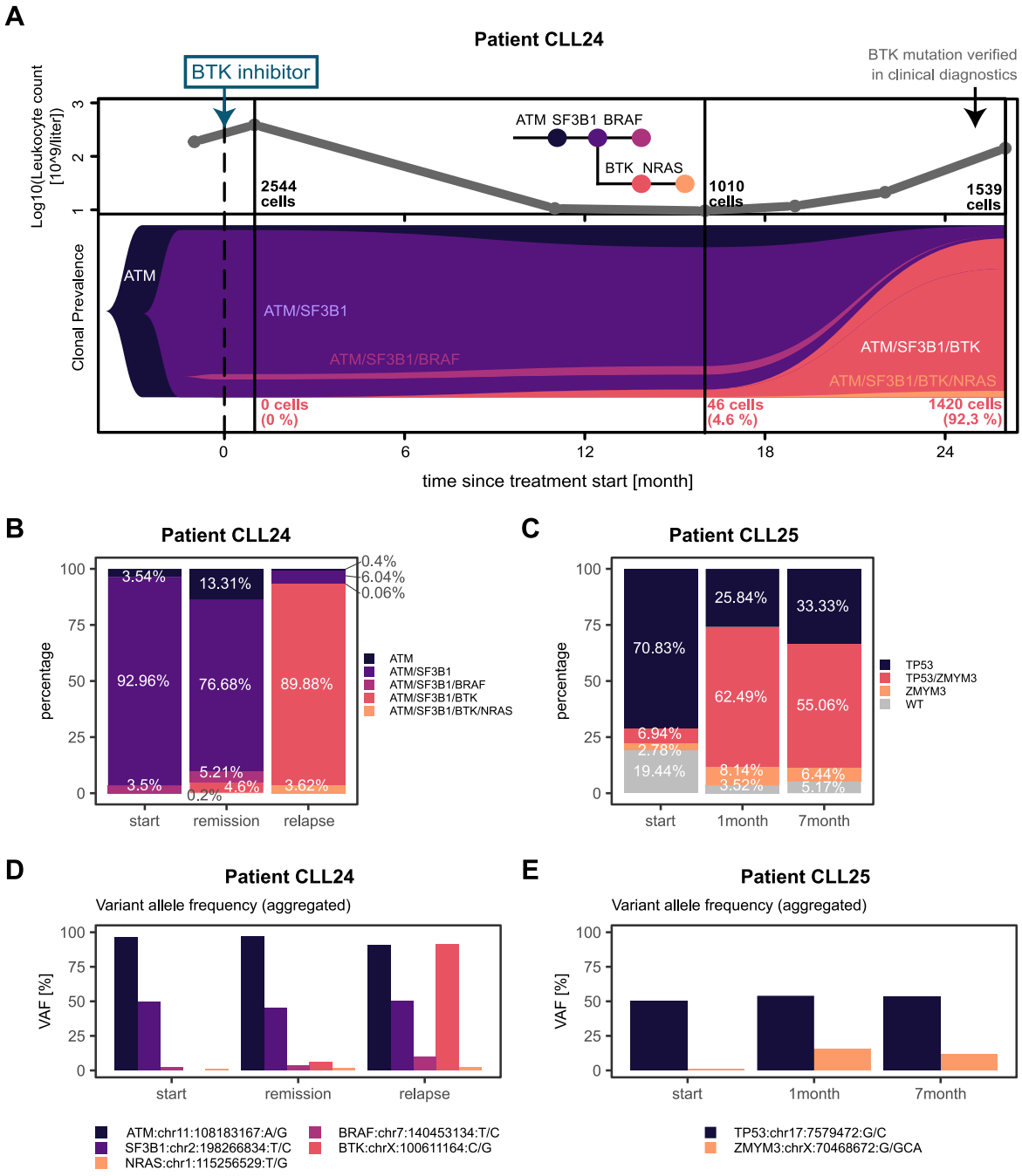


Figure 32 Clonal composition changes during ibrutinib treatment of two CLL patients.
A Top: Tumor load of a CLL patient (CLL24) during ibrutinib treatment. Bottom: FISH-plot shows the clonal composition of CLL cells in the peripheral blood of the patient before ibrutinib treatment start, during remission and at relapse. Clonal prevalence was detected by targeted DNA-sequencing of single cells with the Tapestri instrument (Mission Bio). Cells were only included in the analysis if all indicated loci were detected in that cell. Ibrutinib (BTK inhibitor) treatment start and timepoint of first detection of the patients' *BTK* mutation by clinical diagnostics are indicated with an arrow. **B** Fractions of subclones and their mutational co-occurrence at three timepoints during ibrutinib treatment in patient CLL24. **C** Same as B in a second CLL patient (CLL25). **D** Aggregated variant allele frequencies (VAF) calculated from single cell data for patient CLL24. **E** Same as D for patient CLL25. Gene names in panels A-C represent distinct homozygous or heterozygous single nucleotide substitutions (see legend D and E) within the indicated gene.

In summary, targeted scDNA-sequencing was employed to identify small subclones and resolve mutational co-occurrences in individual cells during ibrutinib treatment of CLL patients. This enabled the detection of resistant subclones and their mutational co-occurrences at an early timepoint in patient CLL24. The possibility to detect mutational co-occurrences further allowed to identify the ibrutinib-induced increase in the prevalence of a double mutant clone in a second CLL patient. These results show the value of scDNA-sequencing for the early detection of resistant subclones and their patient-specific molecular vulnerabilities determined from co-occurring mutations, which holds the potential to open new treatment options.

3.3.3 Gene expression programs change dynamically in a CLL patients under ibrutinib treatment with acquired resistance

After dissection of the clonal evolution during ibrutinib treatment and resistance development of CLL patient CLL24, the effect of ibrutinib in this patient was further studied with respect to transcriptomic changes. The analysis was conducted on the single cell level with a droplet-based approach to account for the observed genetic clonal heterogeneity. In total 11,251 cells from five timepoints passed the quality filtering criteria. The filtered cells had a median of 3,899 detected UMI count per cell and 1,344 detected genes. I identified six clusters of cells from all timepoints by their transcriptome signature in a Seurat analysis (Stuart *et al.*, 2019). Clusters of cells were visualized in a two-dimensional representation using uniform manifold approximation and projection (UMAP) (Figure 33 A). T-cell (*CD3D, CD3E*) and monocytes (*CD14*) clusters were manually assigned on the basis of established markers and contain cells from various timepoints. The three CLL clusters contained primarily cells from only one timepoint (Figure 33 A, B) except for cluster C2, which comprised cells of both the timepoint shortly after treatment initiation (1 month) and the relapse timepoint. All timepoints form distinct populations with some overlapping cells especially between the before treatment and the two relapse timepoints comprised in clusters C2 and C3. I then determined the 10 most important marker genes for each cluster from Figure 33 A and visualized their expression in a heatmap (Figure 33 C). The genes well separate the clusters and contain cell type defining markers for non-malignant clusters (*CD3D* and *LST1, FTL*)(Hay *et al.*, 2018). Cells before treatment were characterized by *ID3* expression, which is a described target for repression of EBF1 in normal B-cell development (Thal *et al.*, 2009). The high expression of *ID3* thus fits to the loss of EBF1 activity found in the ATAC-seq analysis. *ID3* expression was lost in all ibrutinib treated cells of patient CLL24 (Figure 33 C). *CXCR4* and B-cell signaling related *FOXP1* (Mraz *et al.*, 2014) were detected as markers for cluster C2, which mostly contains cells from both very early during treatment and relapse, and was implicated in BTK inhibition induced defects in homing to spleen (Chen *et al.*, 2016). *CXCR4* is only moderately expressed before treatment, but increased shortly (1 month) after treatment initiation (cluster C2). The increase was not maintained but drastically reduced till early relapse (cluster C4 and C3). This was followed by a high expression in relapse (cluster C2) except for some cells present in cluster C3 with low *CXCR4* and low

FOXP1 expression. This dynamic was confirmed in bulk RNA-seq data of the same timepoints (data not shown). This increase at 30 days of ibrutinib treatment has also been observed in another study in peripheral blood cells of CLL patients, but not in lymph nodes (Palma *et al.*, 2018). The initial rise in expression was however not present in two other CLL patients which I analyzed by bulk RNA-seq. They exhibit a general decrease of *CXCR4* with ibrutinib treatment. Cells during remission (CLL cluster C4) were characterized by a higher expression of several mitochondrial genes of the energy metabolism including ND1 and ND2 (top 20 genes; not shown). The downregulation of mitochondrial genes is often observed in tumors (Reznik *et al.*, 2017). Cells in remission also expressed ribosomal proteins at high levels including *RPS17*, which is often the target of mutations in AML. Ribosomal proteins are involved in various biological functions besides protein synthesis, which includes cell cycle arrest and apoptosis but also proliferation (Xu *et al.*, 2016).

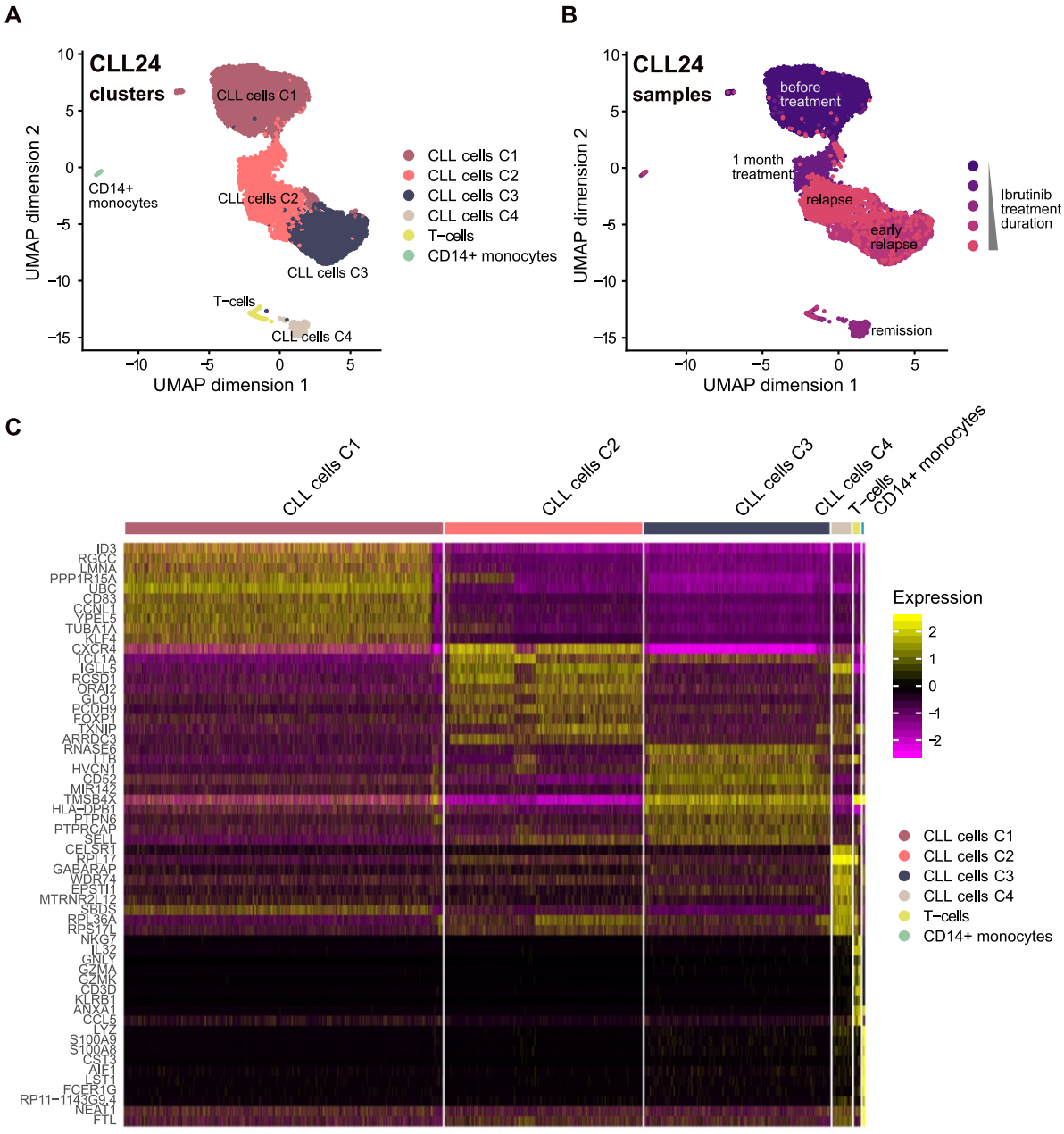


Figure 33 Gene expression dynamics on the single cell level in a CLL patient developing resistance under ibrutinib treatment.
UMAP embedding of transcriptome signatures of cells from five timepoints during ibrutinib treatment of patient CLL24 that develops resistance due to a *BTK* mutation, which is first detectable in a view cells from peripheral blood at the remission timepoint colored by **A** cluster or **B** sampling timepoint. **C** Heatmap of gene expression levels of the top ten maker genes of each cluster.

4 Response to the chromatin modifying drug panobinostat in multiple myeloma cells

Disease-induced chromatin state changes are potentially reversible. The mutant IDH1 inhibitor was able to partially revert the histone hypermethylation phenotype observed in IDH1 mutated AML cells (see Results chapter 1.3). Similarly, ibrutinib treatment of CLL showed the potential to revert deregulated epigenetic patterns at regulatory elements (see Results chapter 3.3.1). One class of drugs that directly targets aberrant epigenetic modifications are histone deacetylase inhibitors (Cai *et al.*, 2015). This class includes the histone deacetylase (HDAC) inhibitor panobinostat, which was studied here. This HDAC inhibitor was recently approved by the FDA for the therapy of multiple myeloma (MM) patients (Cai *et al.*, 2015). Therefore, I evaluated the effect of panobinostat on the chromatin state of both MM cell lines representing distinct clinical subgroups of MM (n=3) and MM patients (n=4).

4.1 Chromatin changes during panobinostat treatment

4.1.1 The histone deacetylase inhibitor panobinostat induces global histone hyper-acetylation in multiple myeloma cell lines

The response to panobinostat was evaluated in the three multiple myeloma cell lines KMS-11, MM1.S and U266. Panobinostat is a histone deacetylase (HDAC) inhibitor. First, the effect on global histone acetylation levels was analyzed by immunoblotting. The cell lines were treated with 8 nM panobinostat for 24 hours to 96 hours and changes in global acetylation levels of histone H3 at lysine residue 27 in relation to GAPDH levels were measured (Figure 34 A). Global H3K27ac levels increased already after 24 hours treatment with 8 nM panobinostat in all three cell lines. In MM1.S and U266 cell lines H3K27ac further increased up to 96 hours. Additionally, hyperacetylation was found at histone H3 lysine residue 9 after 24 hours treatment with 8 nM panobinostat in all cell lines (Dreßler, 2017). Acetylation levels of histone H3 at different lysine residues (H3K27, H3K9) and general histone H4 acetylation were further assessed after 24 hours treatment with higher panobinostat concentrations of 2.5 µM and 5 µM by immunoblotting. Increased concentrations of panobinostat lead to a strong global hyperacetylation of histone H4 and at both analyzed lysine residues of histone H3 (H3K27, H3K9) after 24 hours (Figure 34 B). Thus, panobinostat induced global histone hyperacetylation after 24 hours treatment already at a concentration of 8 nM, which was further intensified with increased panobinostat concentration and in MM1.S and U266 also with prolonged treatment.

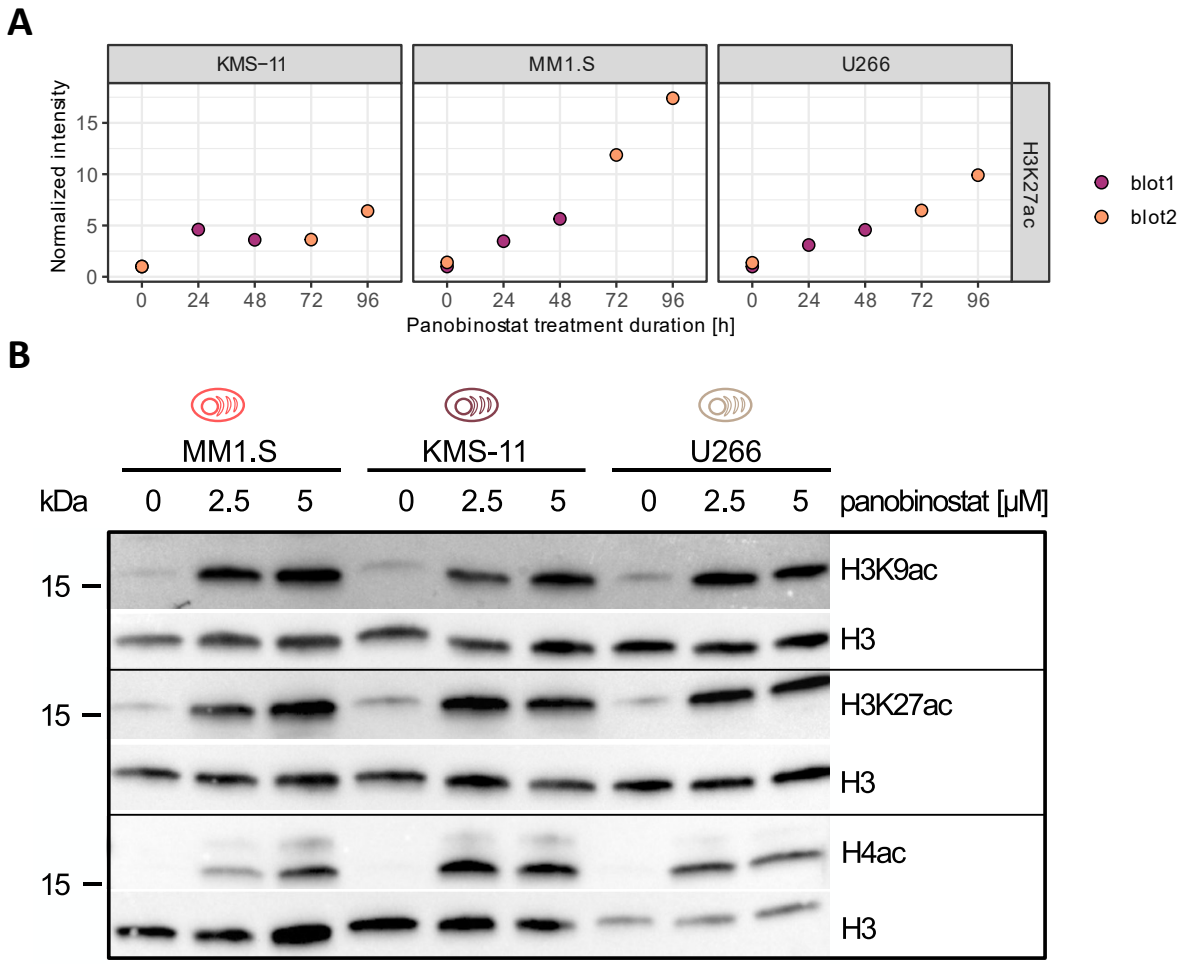


Figure 34 Global histone acetylation levels after panobinostat treatment in MM cell lines assessed by immunoblotting.
A Quantification dot plots of global histone acetylation levels of H3K27 in reference to GAPDH levels in MM cell lines (KMS-11, MM1.S, U266) treated for 0 h - 96 h with 8 nM panobinostat assessed by immunoblotting. The integrated intensity of bands was used to calculate the ratio to untreated KMS-11. The fold change between these ratios for H3K27ac and GABDH is plotted. **B** Immunoblots for global acetylation levels at histone H3 at residues K27 and K9 and at histone H4 in relation to the total amount of histone H3 in MM cell lines treated with 2.5 µM and 5 µM panobinostat for 24 hours. Experimental part of immunoblot analysis was conducted by Sabrina Schumacher and Caroline Bauer (Division Chromatin Networks, DKFZ, Heidelberg).

4.1.2 Chromatin accessibility remains unchanged at low panobinostat concentration

H3K27ac, which shows a panobinostat-induced global increase, is a mark associated with active cis-regulatory elements. Therefore, I analyzed the effect of panobinostat on the activity of regulatory elements by ATAC-seq, RNA-seq and ChIP-seq of the enhancer marks H3K27ac and H3K4me1 in MM cell lines and MM patients (Figure 35). In these experiments, cells were treated *ex vivo* with 8 nM or 5 µM panobinostat for 24 hours.

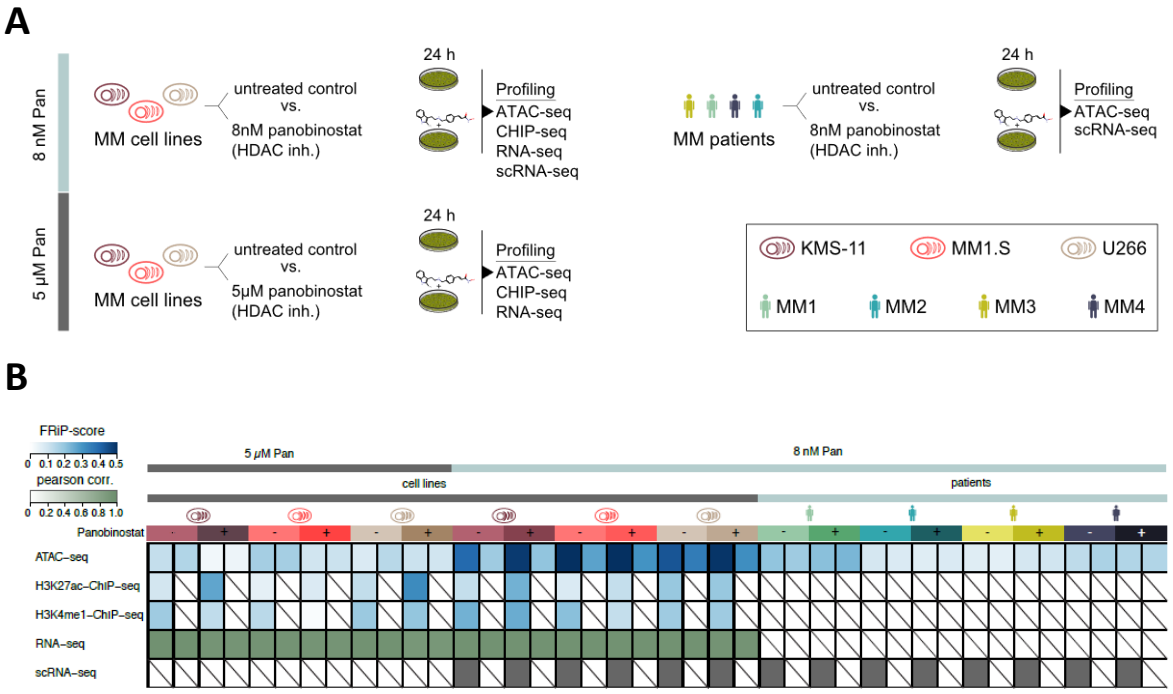
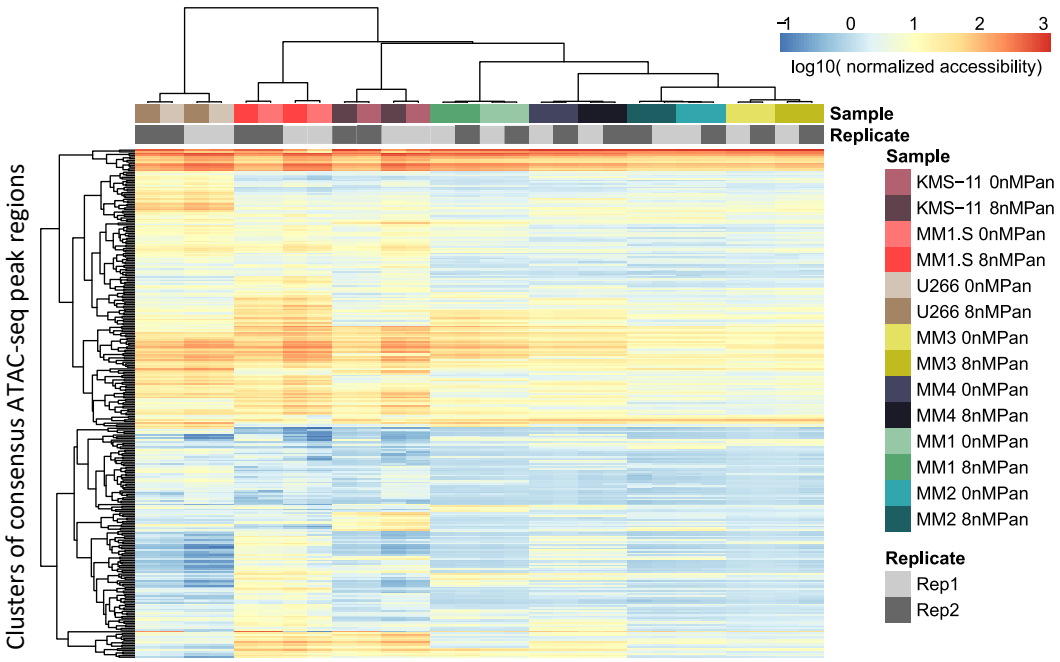


Figure 35 Overview of conducted read-outs to analyze the effect of 8 nM and 5 µM panobinostat on enhancer chromatin states and gene expression in multiple myeloma cell line and patient cells.
A Scheme of *ex vivo* panobinostat treatment of three MM cell lines and four patients at two different concentrations for 24 h. Figure adapted from Mallm et al. (Mallm *et al.*, 2019) **B** Quality metrics of the indicated chromatin or transcriptome read-out. Higher values of fraction-of-reads-in-peak (FRiP) scores indicate better quality of ATAC-seq and ChIP-seq technical replicates (blue). The Pearson correlation of technical replicates was calculated to access the quality of RNA-seq data (green). Grey: samples for which single cell RNA-seq was conducted. Pan: panobinostat. ChIP-seq and RNA-seq experiments with 5 µM panobinostat treatment and scRNA-seq of MM cell lines was conducted by Lisa Dreßler (Dreßler, 2017).

The effect on the activity of regulatory elements by panobinostat treatment was first addressed by the identification of changes in chromatin accessibility. ATAC-seq libraries were generated in replicates from independent experiments for MM cell lines (MM1.S, KMS-11, U266) treated *ex vivo* with 8 nM panobinostat for 24 hours. Additionally, ATAC-seq was conducted in technical replicates from cells of four MM patients, which were subjected to an *ex vivo* treatment with 8 nM panobinostat for 24 hours. Clustering of chromatin accessibility patterns of all MM cell lines and patients revealed accessibility signatures specific for individual cell lines or patients (Figure 36 A). This is in line with the notion that MM is a heterogeneous disease with high variation between MM patients (Kumar *et al.*, 2017). Differences observed between single cell lines and patients were more pronounced than changes induced by 8 nM treatment with panobinostat for 24 hours. Changes in chromatin accessibility induced by treatment were further assessed by a DiffBind analysis (Ross-Innes *et al.*, 2012). Regions with changed accessibility upon panobinostat treatment were identified separately for each MM cell line or patient (Figure 36 B). Regions with a significant (adjusted p-value < 0.05) change in chromatin accessibility were very rare or absent, despite of the globally increased histone acetylation levels. Common regions with changed accessibility induced by panobinostat could not be detected. In summary, there were no major changes detectable in the accessibility pattern of the three cell lines and four analyzed patients induced by the 24 hours treatment with 8 nM panobinostat, which corresponds to the range of IC50-values of panobinostat for most HDACs (Atadja, 2009). Therefore, the effect of a higher dose of panobinostat on the chromatin accessibility was assessed in the cell line MM1.S.

A



B

	Number of regions with gained accessibility (FDR < 0.05)	Number of regions with lost accessibility (FDR < 0.05)
KMS-11	0	0
MM1.S	10	0
U266	53	0
MM1	138	883
MM2	8	35
MM3	0	28
MM4	4	1

Figure 36 Chromatin accessibility changes at genomic loci genome-wide between MM cell line and patient cells treated *ex vivo* with 8 nM panobinostat or DMSO as control.
A The MM cell lines KMS-11, MM1.S and U266 as well as cells from four MM patients were treated for 24 h with 8 nM panobinostat or DMSO (control) before ATAC-seq was performed in replicates. Heatmap of chromatin accessibility signal (normalized using DiffBind) in 195,368 genomic regions aggregated to 300 clusters by k-means clustering (rows) is shown for all samples (columns). Samples were hierarchical clustered based on Pearson correlation using complete linkage. **B** Number of regions which show a significant gain or loss in accessibility in panobinostat treated compared to untreated cells of individual cell lines or patients identified using DiffBind (Ross-Innes *et al.*, 2012).

4.1.3 Local chromatin accessibility changes are induced by treatment with high panobinostat concentration in the MM cell line MM1.S

The cell line MM1.S was treated for 24 hours with 2.5 and 5 µM panobinostat, before chromatin accessibility changes compared to untreated cells were assessed by ATAC-seq. Around 20,000 regions were identified that exhibited a differential accessibility between treated and untreated MM1.S cells. Only ~6 % of those gained ATAC-seq signal. Among the regions with local loss of accessibility were several regions located within the super-enhancer region of the *IgH* locus, which is causing increased *MYC* expression in MM1.S cells due to a translocation event (Figure 37 A) (Dib *et al.*, 2008). This change in chromatin accessibility was not observed at lower panobinostat concentrations. The panobinostat-induced loss of chromatin accessibility in the *IgH* super-enhancer was accompanied by a local loss of the active enhancer marks H3K27ac and H3K4me1 as well as the expression of MYC (Figure 37 A). The downregulation of MYC expression in MM1.S cells in response to panobinostat treatment is in concordance with previous finding (Hideshima *et al.*, 2015). Gene regulation changes affecting MYC expression are especially interesting as its expression deregulation is frequently observed in MM patients (Chng *et al.*, 2011). Furthermore, regions with *de novo* accessible chromatin due to panobinostat treatment were found in proximity to the *COPS7A* locus, which encodes a subunit of the COP9 signalosome (Figure 37 B) (Li *et al.*, 2017). Next, I conducted a motif enrichment analysis in regions with accessibility alterations induced by high panobinostat concentrations. Non-promoter regions (outside RefSeq TSS +/- 1 kb) with lost accessibility exhibited an enrichment of transcription factor binding motifs of TCF3, IRF4, RUNX and MEF2, while non-promoter regions with gained accessibility were enriched for motifs of JDP2, cJUN and KLF. JDP2 has been previously described to inhibit histone acetylation (Jin *et al.*, 2006). Inhibition of histone acetylation would be a possible mechanism of cells to counteract HDAC inhibition.

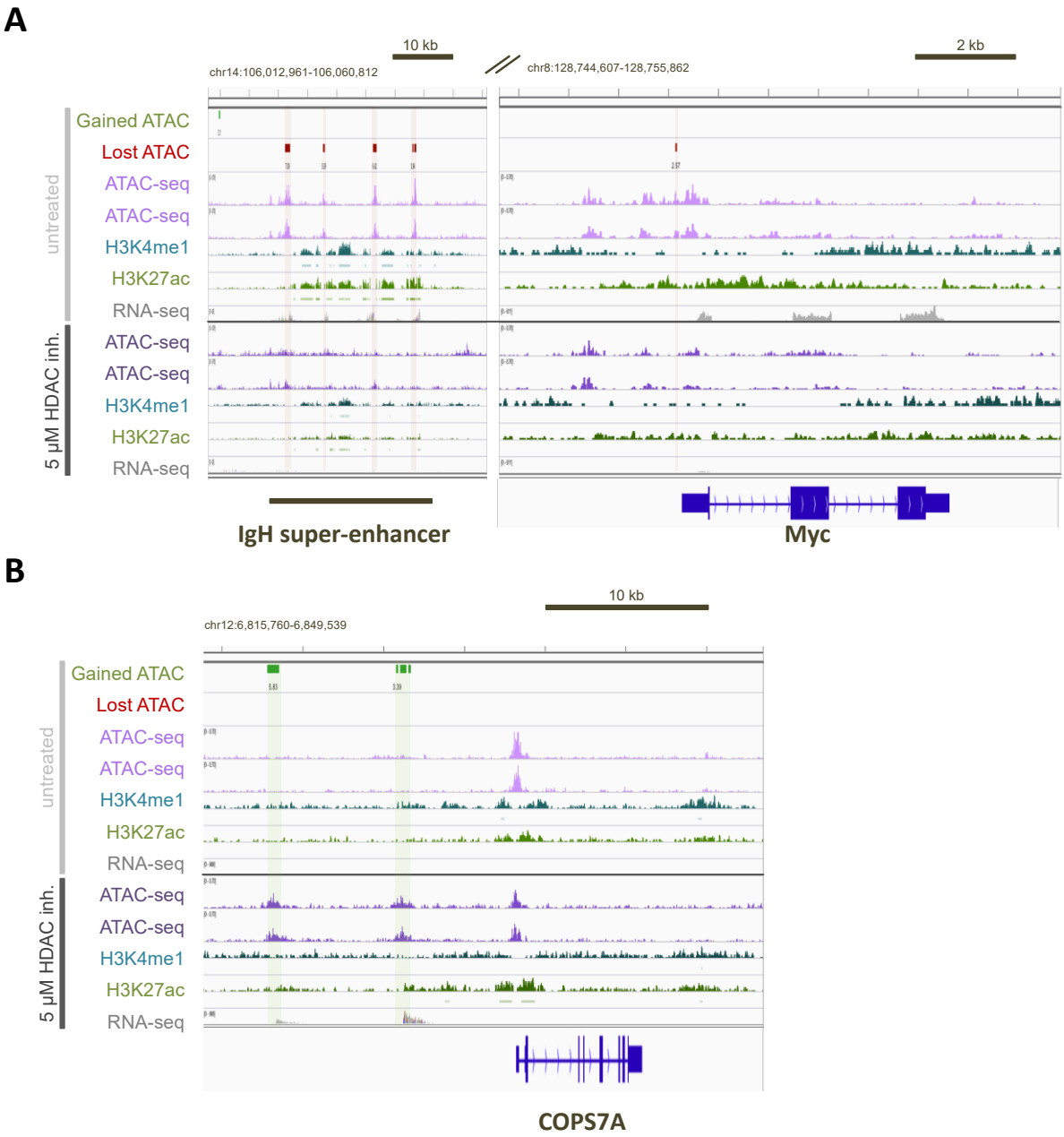
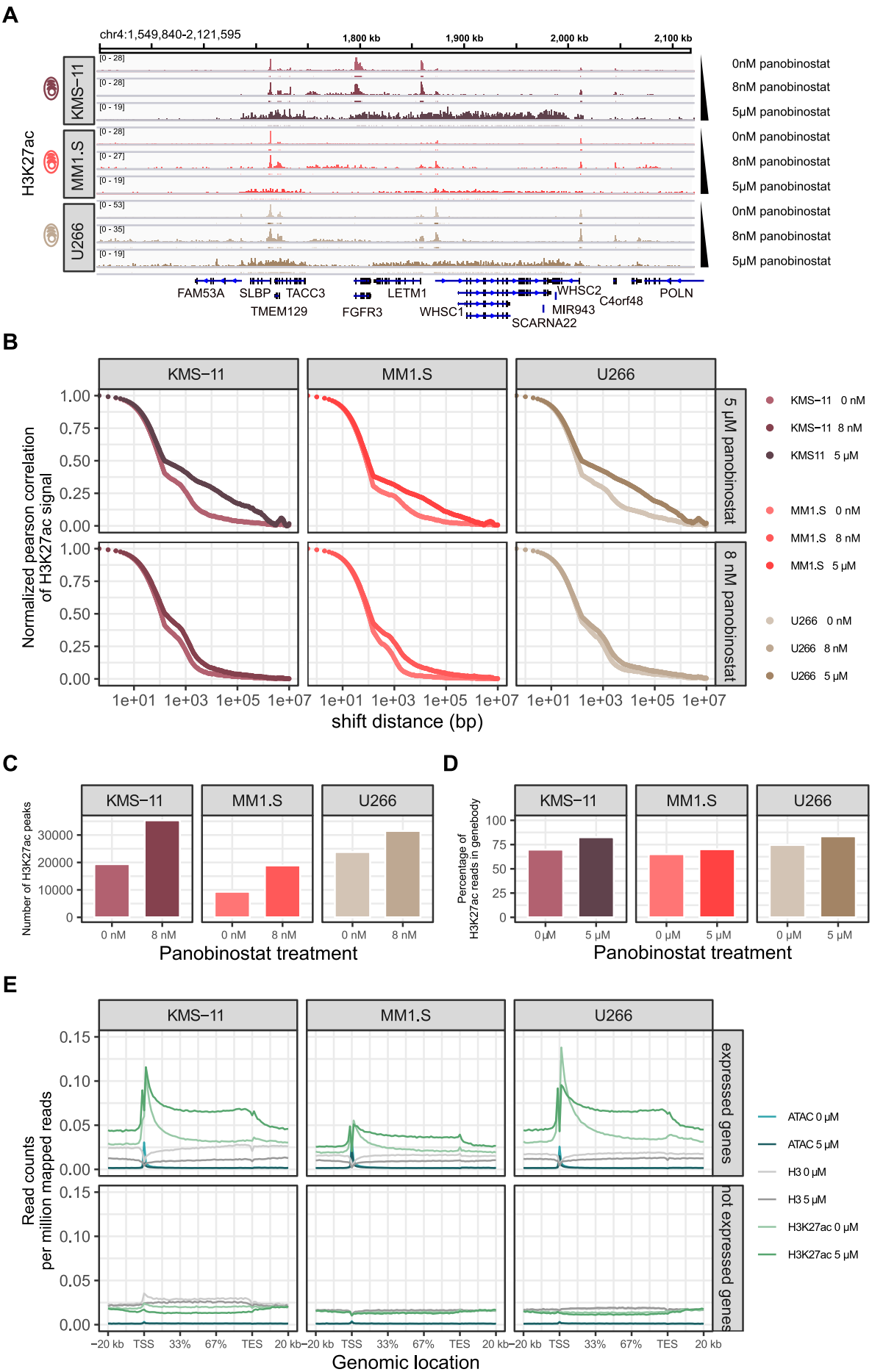


Figure 37 Genomic regions exhibiting local chromatin changes regarding accessibility and post-translational histone modifications in panobinostat (5 µM) treated MM1.S cells compared to untreated controls.
Exemplary genomic regions with **A** lost or **B** gained accessibility in MM1.S cells treated *ex vivo* for 24 hours with 5 µM panobinostat (HDAC inhibitor) in comparison to cells treated with DMSO (untreated). ”Gained ATAC” and “lost ATAC” tracks denote regions which exhibited a significant increase or decrease in accessibility measured by ATAC-seq. Regions were determined in a DiffBind analysis (Ross-Innes *et al.*, 2012). ChIP-seq and RNA-seq experiments were conducted by Lisa Dreßler (Dreßler, 2017).

4.1.4 Panobinostat induces broadening of H3K27ac over gene bodies of active genes

Although global H3K27ac levels were increased in all MM cell lines, no major changes in the chromatin accessibility could be detected upon treatment with 8 nM panobinostat for 24 hours. Only treatment with higher panobinostat concentrations induced detectable changes in chromatin accessibility. Therefore, I analyzed the genome-wide location of histone H3 acetylation at K27 by chromatin immunoprecipitation sequencing (ChIP-seq). ChIP-seq data revealed a broadening of previously existing H3K27ac induced by the treatment with panobinostat in all three MM cell lines (Figure 38 A). The typical narrow shape of H3K27ac peaks in the genomic read distribution even nearly disappeared after treatment with 5 μM panobinostat. This observation was quantified by computing genome-wide correlation functions. The Pearson correlation between normalized H3K27ac signal profiles shifted to each other with respect to their genomic location was calculated as a function of their shift distance with the tool MCORE (Molitor *et al.*, 2017). Correlation functions calculated from two genomic read profiles of the same condition allow to assess the topology of a certain domain. They were therefore used to compare the size of H3K27ac domains between panobinostat treated and untreated cell lines. Sharp decays in the function (inflexion points) arise at shift distances equal to half domain sizes. The initial decay in the correlation at small shift distances was still rather similar between treated and untreated cells for each cell lines and panobinostat concentration (Figure 38 B). This indicates that the fragmentation of the genome was similar between treated and untreated cells. The shift distance of the inflexion point representing the half fragment length equals the expected length of the DNA wrapped around a nucleosome. The shift distance at the inflexion point of the second steep decay. The shift to the right which is visible already after treatment with 8 nM panobinostat confirms the observed broadening of H3K27ac on a genome-wide scale. This phenomenon further increases with higher panobinostat concentrations (5 μM). This is probably attributable to the formation of broad H3K27ac domains spanning multiple peak regions present in untreated control cells (Figure 38 A). Simultaneously, the number of regions with significantly enriched acetylation (called peaks) increased with 8 nM panobinostat treatment after 24 hours (Figure 38 C). Differences in the quality of ChIP-seq data as technical cause for this observation could be excluded (judged from NSC and RSC values; Appendix Suppl. table S2). The broadening of H3K27ac further seemed to coincide with gene body annotations (Figure 38 A). Quantifying the fraction of H3K27ac reads in gene bodies showed an increase with panobinostat treatment (Figure 38 D). To test if the expression of a gene influences its gene body acetylation at H3K27, genes were separated into expressed and not expressed in each of the three cell lines from gene expression data. Then, the distribution of H3K27ac over gene bodies of expressed and silent genes was analyzed for cells treated with 5 μM panobinostat in comparison to untreated control cells (Figure 38 E). The typical sharp H3K27ac peak at the TSS of expressed genes is visible in untreated cells.


A
chr4:1,549,840-2,121,595
1,800 kb
1,900 kb
2,000 kb
2,100 kb
H3K27ac
KMS−11
MM1.S
U266
0nM panobinostat
8nM panobinostat
5μM panobinostat
FAM53A
SLBP
TACC3
TMEM129
LETM1
FGFR3
WHSC1
WHSC2
MIR943
SCARNA22
C4orf48
POLN
B
KMS−11
MM1.S
U266
5 μM panobinostat
8 nM panobinostat
Normalized pearson correlation of H3K27ac signal
shift distance (bp)
KMS−11 0 nM
KMS−11 8 nM
KMS11 5 μM
MM1.S 0 nM
MM1.S 8 nM
MM1.S 5 μM
U266 0 nM
U266 8 nM
U266 5 μM
C
Number of H3K27ac peaks
Panobinostat treatment
D
Percentage of H3K27ac reads in genebody
Panobinostat treatment
E
Read counts per million mapped reads
expressed genes
not expressed genes
Genomic location
ATAC 0 μM
ATAC 5 μM
H3 0 μM
H3 5 μM
H3K27ac 0 μM
H3K27ac 5 μM

◀ **Figure 38 Broadening of H3K27ac over gene bodies of expressed genes induced by the HDAC inhibitor panobinostat in MM cell lines.**
A Distribution of H3K27 acetylation at a locus on chromosome 4 containing *WHSC1* of MM cell line cells treated with 8 nM or 5 µM panobinostat for 24 hours and untreated control cells. **B** Correlation functions calculated with MCORE display the normalized Pearson correlation of H3K27ac signal in dependence of the shift distance of the genomic position. Correlation functions were calculated for the indicated MM cell lines treated with 8 nM (top) or 5 µM (bottom) for 24 hours and the untreated control cells from the same experiment. **C** Barplots of detected H3K27ac peaks for panobinostat treated (8 nM) and untreated cell. **D** Barplots of percentage of H3K27ac reads located within gene bodies of panobinostat treated (5 µM, 24 hours) and untreated cells. **E** Read coverage of H3K27ac, ATAC-seq and control H3 at expressed (top) or not expressed (bottom) genes in the respective cell line. Cells of MM cell lines were treated with 5 µM or left untreated for 24 hours. Expressed genes were identified by fitting two Gaussian distributions to the distribution logarithmized TPM values (base2) and the threshold determined as x-value of the intersection point for every condition separately.

The H3K27ac signal smears out into the whole gene body in all MM cell lines, while the genomic region upstream of the TSS stays at a lower acetylation level (Figure 38 E top). The decay of the H3K27ac signal downstream of the TSS in panobinostat treated cells in comparison to untreated cells showed a reduction in both the steepness and the amplitude. The slightly elevated basal levels in treated cells outside gene bodies, probably originates from gene body acetylation of nearby neighboring active genes. A spiking ATAC-seq signal at the TSS confirmed the selection of active genes. The acetylation difference between panobinostat-treated and untreated cells was not present for inactive genes in which acetylation was generally absent (Figure 38 E bottom). In summary, the HDAC-inhibitor panobinostat induces acetylation at histone H3 lysine residue K27 in gene bodies of transcribed genes as a major contribution to the observed global increase in acetylation detected by immunoblotting.

4.2 Response difference to panobinostat treatment between multiple myeloma cells of different genetic background

4.2.1 KMS-11 shows no major transcriptional response to panobinostat treatment

The observed global increase of H3K27ac levels in the MM cell lines KMS-11, MM1.S and U266 could also induce gene expression changes. A differential gene expression analysis was conducted with DESeq2 (Love *et al.*, 2014) for a systematic analysis of global gene expression alteration caused by the treatment of the MM cell lines with 8 nM panobinostat (Figure 39). Each cell line showed both up and down regulated genes and no clear tendency towards gene activation was observed. However, the three cell lines exhibited a remarkable difference in both the degree and the quantity of response. About 4,000-5,000 genes were each significantly up and downregulated in MM1.S and U266, while only about 300 differentially expressed genes were detected in total in KMS-11 (Figure 39).

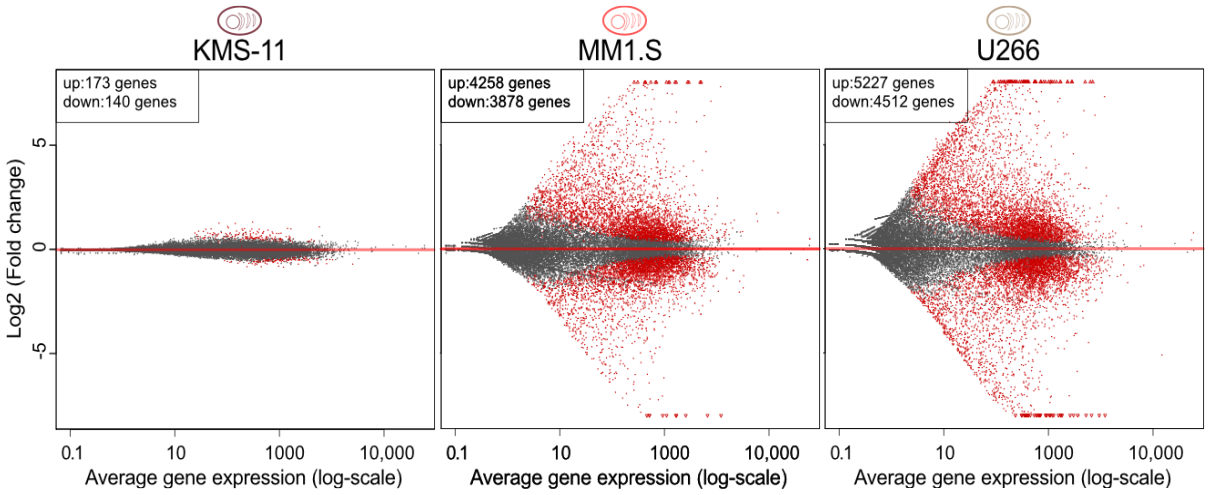


Figure 39 Differential gene expression analysis results for MM cell lines treated with 8 nM panobinostat compared to untreated control cells.
MA plots generated with DESeq2 (Love *et al.*, 2014) show log2 fold changes (y-axis) of the expression of individual genes between cells treated with 8 nM panobinostat (24 hours) and untreated cells plotted against the mean expression of the gene over all conditions (x-axis). Red dots denote genes exhibiting a significant (adjusted p-value<0.05) expression change between treated and untreated cells.

To further characterize this difference, I identified genes which showed a significantly different response to panobinostat treatment in the cell lines MM1.S and U266 compared to KMS-11 followed by a KEGG pathway enrichment analysis (Kanehisa & Goto, 2000). No KEGG pathway enrichment (p-value < 0.05) was found for genes identified to be more upregulated in the transcriptionally responsive cell lines compared to KMS-11. However, genes, which were significantly stronger downregulated in the two transcriptionally responsive MM1.S and U266 cell lines, were enriched in pathways such as “Cell cycle”, “DNA replication”, “p53 signaling” as well as “cellular senescence”. The differentially expressed genes belonging to the cellular senescence pathway were *E2F2, MYBL2, FOXM1, CCNA2, E2F1, CCNB1, CCNB2, CDK1*. The decrease in FOXM1 has for instance been associated with the induction of replicative-induced senescence (Smirnov *et al.*, 2016).

A visual representation with a KEGG pathway map was employed to evaluate the expression changes of genes induced by panobinostat in the whole cellular senescence pathway (Figure 40). It revealed a concerted change in gene expression in many genes linked to cellular senescence in the transcriptionally responsive cell lines (MM1.S and U266), while KMS-11 showed nearly no response especially of genes in the downstream part (Figure 40 right) of the pathway including *MYBL2* (*B-MYB*). This expression dynamic of *MYBL2* on the RNA level (Suppl. figure S 2) was not detectable on the protein level as tested by immunoblotting (Suppl. figure S 3). In summary, the response between the three cell lines on the transcriptional level showed substantial differences. These differences were associated with the regulation of cell cycle and senescence induction.

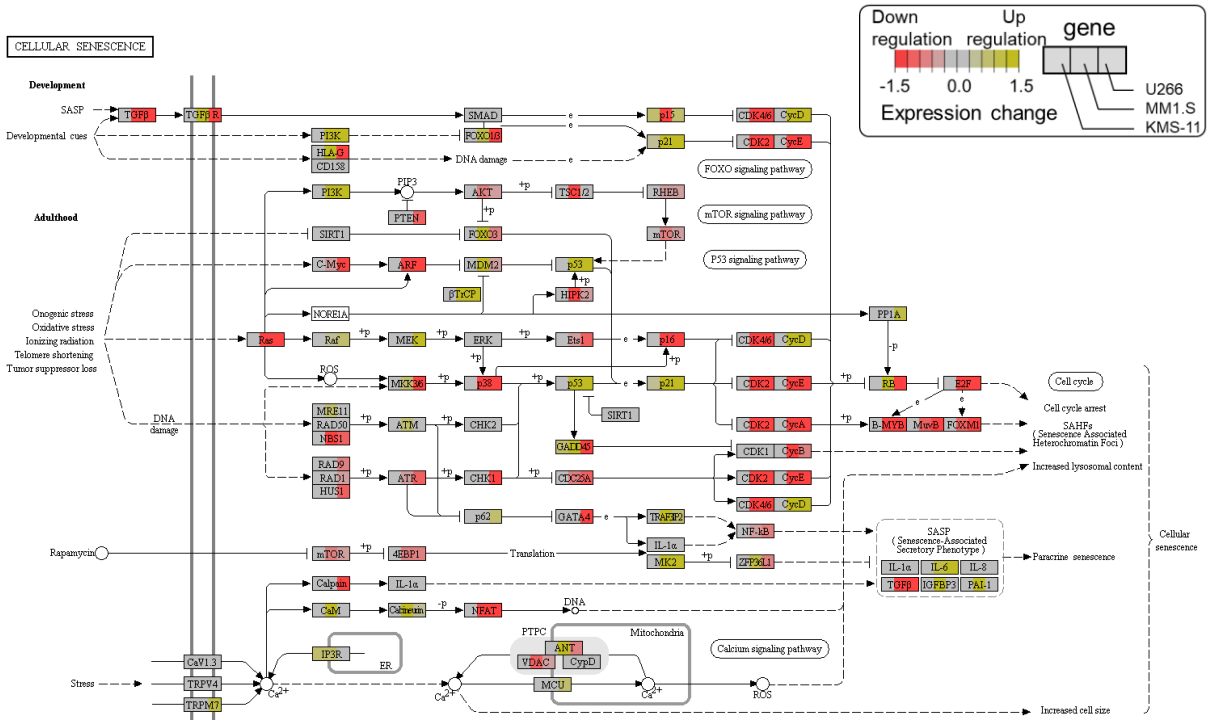


Figure 40 Cellular senescence pathway by KEGG with genes colored according to panobinostat-induced expression change in MM cell lines.
KEGG pathway map of cellular senescence (Kanehisa & Goto, 2000) colored by the expression change (log2) of genes or gene groups in cells treated with 8 nM panobinostat compared to untreated cells of MM cell lines KMS-11 (left), MM1.S (middle) and U266 (right). Visualization was conducted with pathview (Luo & Brouwer, 2013).

4.2.2 G1 arrest and senescence are induced by panobinostat in transcriptionally responsive cell lines

The differential gene expression analysis indicated that differences between the three MM cell lines in their response to panobinostat could be reflected in cell cycle regulation. To further evaluate changes in the cell cycle stage distribution induced by panobinostat, I analyzed scRNA-seq data of the three MM cell lines treated with 8 nM panobinostat or DMSO for 24 hours (Dreßler, 2017). The information about the gene expression patterns in individual cells allows to infer the cell cycle stage on the single cell level and thus the percentage of cells in a specific stage. Cell cycle stage inference conducted with the R package Seurat (Stuart *et al.*, 2019) revealed an unchanged cell cycle distribution in panobinostat treated cells compared to untreated ones for the transcriptionally unresponsive cell line KMS-11, while MM1.S and U266 show an increase of cells in G0/G1 phase (Figure 41 A). The increase of cells in G1 phase with panobinostat treatment only in the transcriptionally responsive cell lines could be confirmed by flow cytometry (Figure 41 B). The enlarged percentage of G1 was stable or even further increased with prolonged treatment till 72 hours.

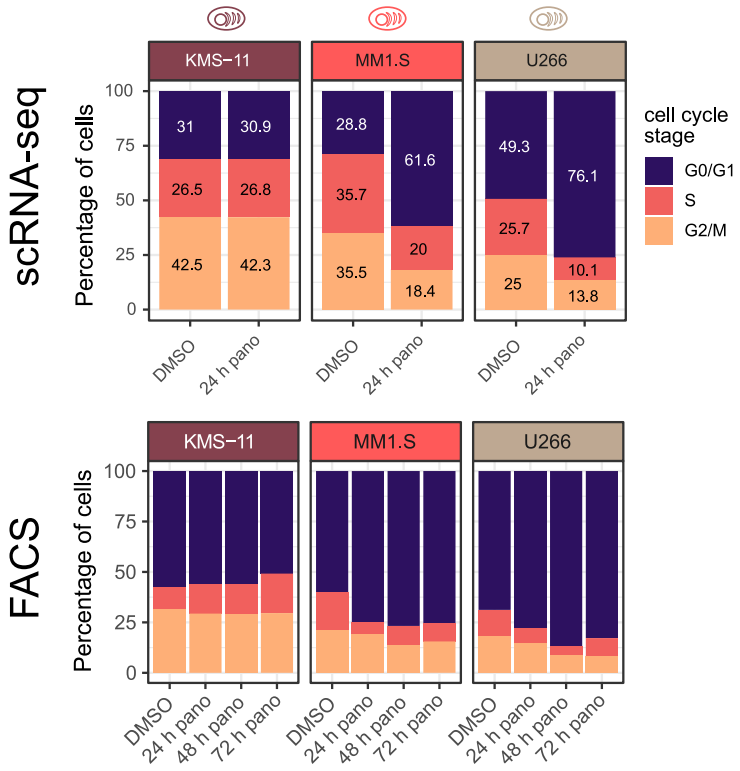


Figure 41 Cell cycle stage distribution in MM cell lines after treatment with panobinostat determined from a scRNA-seq or FACS analysis.
Percentage of cells in G0/G1, S or G2/M cell cycle phase were determined by **A** scRNA-seq using Seurat or **B** by flow cytometry (FACS). Pano: 8 nM panobinostat, DMSO: 24 hours (scRNA-seq) or 72 hours (FACS) treatment with DMSO as control.

We further evaluated the hypothesis that senescence is induced by panobinostat at different percentages between the cell lines. The scRNA-seq data was employed to determine the fraction of cells which show signs of senescence induction after 24 hours panobinostat treatment. Early marker genes of senescence (Zirkel *et al.*, 2018) were used to calculate for every cell two expression scores one each for up- and downregulated genes (see Appendix Suppl. table S3). Cells separated based on the two calculated score into two clusters (Figure 42 A). A k-means clustering was performed to separate cells which showed signs of senescence induction in their expression profile from cells that did not. Cells of the cluster with a high score for the expression of genes upregulated and a low score for genes downregulated during early senescence (Figure 42 A upper left corner) were classified as being in an early state of senescence. Quantification of this cell fraction revealed an increase of the senescent cells with panobinostat treatment only in MM1.S and U266 cells (Figure 42 B).

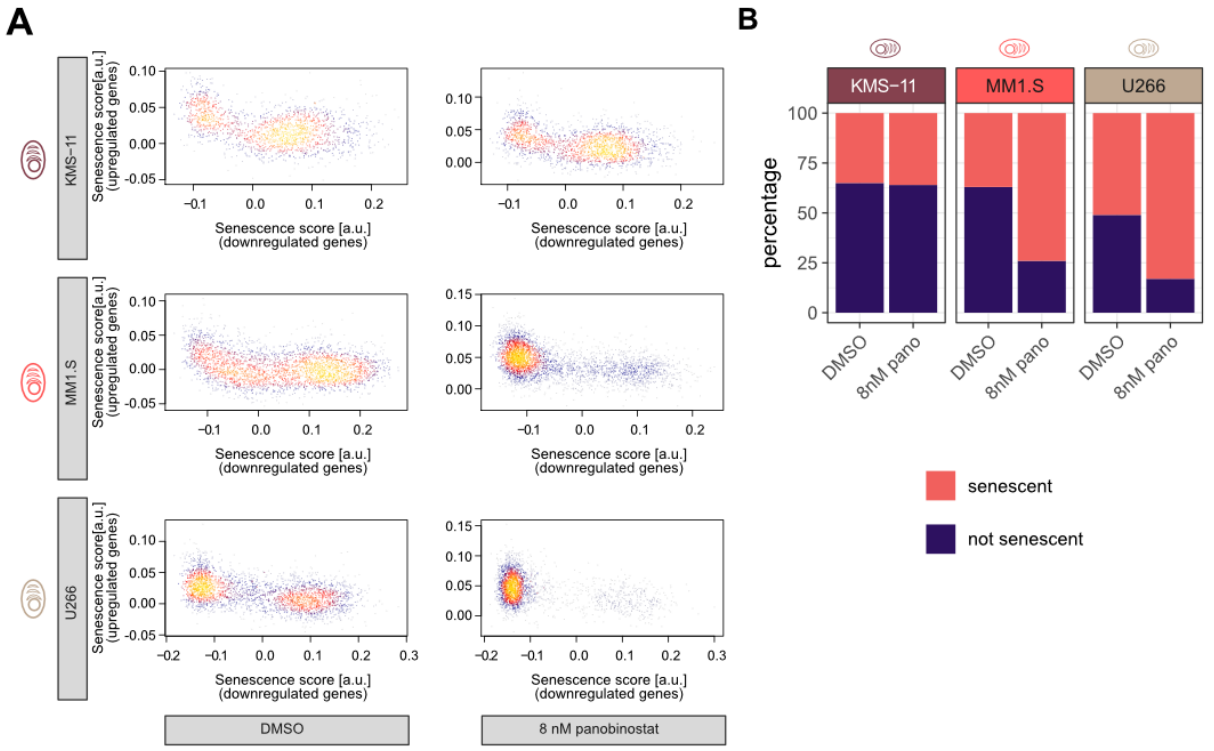


Figure 42 Senescence scores calculated for individual cells from scRNA-seq data of panobinostat treated MM cell lines.
A Expression scores calculated separately for genes up- and downregulated during early senescence are plotted against each other for individual MM cells treated with panobinostat or DMSO as control for 24 hours. **B** Barplot of fraction of cells belonging to the cluster which was determined to exhibit an early senescence expression phenotype.

An increased ß-galactosidase activity measurable ~7 days after senescence induction is a characteristic of senescence (Dimri *et al.*, 1995; Ewald *et al.*, 2010; Kuilman *et al.*, 2010). Therefore, senescence induction was assessed by the quantification of ß-galactosidase activity by flow cytometry after treatment of MM cell lines with 5 nM or 8 nM panobinostat for 7 days (Figure 43 A, B). The experiment was conducted in three independent replicates. MM1.S and U266 both showed an increase of cells with high ß-galactosidase activity with increased concentration of panobinostat. The shape of the distribution of ß-galactosidase activity of KMS-11 cells varied between experiments. By defining senescent cells as having a ß-galactosidase activity above the activity measured in untreated cells, the percentage of senescent cells in KMS-11 did not increase with panobinostat treatment. The measurement of cell cycle distribution by DNA content of the same cells revealed an unchanged cell cycle distribution for KMS-11 cells. Transcriptionally responsive cell lines in contrast showed a loss of cells with high DNA content and thus reduced percentages of cells in S or G2/M phase. Further experiments are needed to verify that these cells are in an irreversible cell cycle arrest, which is a feature of senescent cells. In summary, transcriptionally responsive cell lines (MM1.S and U266) showed an increase of cells in G1 after panobinostat treatment for 24 hours to 7 days, which was not detected in KMS-11. Furthermore, only the transcriptionally responsive cell lines MM1.S and U266 showed an increase of cells with

ß-galactosidase activity above the levels seen in untreated cells after 7 days panobinostat treatment, which is indicative of senescent cells.

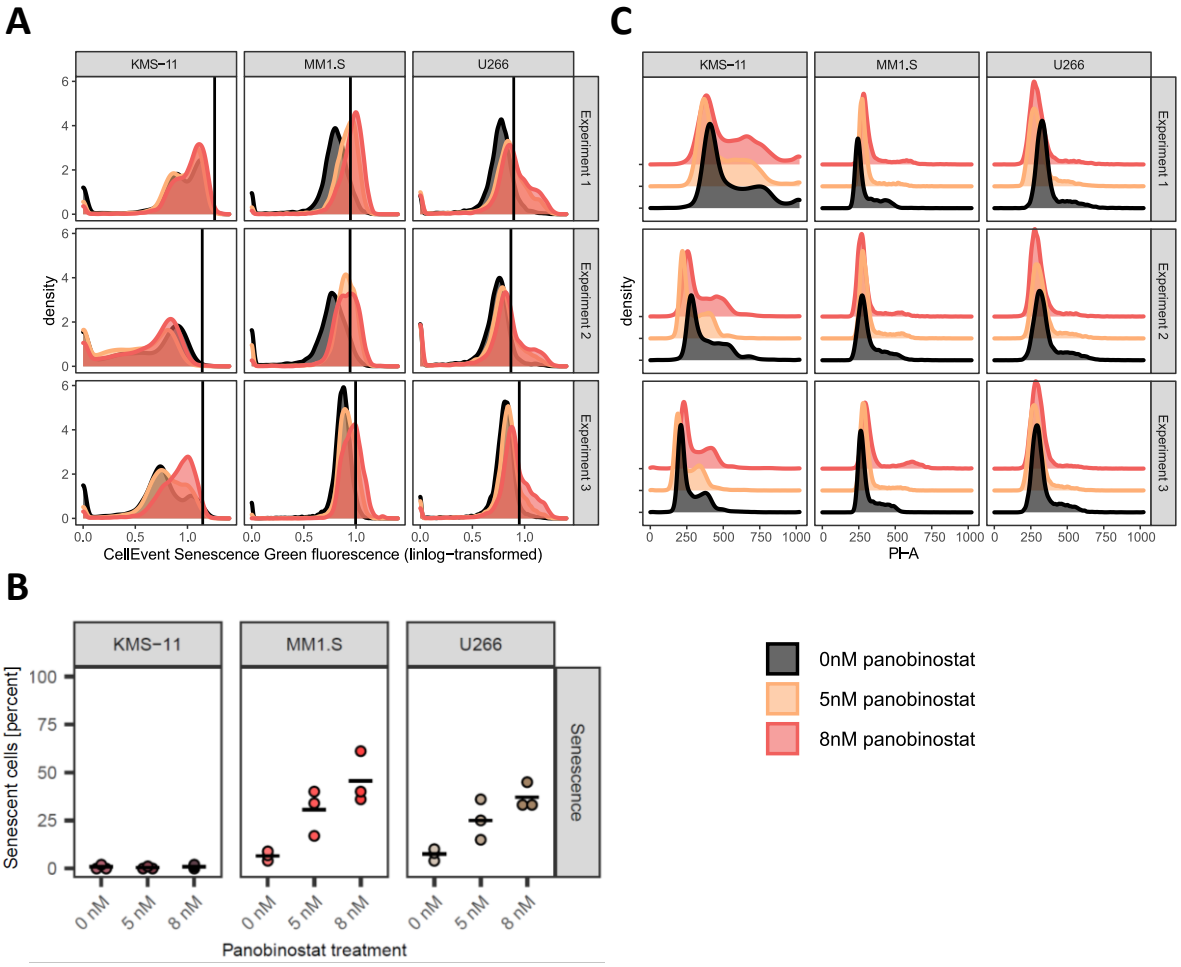


Figure 43 ß-Galactosidase activity and DNA content measurement by flow cytometry in panobinostat-treated MM cell lines.
A Distribution of ß-galactosidase activity levels (linear-log10 transformed) measured in three independent experiments after cultivating cells of the indicated MM cell line in the presence of panobinostat at the indicated concentration for 7 days. The indicated threshold (vertical black line) for the definition of senescent cells was individually calculated for every experiment and cell line from the distribution of untreated cells by the identification of the upper peak limit. **B** Quantification dot plot of percentage of cells with ß-galactosidase activity levels above the calculated threshold. **C** Quantification of DNA content using propidium iodide (PI-A) for cell cycle determination by flow cytometry.

4.2.3 Transcriptional non-responsiveness to panobinostat treatment is also observed in multiple myeloma patients

The three analyzed MM cell lines exhibited a strong difference in the degree of transcriptional response to panobinostat treatment. We then analyzed if such differences also existed between cells from MM patients. We studied the transcriptional change in cells of four MM patients treated *ex vivo* with 8 nM panobinostat compared to untreated control cells by scRNA-seq. In total 56,542 cells remained after quality filtering and were analyzed by the Seurat software (Stuart *et al.*, 2019). Cells of individual patients clustered separately (Figure 44 A, B). Clustering further separated treated and untreated cells of each patient with the exception of patient MM 2. Therefore, the number of differentially expressed genes between panobinostat treated and untreated cells was calculated for every patient individually (Figure 44 C). MM 2, one of the four analyzed MM patients, exhibited a minor response to panobinostat in terms of number of differentially expressed genes similarly as observed for the cell line KMS-11.

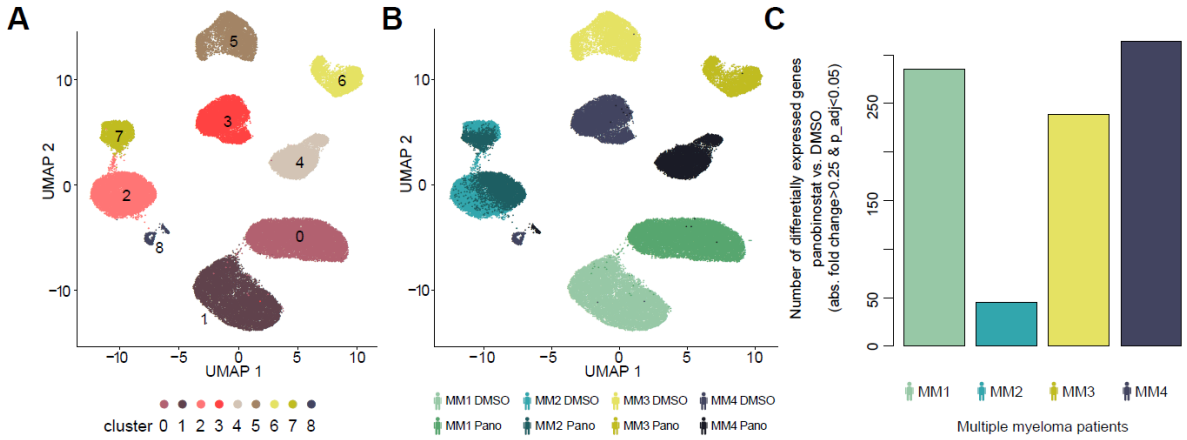


Figure 44 Panobinostat-induced gene expression changes in multiple myeloma patients analyzed by scRNA-seq using a droplet-based approach.
UMAP representation of transcriptome signatures of individual cells of four MM patients after *ex vivo* treatment with 8 nM panobinostat or DMSO for 24 hours as control colored by **A** cluster assignment or **B** sample. **C** Number of differentially expressed gene between treated and untreated cells of each patient.

Marker genes for the transcriptionally non-responsive patient MM 2 included *RAC3, PAGE1, LY6E, NME4, PTMS, RAB13, NGFRAP1, S100A10, PAGE5, GAGE5, FADS1, ANXA1, IFI27, GAL, UBE2S, EIF4EBP1 and RPS2*. High expression of *RAC3* was also observed in the transcriptionally non-responsive cell line KMS-11, while it was basically absent in all transcriptionally responsive cell lines and patients irrespective of treatment (Figure 45). Next, genes were determined, which were upregulated upon panobinostat treatment in transcriptionally responsive but not in the transcriptionally unresponsive cell lines and patients (KMS-11 and MM 2). Commonly upregulated genes were *TXNIP*, *BTG1* and *CYSTM1*. *BTG1* expression determined by single cell sequencing showed low levels in transcriptionally non-responding cells, while its expression increased with panobinostat treatment in

cells of transcriptionally responding patients and cell lines (Figure 45). An immunoblotting analysis could not detect any change on the protein level of BTG1 (Suppl. Figure S 3). Commonly downregulated genes only found in responsive cells were *NDUFB11*, *PHPT1*, *FAM136A*, *NDUFS6*, *PLP2*, *UFC1*, *HMGN2*, *NDUFB2*, *SRM* and *C12orf57*. In summary, the phenomenon of transcriptional unresponsiveness to panobinostat (low number of genes which change their expression) is not only present in MM cell lines but was also found in cells from MM patients.

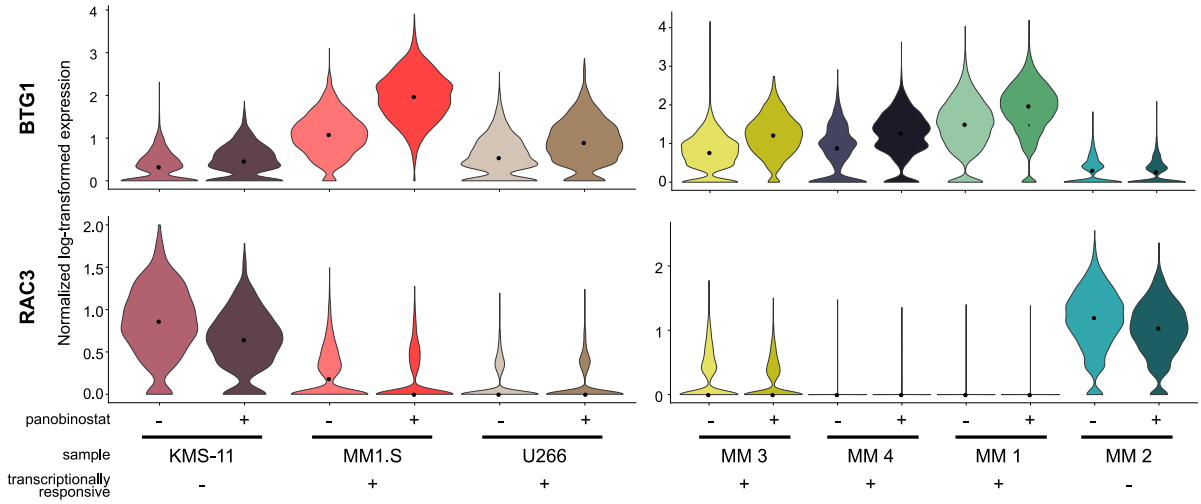


Figure 45 Violin plots of single cell gene expression of *BTG1* and *RAC3* in MM cell lines and patient cells treated *ex vivo* with panobinostat for 24 hours.
UMAP representation of transcriptome signatures of individual cells of four MM patients after *ex vivo* treatment with 8 nM panobinostat or DMSO for 24 hours as control colored by **A** cluster assignment or **B** sample. **C** Number of differentially expressed gene between treated and untreated cells of each patient.

Discussion

Every stage of B-cell differentiation is characterized by a certain epigenome profile to enable gene regulatory programs necessary for this developmental stage (Satpathy *et al.*, 2019; Smith & Meissner, 2013; Wu & Sun, 2006). Cancer is emerging as a disease of genetic and epigenetic events, that establish and maintain in a complex interplay deregulated gene expression circuits (Baylin & Jones, 2011). Depending on the differentiation stage at which it occurs, the same genetic alteration can have diverging consequences. Thus, the underlying epigenetic landscape of a cell might contribute to the phenotypic impact of a genetic alteration (Krivtsov *et al.*, 2013; Visvader, 2011). Along those lines, the most common translocation in multiple myeloma (MM) t(11;14) is also frequently observed in mantle cell lymphoma (Barwick *et al.*, 2019). Although the same genetic alteration contributes to the formation of phenotypically differing disease, it is unclear what molecular mechanisms are required to discriminate these different developments. One decisive factor could be that the cells-of-origins for both diseases are in different developmental stages and thus harbor differing epigenetic signatures (Barwick *et al.*, 2019). Thus, knowledge of the epigenetic state of the cell-of-origin is crucial to reliably identify and characterize epigenetic deregulation in cancer. The importance of the analysis of epigenetic alterations in disease is further emphasized by their potential reversibility.

In this thesis, I dissected epigenetic deregulation by the analysis of chromatin accessibility patterns and their functional implications during normal B-cell maturation, tumorigenesis and under the influence of drugs in chronic lymphocytic leukemia (CLL) and multiple myeloma (MM) (Figure 46). By establishing the required experimental and bioinformatic procedures for ATAC-seq, I could dissect developmental, tumorigenic and drug-induced activity changes of cis-regulatory elements linked to transcription factor binding, which guides gene regulatory programs. I showed here exemplary for B-cell subsets from the spectrum of potential tumor initiating cells of CLL that chromatin accessibility measured by ATAC-seq can be used to reconstruct cellular phylogenies of distinct disease-precursor subpopulations. The approach based on principal component regression to map the tumor-initiating cell for individual CLL patients enabled to determine their developmental stage and thus their chromatin accessibility pattern. The cells-of-origin of individual CLL patients could be assigned to the range of CD5 memory and IgM memory cells on the main maturation trajectory (Figure 46).

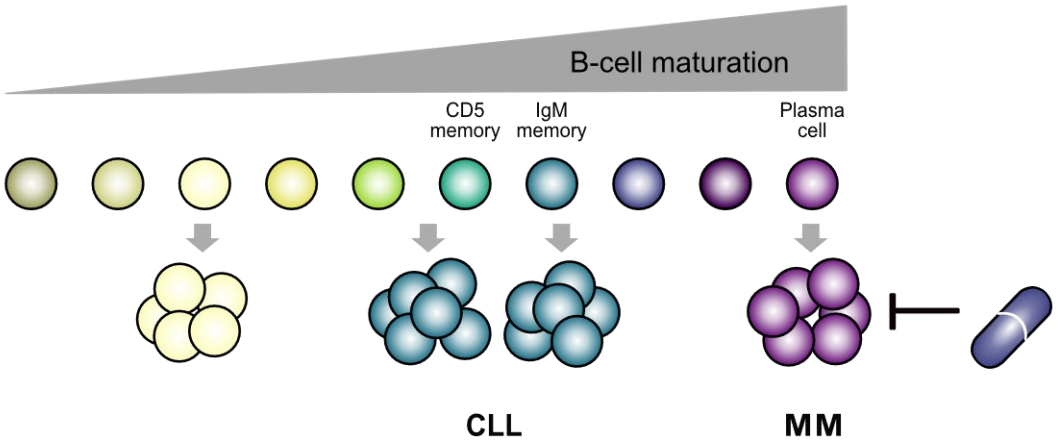


Figure 46 Analysis of chromatin accessibility changes during B-cell maturation, malignant transformation and drug treatment.
Chronic lymphocytic leukemia (CLL) and multiple myeloma (MM) originate from a continuum of B-cell maturation stages within the lymphoid lineage characterized by certain epigenetic profiles. The development of tumor entities from several adjacent B-cell developmental stages, which are summarized as a single clinical disorder, is illustrated for CLL. Based on the analysis of chromatin accessibility in this thesis, CLL was found to originate from the range of CD5 memory and IgM memory cells. Malignant transformation events are indicated by grey arrows. Drugs (right) can indirectly or directly affect the epigenetic state of malignant cells. Figure adapted from Kulis et al. (Kulis *et al.*, 2015).

With this approach disease-induced epigenetic patterns were successfully identified. The analysis of disease-induced chromatin accessibility changes by the conventional comparison of CLL to $CD19^+$ B cells of healthy donors (Landau *et al.*, 2014; Oakes *et al.*, 2014; Pei *et al.*, 2012), revealed a disturbed enhancer activity in CLL and linked these changes to the deregulation of transcription factors like EBF1. By comparing chromatin accessibility signatures of CLL patients to their corresponding cell-of-origin, CLL-specific alterations could be further delineated by dissecting them from preserved "imprints" of normal epigenetic development (Figure 46). The identification of these neoplastic chromatin accessibility alterations allowed to study their reversibility by the treatment with the B-cell receptor signaling inhibitor ibrutinib. Treatment with ibrutinib partly reverted disease-induced chromatin changes and induced alterations in the clonal intratumor heterogeneity with implications for drug response, which could be resolved at unprecedented depth with single cell techniques. I further demonstrated that the direct chromatin-modifying drug panobinostat induced broad global histone acetylation changes mainly in gene bodies of expressed regions in MM cell lines. Interestingly, these alterations did however not lead to a change in chromatin accessibility signatures. With the development of the tumor cell-of-origin mapping approach based on ATAC-seq, the identification of disease-specific chromatin accessibility changes in CLL, and the analysis of the effect of drugs on epigenetic signatures, this thesis lays the foundation for a better understanding of disturbed chromatin regulatory mechanisms in lymphoid malignancies and their reversibility by drugs.

1 Establishment of experimental and data analysis procedures to map accessible chromatin regions

ATAC-seq has already been used for the detection of activity changes of regulatory elements in health and disease. It has proven to be very valuable to study the regulatory landscape of normal hematopoiesis in distinct populations and continuously as a whole (Buenrostro *et al.*, 2018; Corces *et al.*, 2016; Satpathy *et al.*, 2019) or the parts relevant for a specific disease (Beekman *et al.*, 2018; George *et al.*, 2016). ATAC-seq is a relatively new chromatin readout (Buenrostro *et al.*, 2013) and requires appropriate experimental and data analysis procedures. Currently, only a limited number of software tools are available that are specifically designed for the analysis of ATAC-seq data (Yan *et al.*, 2020). Existing tools developed for other readouts like MACS2 for peak calling can be used for ATAC-seq analysis but need a careful selection and adjustment of parameters to yield correct results. Accordingly, I adjusted the settings for tools designed for other data types for the analysis of chromatin accessibility. Now, more and more guides for the analysis of ATAC-seq data are available and even an increasing number of publicly available pipelines exist. These pipelines are similar in basic concepts, differ however in detailed settings and selection of integrated tool, the range of analyses that are possible and computational expertise of the targeted user (Smith & Sheffield, 2020). The pipeline I set up for the analysis of ATAC-seq data is in line with the publicly available pipelines.

The establishment of an qPCR quality control step enables to assess the quality of ATAC-seq libraries before sequencing. I demonstrated that the qPCR results positively correlated with the quality of the sequenced library. Consequently, the qPCR control step enabled to get a qualitative measure of library quality before sequencing, which classic library quality assessment by fragment size distribution alone cannot accomplish. Classically, only the fragment size distribution is assessed for quality control by gel electrophoresis (Buenrostro *et al.*, 2015a). As illustrated in my thesis, reads originating from the mitochondrial genome can however largely obscure the observation of a nucleosomal pattern by gel electrophoresis. Nucleosomal patterns are considered to be a characteristic of a good quality ATAC-seq library (Buenrostro *et al.*, 2015a). Factors that could prevent a more accurate prediction include the typically broad library size of ATAC-seq. Fragments of all length are considered in qPCR, while the sequencing process eliminates long fragments. Vice versa small fragments which account for a huge proportion of ATAC-seq libraries can be too small to span the region (~100 bp) between the primer pairs and are thus not amplifiable by PCR. Further, only single regions are considered which are subjected to stochastic and possibly also slight biological variations. Finally, the FRiP-score is not an absolute measure of sequencing quality, but dependents on the number and size of peaks, which is

influenced by library quality and by the settings of the peak calling tool. Although it cannot exactly predict the library quality measured by FRiP-score, the qPCR step prevents either unnecessary sequencing times and costs or discard of precious primary samples. In ongoing projects with established protocols, extensive library quality assessment before sequencing might not be required also considering further published optimizations of the original ATAC-seq protocol with respect to mitochondrial contamination percentage (Corces *et al.*, 2016; Corces *et al.*, 2017; Montefiori *et al.*, 2017). However, a reliable pre-sequencing measurement to predicted the quality of the data set is important for new cell types and especially primary material, which often shows varying quality levels even within a project.

2 Modeling chromatin accessibility signatures of personal tumor cells-of-origin exemplary for CLL

The knowledge of the non-malignant reference epigenetic state at best the epigenetic state of the cell-of-origin is highly important to bring epigenetic deregulation in disease and drug response into context. Consequently, I developed an approach to model the chromatin accessibility patterns of patient-specific cells-of-origin for the analysis of epigenetic deregulation of regulatory elements in CLL. Chromatin accessibility is considered to be a rather stable cell type mark which progressively changes during development and has been used for cell and cancer type assignment (Corces *et al.*, 2016; Corces *et al.*, 2018; Erarslan-Uysal *et al.*, 2020) and even to disentangle differentiation trajectories and cellular hierarchies (Corces *et al.*, 2016; Satpathy *et al.*, 2019). I employed chromatin accessibility data measured by ATAC-seq and showed that it can also be used to reconstruct the cellular hierarchy of healthy potential disease precursors of CLL. I decided to “freeze” the phylogenetic tree and thus the calculated maturation stage for the six distinct non-malignant B-cell subsets for the calculation of the cell-of-origin for individual CLL patients. This differs from the previously described approach for DNA methylation data (Wierzbinska *et al.*, 2020). It has the advantage that the calculated maturation scores for the healthy subpopulations are not affected by the analyzed patient cohort and yield comparable results between different cohorts.

I further used an approach based on principal component regression to identify the cell-of-origin defined as the closest virtual cell on the main trajectory. Simple multiple linear regression was not possible, due to the naturally occurring multicollinearity in the data, which arose from the coregulation of genes. The determination of the maturation stage of the cell-of-origin of individual CLL patients from ATAC-seq is of relevance for clinical assessment of patients due to the prognostic value of the maturation stage. Reasons that might prevent a fast implementation in clinical routine include the demands on sample quality and conservation, which are higher than for instance for analysis of DNA methylation. The

modelling of the developmental stage of the tumor cell-of-origin from ATAC-seq data allows however the inference of its chromatin accessibility state. Thus, the described approach enables the computational comparison of the deregulated accessibility of patients' cell to the state in its cell-of-origin. These accessibility changes are of high functional relevance, as chromatin accessibility contains information about gene regulation on several layers. Thus, it complements the analysis of epigenetic deregulation in disease from DNA methylation data.

So far different approaches were used to analyze the diseased state. Some at least partially accounted for changes occurring during normal differentiation. Beekman and coworkers defined changes occurring in the epigenetic state of CLL as diverting from all the profiles of distinct healthy subpopulations from a broad range of B-cell development (Beekman *et al.*, 2018). The usage of all possible healthy references simultaneously is a rather conservative approach. Disease-induced changes within the physiological spectrum are per se not detectable. However, important components for tumor maintenance and progression are often factors involved in developmental processes, which confer self-renewal capacities or prevent differentiation (Dormoy *et al.*, 2012; Look, 1997; Nagel *et al.*, 2015). The identification of the cell-of-origin maturation stage and the modeling of its accessibility pattern like presented in this thesis also allowed to identify changes of developmentally dynamic regions as shown for the promoter region of *GAB1* (class A region). Corces et al. used 13 distinct healthy subsets distributed over complete hematopoiesis to compute the contribution of the accessibility profile of each cell type to an unknown population (Corces *et al.*, 2016). They employed this on AML cell types to calculate the fractional impacts. They further generated *in silico* normal cells composed of a mixture of the distinct healthy subset profiles according to the before calculated fractions for the analysis of deregulation in AML. This approach is designed to identify very distinct profiles of a wide range of cell types in leukemia cells. It is very well suited for the heterogeneous nature of the analyzed AML samples, which display profiles not present in normal development, but a mixture of regulatory profiles. However, it is not designed for more similar populations and does not account for continuous changes between the distinct measured healthy states.

The here described approach for CLL might serve as a model for the analysis of other tumor entities. The importance of the epigenetic state of the cell-of-origin for tumor manifestation also in other tumor entities is emphasized by an example of *MLL*-rearranged AML. Here, the same genetic aberration was experimentally induced in different cells-of-origin and led to phenotypically differing tumor subtypes (Krivtsov *et al.*, 2013). Krivtsov et al. employed a model for a cytogenetically homogeneous but clinically divers AML subgroup. They could show that, depending on the cell-of-origin (hematopoietic stem cells or granulocyte–macrophage progenitors), tumorigenesis induced by MLL-AF9 leads to differing leukemia cells with respect to transcriptome and DNA methylation profile. Leukemia cells derived from the more immature stem cells display signatures correlated with poorer prognosis and are less responsive to drug treatment. Georg et al (George *et al.*, 2016) is in line with these finding by an

extended analysis and found unique open chromatin signatures preserved in the leukemia cells, which allowed to identify the cell-of-origin from the four used healthy precursor populations.

Several issues need to be considered for applying the cell-of-origin calculation described in this thesis to other tumor entities. As described above, for very heterogeneous and lineage mixed disease other approaches might be more suitable. Additionally, cell types which keep a high proliferative potential during differentiation, like in the lymphoid lineage, might be more prone to give rise to neoplasms throughout differentiation including late stages of development. For other cancer types the potential cell-of-origin is thought to be only the tissue stem cell, which already possesses self-renewal capacities, or a restricted progenitor cell (Visvader, 2011). In cases where the oncogenic event happened at an early differentiation stage, the malignant cell might further differentiate along the developmental axis until arrested at possibly different developmental stages (Cobaleda & Sanchez-Garcia, 2009). This is likely to be relevant for neuroblastoma, where the tumor initiating cell is thought to be an immature precursor, which further differentiates to different extends (Jogi *et al.*, 2012). In such cases the closest cell on the trajectory might not equal the cell-of-origin. Likewise, the loss of a lineage-specifying transcription factor was reported to induce a dedifferentiation from the tumor initiating cell (Cobaleda *et al.*, 2007), which could also affect the prediction.

3 Deregulated chromatin accessibility and transcription factor binding at enhancers in CLL

The analysis of chromatin accessibility changes in CLL revealed a deregulation of enhancer activity, which probably drives pathogenic transcriptional programs. Epigenetic enhancer deregulation has also been described on the level of DNA methylation (Wierzbinska *et al.*, 2020). Locations with CLL-specific methylation changes were found to be enriched for annotated enhancer regions. Furthermore, described super-enhancer regions presented DNA methylation changes. CLL-specific hypomethylation events were found to be enriched for *de novo* gained super-enhancers in CLL (Mallm *et al.*, 2019; Wierzbinska *et al.*, 2020). The integration with several other chromatin readouts revealed that 81 % of gene expression changes could be explained by a change of chromatin features at either promoters or enhancer regions (Mallm *et al.*, 2019). A transcription factor network, derived from the integrated analysis, contained the SWI/SNF chromatin remodeling complexes as one connectivity hub. Those complexes might play a role in the translocation of nucleosomes at regions with changed accessibility. An additional single cell ATAC-seq analysis further dissected enhancer promoter connections and revealed a rewiring which might indicate that enhancer deregulation in CLL also occurs via connectivity changes (Mallm *et al.*, 2019). Enhancer deregulation in CLL has been confirmed by

others (Beekman *et al.*, 2018). Deregulation of enhancer activity measured by chromatin accessibility was also found in other tumor entities. Jin *et al.* found about 20,000 enhancers with change activity in multiple myeloma (MM) compared to memory B-cells or plasma cells (Jin *et al.*, 2018). Enhancers with changed activity in MM are linked to gene expression changes and contain many regions previously described to be important for the pathogenesis of MM. Along those lines, the accessibility at HOX binding regions in regulatory elements could be linked to a differentiation block in AML (Corces *et al.*, 2016). The loss of chromatin accessibility at bivalent regions alongside enhancer and promoter regions might be indicative of a loss of plasticity in CLL.

The most prominent change in inferred transcription factor activity exhibited EBF1, which showed a clear enrichment in regions with lost accessibility. EBF1 is an important transcription factor for early B-cell development and has activating and repressing functions (Boller *et al.*, 2018). EBF1 has been described to possess pioneer factor activity linked to its C-terminal domain. This means it is able to induce chromatin accessibility and DNA demethylation of chromatin in closed confirmation thus allowing other transcription factors to bind (Boller *et al.*, 2016). Besides, EBF1 is known to be involved in the activation of the promoter of the B-cell receptor component *CD79A* (Maier *et al.*, 2004), which is contained in class B regions in my analysis (lost accessibility in CLL). Furthermore, EBF1 has been recently described alongside PAX5 to prevent transformation in the context of B-cell acute lymphoblastic leukemia linked to the inhibition of IL-7 signaling (Ramamoorthy *et al.*, 2020). This emphasizes the important role of EBF1 in CLL biology. In summary I could identify disease-induced changes in chromatin accessibility and showed that they can be linked to a deregulation of transcription factor activity.

4 Effects of targeted drugs in blood cancer therapy on epigenetic deregulation

The enzymatic regulation of epigenetic processes bears the potential for interesting new intervention points for targeted cancer therapy. Drugs might indirectly or directly influence the deregulated epigenetic state of a tumor cell. I showed that the treatment of CLL patients with ibrutinib could partially reduce the disease-induced accessibility changes. The BTK inhibitor ibrutinib only indirectly affects the epigenetic state. Thus, most regions with restored physiological state are probably attributable to the normalization of B-cell receptor signaling. However, only a fraction of regions showed this significant reduction in the deviation from the physiological state. Consequently, the question arises, if some changes could be self-sustained, even though the trigger (e.g., high B-cell receptor signaling) is removed. Further analysis would be needed to address this question.

Although promising results have been achieved, some direct chromatin modifying compounds fell short of general expectations, which can be attributed to their global and pleiotropic effects that are not very well understood (West & Johnstone, 2014). A better understanding of molecular processes of drug action might help to overcome this. Overexpression of HDACs has been described for multiple myeloma (Mithraprabhu *et al.*, 2014), but also for CLL (Mallm *et al.*, 2019), which further emphasizes the value of studying HDAC inhibitor effects. I could show that the direct chromatin modifying drug panobinostat induced global histone acetylation changes in three multiple myeloma cell lines with different genetic background. The increase in acetylated histones was mostly attributable to a broadening of H3K27ac signal in gene bodies of expressed genes, independent of the cell line. This phenomenon has also been described for the HDAC inhibitor largazole, which was not correlated with expression changes of the respective gene (Sanchez *et al.*, 2018). The chromatin accessibility signatures of individual cell lines and patients differs. Treatment with 8 nM panobinostat did however not lead to detectable changes in the chromatin accessibility in all cell lines and patients. Qu and coworkers report an increase of chromatin accessibility in HDAC treated T-cell leukemia patients, which responded to the drug. However, the accessibility changes were not present in patients without clinical response (Qu *et al.*, 2017). The cell line U266 was indeed described as model for clinical resistance to panobinostat in patients mediated by high CXCR4 expression (Beider *et al.*, 2019). CXCR4 expression in U266 could be confirmed in our data, was however basically absent in all other cell lines and patients. This suggests a complex nature of response to panobinostat, which might be mediated by different mechanisms.

Similar to the interconnection of epigenetic and genetic layers, drug treatment influences the epigenetic landscape and clonal genetic composition. HDAC inhibitor treatment is generally associated with an increase of histone acetylation connected to open chromatin. Besides this direct effect on chromatin, HDAC inhibitors can also indirectly influence the chromatin state by the acetylation of non-histone proteins (Choudhary *et al.*, 2009). Thus, HDAC inhibitors can have diverse secondary effects on the epigenetic status of chromatin. The observed local loss of accessibility induced by 5 µM HDAC inhibitor treatment in MM cell lines could be induced by inhibitory effects on non-histones protein acetylation.

Inversely, the effectivity of drugs is also impacted by the genetic and epigenetic background. I could show that cell lines with different genetic and also epigenetic background exhibit different degrees of transcriptional response to panobinostat. Transcriptional responsiveness in two cell lines was accompanied by a G1-arrest and senescence induction. The phenomenon of transcriptional non-responsiveness to panobinostat was also observed in one MM patient, which emphasizes the importance of this observation identified in my exploratory analysis. Candidate genes for a further evaluation in subsequent studies were identified, which included genes with a known role in tumor biology. BTG1 was, for instance, reported to be a tumor suppressor and is frequently mutated in ALL

(Yuniati *et al.*, 2019). How transcriptional responsiveness relates to clinical response, still needs to be determined and requires a bigger patient cohort.

Reported effects of HDAC inhibitors include the induction of cell cycle arrest, apoptosis, differentiation and senescence as well as the inhibition of angiogenesis. The response to HDAC inhibition is highly affected by the genetic background and the selection of HDAC inhibitor (West & Johnstone, 2014). The expression of certain HDACs is, for example, already different between the three cell lines (Suppl. Figure S1). Aspects that could further lead to a diverse response to panobinostat include that HDAC function in multiprotein complexes, which shapes their activity (Cheng *et al.*, 2019). Moreover, mutations in HDACs are described, which leads to resistance to HDAC inhibitors (Ropero *et al.*, 2006). The analysis of HDAC response in different genetic backgrounds emphasizes the need for a better understanding of the interplay of epigenetic modulation and the genetic background. Although epigenetic changes are theoretically reversible, the complex interplay of many factors might prevent simple reversion of deregulated processes. A possibility to address this issue could be to identify specific deregulated events in individual patient groups and target them by more specific drugs (West & Johnstone, 2014).

Aside from intertumor heterogeneity, intratumor heterogeneity might be a further reason for differences in the response to drugs. The intratumor heterogeneity was minor on the transcriptional level. Resistance development resulted in a similar expression profile as seen at the beginning of treatment. Genetic heterogeneity within a tumor was observed in many cancer entities and plays a role in resistance development (McMahon *et al.*, 2019). Therefore, I dissected the clonal evolution of two CLL patients during ibrutinib treatment by a targeted single cell DNA-sequencing approach. The knowledge about the prevalence of a patient's various subclones and the co-occurrence of mutations in single cells as demonstrated in this thesis might help to detect molecular vulnerabilities of certain subclones and guide therapy decisions. Its value for patient care was already demonstrated for *FLT3*-mutated AML patients (McMahon *et al.*, 2019). *FLT3*-mutated AML can be effectively treated with the FLT3 kinase inhibitor gilterinib as long as no resistance due to secondary mutations occurs. The detection of an *NRAS* mutation co-occurring in a *FLT3*-mutated clone could explain the developed resistance as demonstrated in cell lines. Moreover, treatment of double mutant cells with both Gilterinib and a MEK inhibitor, based on the knowledge about the co-occurring *NRAS* mutation, was able to abrogated resistance (McMahon *et al.*, 2019). Further, the detection of resistant subclones early during therapy would enable a fast adaption of therapy, although currently not applicable in large scale in clinical routine. The mutant subclone could first be detected during remission and was not detectable at treatment start in peripheral blood. This cannot rule out the possibility that the mutant subclone was already present before treatment in a niche of the body. In summary, targeted single-cell DNA sequencing is a powerful approach to disentangle clonal evolution during treatment and resistance development. Deep bulk DNA-sequencing could also

achieve high sensitivity, however the detection of mutational co-occurrences in individual cells is not feasible.

5 Conclusions and outlook

This thesis provides new insights into disease and drug-induced chromatin accessibility changes in CLL and MM. The ATAC-seq analysis workflow established here provides a fast and relatively simple condensed readout of the chromatin state and can be directly conducted with patient samples. Thus, it could be applied in the clinic as an important addition to DNA methylation analysis, an already established readout of epigenetics. It yields valuable information on the activity of regulatory regions with high impact on gene regulation as well as transcription factor binding (Buenrostro *et al.*, 2013). Such a chromatin accessibility analysis revealed a deregulated enhancer landscape in CLL and the associated transcription factors. The value of this analysis was confirmed by the validation of the change in chromatin binding of selected factors by chromatin immunoprecipitation.

The view of CLL as a disease originating from a continuum of B-cell maturation stages has evolved in recent years. The previously conducted analysis of DNA methylation patterns in CLL (Kulis *et al.*, 2015; Oakes *et al.*, 2016; Wierzbinska *et al.*, 2020) was extended in this thesis by an evaluation of chromatin accessibility. On the one hand, both read-outs are equivalent for the assignment of CLL patients to different maturation stages of normal lymphoid development showing only small variations. A further investigation of the cellular origin of CLL could include experimental validation by directly testing the ability of cells in the identified maturation stages to form CLL of phenotypically varying occurrence similar to experiments conducted in AML (Krivtsov *et al.*, 2013). A differing phenotypic occurrence would be expected between tumors formed from more or less matured cells in resemblance to differences observed for IGHV unmutated and mutated cases. The continuous maturation stage assignment using epigenetic markers exceeds the simple binary classification into IGHV mutated and unmutated which has a high prognostic impact and high clinical value. The identification of marker chromatin accessible regions from a bigger cohort would allow a fast determination of the maturity stage of the cells-of-origin of CLL patients from an ATAC-seq data-set. With increased understanding of the disease pathology, it could thus in the future become valuable for clinical decision making. A first step to the evaluation of the prognostic potential of the ATAC-seq based analysis for the clinics could include the application to a bigger patient cohort and the correlation of the resulting maturation stage assignment with information on progression free survival of the patients. A further promising avenue of research in the future could be the identification of the tumor cell-of-origin based on the integration of several readouts of chromatin namely chromatin accessibility and DNA methylation with the transcriptomic layer. It remains to be seen, if and how a prediction based on several molecular phenotypes improves the prediction. On the other hand, the identification of the cell-of-origin from ATAC-seq complements similar analysis based on DNA methylation by providing a different local read-out with valuable

functional information. The novel approach to account for the maturation stage dependent chromatin accessibility pattern of the cell-of-origin of individual patients allows to further improve the identification of disease-induced chromatin changes as shown for CLL. Thus, it becomes now possible to further delineate disease-specific alterations of accessible chromatin patterns by dissecting them from "imprints" of normal epigenetic development preserved from the tumor cell-of-origin. The approach to separate open chromatin loci that reflect the cell-of-origin from the tumor specific ones is not only valuable for CLL research, but might serve as a model for the analysis of epigenetic deregulation in other cancer entities.

The reversibility of disease-specific epigenetic deregulation by treatment can be dissected as shown here for ibrutinib that targets aberrant signaling in CLL. The integration of the different single cell measurements during treatment of CLL patients with ibrutinib would further enable a holistic view of therapy response on the molecular level. The integration of different single cell read-outs is challenging, but could be achieved by copy number variation (CNV) calling. CNV calling from different single cell read-outs of the same patient could enable the assignment of identical subclones between data sets. The analysis of the transcriptome and possibly also epigenome of CLL patients during treatment separately for the observed mutational subclones would provide further insights into the functional implications of the evolution of the present subclones.

Finally, the complex interplay of genetics and epigenetics observed in CLL is of general interest and also applies to other blood cancers. It was studied here with respect to the analysis of the effects of drugs in multiple myeloma. In this entity, the treatment with histone deacetylase inhibitor panobinostat had a global impact on the histone acetylation state of the chromatin as well as cell line specific effects on the transcriptome and phenotypic response. Although targeted, drugs can affect multiple molecular levels of gene regulation. The analysis of the effect of drugs in the therapy of hematopoietic neoplasms on multiple "omics" levels as presented in this thesis and the analysis of intratumor heterogeneity might help to better understand individual patients' response to drugs and the development of resistance. Sequencing-based approaches in clinical decision making will further increase in importance and the extend of usage. Aside from DNA-sequencing, which is already used in clinical practice, other sequencing-based read-outs for the analysis of gene regulation will probably emerge as part of the clinical routine. The analysis of gene regulation including the impact of enhancers could contribute to this development. Enhancers are a so far rather neglected aspect of the deregulation of gene expression in cancer although they are potentially good drug targets. Thus, this thesis explores so far unused capacities for clinical application. By dissecting disease- and drug-induced chromatin accessibility changes at enhancers in lymphoid neoplasms and linking them to transcription factor deregulation and gene expression a better understanding of the disease pathology and drug response is obtained that has the potential to improve personalized patient therapies.

References

Acker, D. (2020). gg3D: 3D perspective plots for ggplot2. R package version 0.0.0.9000.

Adey, A., Morrison, H. G., Asan, Xun, X., Kitzman, J. O., Turner, E. H., Stackhouse, B., MacKenzie, A. P., Caruccio, N. C., Zhang, X., Shendure, J. (2010). Rapid, low-input, low-bias construction of shotgun fragment libraries by high-density in vitro transposition. *Genome Biol, 11*(12), R119. https://doi.org/10.1186/gb-2010-11-12-r119

Agirre, X., Castellano, G., Pascual, M., Heath, S., Kulis, M., Segura, V., Bergmann, A., Esteve, A., Merkel, A., Raineri, E., Agueda, L., Blanc, J., Richardson, D., Clarke, L., Datta, A., Russinol, N., Queiros, A. C., Beekman, R., Rodriguez-Madoz, J. R., San Jose-Eneriz, E., Fang, F., Gutierrez, N. C., Garcia-Verdugo, J. M., Robson, M. I., Schirmer, E. C., Guruceaga, E., Martens, J. H., Gut, M., Calasanz, M. J., Flicek, P., Siebert, R., Campo, E., Miguel, J. F., Melnick, A., Stunnenberg, H. G., Gut, I. G., Prosper, F., Martin-Subero, J. I. (2015). Whole-epigenome analysis in multiple myeloma reveals DNA hypermethylation of B cell-specific enhancers. *Genome Res, 25*(4), 478-487. https://doi.org/10.1101/gr.180240.114

Amemiya, H. M., Kundaje, A., Boyle, A. P. (2019). The ENCODE Blacklist: Identification of Problematic Regions of the Genome. *Sci Rep, 9*(1), 9354. https://doi.org/10.1038/s41598-019-45839-z

Andersson, R., Gebhard, C., Miguel-Escalada, I., Hoof, I., Bornholdt, J., Boyd, M., Chen, Y., Zhao, X., Schmidl, C., Suzuki, T., Ntini, E., Arner, E., Valen, E., Li, K., Schwarzfischer, L., Glatz, D., Raithel, J., Lilje, B., Rapin, N., Bagger, F. O., Jorgensen, M., Andersen, P. R., Bertin, N., Rackham, O., Burroughs, A. M., Baillie, J. K., Ishizu, Y., Shimizu, Y., Furuhata, E., Maeda, S., Negishi, Y., Mungall, C. J., Meehan, T. F., Lassmann, T., Itoh, M., Kawaji, H., Kondo, N., Kawai, J., Lennartsson, A., Daub, C. O., Heutink, P., Hume, D. A., Jensen, T. H., Suzuki, H., Hayashizaki, Y., Muller, F., Forrest, A. R. R., Carninci, P., Rehli, M., Sandelin, A. (2014). An atlas of active enhancers across human cell types and tissues. *Nature, 507*(7493), 455-461. https://doi.org/10.1038/nature12787

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

Arnold, K., Gosling, J., Holmes, D. (2005). The Java programming language. Addison Wesley Professional.

Atadja, P. (2009). Development of the pan-DAC inhibitor panobinostat (LBH589): successes and challenges. *Cancer Lett, 280*(2), 233-241. https://doi.org/10.1016/j.canlet.2009.02.019

Bai, L. & Morozov, A. V. (2010). Gene regulation by nucleosome positioning. *Trends Genet, 26*(11), 476-483. https://doi.org/10.1016/j.tig.2010.08.003

Baliakas, P., Mattsson, M., Stamatopoulos, K., Rosenquist, R. (2016). Prognostic indices in chronic lymphocytic leukaemia: where do we stand how do we proceed? *J Intern Med, 279*(4), 347-357. https://doi.org/10.1111/joim.12455

Banerji, J., Rusconi, S., Schaffner, W. (1981). Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences. *Cell, 27*(2 Pt 1), 299-308. https://doi.org/10.1016/0092-8674(81)90413-x

Barlogie, B. & Gale, R. P. (1992). Multiple myeloma and chronic lymphocytic leukemia: parallels and contrasts. *Am J Med, 93*(4), 443-450. https://doi.org/10.1016/0002-9343(92)90176-c

Barski, A., Cuddapah, S., Cui, K., Roh, T. Y., Schones, D. E., Wang, Z., Wei, G., Chepelev, I., Zhao, K. (2007). High-resolution profiling of histone methylations in the human genome. *Cell, 129*(4), 823-837. https://doi.org/10.1016/j.cell.2007.05.009

Barwick, B. G., Gupta, V. A., Vertino, P. M., Boise, L. H. (2019). Cell of Origin and Genetic Alterations in the Pathogenesis of Multiple Myeloma. *Front Immunol, 10*, 1121. https://doi.org/10.3389/fimmu.2019.01121

Baylin, S. B. & Jones, P. A. (2011). A decade of exploring the cancer epigenome - biological and translational implications. *Nat Rev Cancer, 11*(10), 726-734. https://doi.org/10.1038/nrc3130

Becht, E., McInnes, L., Healy, J., Dutertre, C. A., Kwok, I. W. H., Ng, L. G., Ginhoux, F., Newell, E. W. (2018). Dimensionality reduction for visualizing single-cell data using UMAP. *Nat Biotechnol*. https://doi.org/10.1038/nbt.4314

Beekman, R., Chapaprieta, V., Russinol, N., Vilarrasa-Blasi, R., Verdaguer-Dot, N., Martens, J. H. A., Duran-Ferrer, M., Kulis, M., Serra, F., Javierre, B. M., Wingett, S. W., Clot, G., Queiros, A. C., Castellano, G., Blanc, J., Gut, M., Merkel, A., Heath, S., Vlasova, A., Ullrich, S., Palumbo, E., Enjuanes, A., Martin-Garcia, D., Bea, S., Pinyol, M., Aymerich, M., Royo, R., Puiggros, M., Torrents, D., Datta, A., Lowy, E., Kostadima, M., Roller, M., Clarke, L., Flicek, P., Agirre, X., Prosper, F., Baumann, T., Delgado, J., Lopez-Guillermo, A., Fraser, P., Yaspo, M. L., Guigo, R., Siebert, R., Marti-Renom, M. A., Puente, X. S., Lopez-Otin, C., Gut, I., Stunnenberg, H. G., Campo, E., Martin-Subero, J. I. (2018). The reference epigenome and regulatory chromatin landscape of chronic lymphocytic leukemia. *Nat Med, 24*(6), 868-880. https://doi.org/10.1038/s41591-018-0028-4

Beider, K., Bitner, H., Voevoda-Dimenshtein, V., Rosenberg, E., Sirovsky, Y., Magen, H., Canaani, J., Ostrovsky, O., Shilo, N., Shimoni, A., Abraham, M., Weiss, L., Milyavsky, M., Peled, A., Nagler, A. (2019). The mTOR inhibitor everolimus overcomes CXCR4-mediated resistance to histone deacetylase inhibitor panobinostat through inhibition of p21 and mitotic regulators. *Biochem Pharmacol, 168*, 412-428. https://doi.org/10.1016/j.bcp.2019.07.016

Bembom, O. & Ivanek, R. (2020). seqLogo: Sequence logos for DNA sequence alignments. R package version 1.56.0.

Bernstein, B. E., Stamatoyannopoulos, J. A., Costello, J. F., Ren, B., Milosavljevic, A., Meissner, A., Kellis, M., Marra, M. A., Beaudet, A. L., Ecker, J. R., Farnham, P. J., Hirst, M., Lander, E. S., Mikkelsen, T. S., Thomson, J. A. (2010). The NIH Roadmap Epigenomics Mapping Consortium. *Nat Biotechnol, 28*(10), 1045-1048. https://doi.org/10.1038/nbt1010-1045

Binet, J. L., Auquier, A., Dighiero, G., Chastang, C., Piguet, H., Goasguen, J., Vaugier, G., Potron, G., Colona, P., Oberling, F., Thomas, M., Tchernia, G., Jacquillat, C., Boivin, P., Lesty, C., Duault, M. T., Monconduit, M., Belabbes, S., Gremy, F. (1981). A new prognostic classification of chronic lymphocytic leukemia derived from a multivariate survival analysis. *Cancer, 48*(1), 198-206. https://doi.org/10.1002/1097-0142(19810701)48:1<198::aid-cncr2820480131>3.0.co;2-v

Bogenhagen, D. F. (2012). Mitochondrial DNA nucleoid structure. *Biochim Biophys Acta, 1819*(9-10), 914-920. https://doi.org/10.1016/j.bbagrm.2011.11.005

Bolger, A. M., Lohse, M., Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics, 30*(15), 2114-2120. https://doi.org/10.1093/bioinformatics/btu170

Bolker, B., Butler, M., Cowan, P., de Vienne, D., Eddelbuettel, D., Holder, M., Jombart, T., Kembel, S., Michonneau, F., Orme, D., O'Meara, B., Paradis, E., Regetz, J., Zwickl, D. (2020). phylobase: Base Package for Phylogenetic Structures and Comparative Data. R package version 0.8.10.

Boller, S., Li, R., Grosschedl, R. (2018). Defining B Cell Chromatin: Lessons from EBF1. *Trends Genet, 34*(4), 257-269. https://doi.org/10.1016/j.tig.2017.12.014

Boller, S., Ramamoorthy, S., Akbas, D., Nechanitzky, R., Burger, L., Murr, R., Schubeler, D., Grosschedl, R. (2016). Pioneering Activity of the C-Terminal Domain of EBF1 Shapes the Chromatin Landscape for B Cell Programming. *Immunity, 44*(3), 527-541. https://doi.org/10.1016/j.immuni.2016.02.021

Bolli, N., Avet-Loiseau, H., Wedge, D. C., Van Loo, P., Alexandrov, L. B., Martincorena, I., Dawson, K. J., Iorio, F., Nik-Zainal, S., Bignell, G. R., Hinton, J. W., Li, Y., Tubio, J. M., McLaren, S., S, O. M., Butler, A. P., Teague, J. W., Mudie, L., Anderson, E., Rashid, N., Tai, Y. T., Shammas, M. A., Sperling, A. S., Fulciniti, M., Richardson, P. G., Parmigiani, G., Magrangeas, F., Minvielle, S., Moreau, P., Attal, M., Facon, T., Futreal, P. A., Anderson, K. C., Campbell, P. J., Munshi, N. C. (2014). Heterogeneity of genomic evolution and mutational profiles in multiple myeloma. *Nat Commun, 5*, 2997. https://doi.org/10.1038/ncomms3997

Bolstad, B. (2020). preprocessCore: A collection of pre-processing functions. R package version 1.50.0.

Brien, G. L., Valerio, D. G., Armstrong, S. A. (2016). Exploiting the Epigenome to Control Cancer-Promoting Gene-Expression Programs. *Cancer Cell, 29*(4), 464-476. https://doi.org/10.1016/j.ccell.2016.03.007

Brookes, E. & Shi, Y. (2014). Diverse epigenetic mechanisms of human disease. *Annu Rev Genet, 48*, 237-268. https://doi.org/10.1146/annurev-genet-120213-092518

Buenrostro, J. D., Corces, M. R., Lareau, C. A., Wu, B., Schep, A. N., Aryee, M. J., Majeti, R., Chang, H. Y., Greenleaf, W. J. (2018). Integrated Single-Cell Analysis Maps the Continuous Regulatory Landscape of Human Hematopoietic Differentiation. *Cell, 173*(6), 1535-1548 e1516. https://doi.org/10.1016/j.cell.2018.03.074

Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y., Greenleaf, W. J. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Methods, 10*(12), 1213-1218. https://doi.org/10.1038/nmeth.2688

Buenrostro, J. D., Wu, B., Chang, H. Y., Greenleaf, W. J. (2015a). ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. *Curr Protoc Mol Biol, 109*, 21 29 21-21 29 29. https://doi.org/10.1002/0471142727.mb2129s109

Buenrostro, J. D., Wu, B., Litzenburger, U. M., Ruff, D., Gonzales, M. L., Snyder, M. P., Chang, H. Y., Greenleaf, W. J. (2015b). Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature, 523*(7561), 486-490. https://doi.org/10.1038/nature14590

Burger, J. A. & Chiorazzi, N. (2013). B cell receptor signaling in chronic lymphocytic leukemia. *Trends Immunol, 34*(12), 592-601. https://doi.org/10.1016/j.it.2013.07.002

Burger, J. A., Landau, D. A., Taylor-Weiner, A., Bozic, I., Zhang, H., Sarosiek, K., Wang, L., Stewart, C., Fan, J., Hoellenriegel, J., Sivina, M., Dubuc, A. M., Fraser, C., Han, Y., Li, S., Livak, K. J., Zou, L., Wan, Y., Konoplev, S., Sougnez, C., Brown, J. R., Abruzzo, L. V., Carter, S. L., Keating, M. J., Davids, M. S., Wierda, W. G., Cibulskis, K., Zenz, T., Werner, L., Dal Cin, P., Kharchencko, P., Neuberg, D., Kantarjian, H., Lander, E., Gabriel, S., O'Brien, S., Letai, A., Weitz, D. A., Nowak, M. A., Getz, G., Wu, C. J. (2016). Clonal evolution in patients with chronic lymphocytic leukaemia developing resistance to BTK inhibition. *Nat Commun, 7*, 11589. https://doi.org/10.1038/ncomms11589

Burger, J. A. & O'Brien, S. (2018). Evolution of CLL treatment - from chemoimmunotherapy to targeted and individualized therapy. *Nat Rev Clin Oncol, 15*(8), 510-527. https://doi.org/10.1038/s41571-018-0037-8

Cai, S. F., Chen, C. W., Armstrong, S. A. (2015). Drugging Chromatin in Cancer: Recent Advances and Novel Approaches. *Mol Cell, 60*(4), 561-570. https://doi.org/10.1016/j.molcel.2015.10.042

Carlson, M. (2015). TxDb.Hsapiens.UCSC.hg19.knownGene: Annotation package for TxDb object(s). R package version 3.2.2.

Carlson, M. (2020). org.Hs.eg.db: Genome wide annotation for Human. R package version 3.11.4.

Carter, D., Chakalova, L., Osborne, C. S., Dai, Y. F., Fraser, P. (2002). Long-range chromatin regulatory interactions in vivo. *Nat Genet, 32*(4), 623-626. https://doi.org/10.1038/ng1051

Chaturvedi, A., Herbst, L., Pusch, S., Klett, L., Goparaju, R., Stichel, D., Kaulfuss, S., Panknin, O., Zimmermann, K., Toschi, L., Neuhaus, R., Haegebarth, A., Rehwinkel, H., Hess-Stumpp, H., Bauser, M., Bochtler, T., Struys, E. A., Sharma, A., Bakkali, A., Geffers, R., Araujo-Cruz, M. M., Thol, F., Gabdoulline, R., Ganser, A., Ho, A. D., von Deimling, A., Rippe, K., Heuser, M., Kramer, A. (2017). Pan-mutant-IDH1 inhibitor BAY1436032 is highly effective against human IDH1 mutant acute myeloid leukemia in vivo. *Leukemia, 31*(10), 2020-2028. https://doi.org/10.1038/leu.2017.46

Chen, H., Li, C., Peng, X., Zhou, Z., Weinstein, J. N., Cancer Genome Atlas Research, N., Liang, H. (2018). A Pan-Cancer Analysis of Enhancer Expression in Nearly 9000 Patient Samples. *Cell, 173*(2), 386-399 e312. https://doi.org/10.1016/j.cell.2018.03.027

Chen, L., Glover, J. N., Hogan, P. G., Rao, A., Harrison, S. C. (1998). Structure of the DNA-binding domains from NFAT, Fos and Jun bound specifically to DNA. *Nature, 392*(6671), 42-48. https://doi.org/10.1038/32100

Chen, S. S., Chang, B. Y., Chang, S., Tong, T., Ham, S., Sherry, B., Burger, J. A., Rai, K. R., Chiorazzi, N. (2016). BTK inhibition results in impaired CXCR4 chemokine receptor surface expression, signaling and function in chronic lymphocytic leukemia. *Leukemia, 30*(4), 833-843. https://doi.org/10.1038/leu.2015.316

Chen, S. S., Raval, A., Johnson, A. J., Hertlein, E., Liu, T. H., Jin, V. X., Sherman, M. H., Liu, S. J., Dawson, D. W., Williams, K. E., Lanasa, M., Liyanarachchi, S., Lin, T. S., Marcucci, G., Pekarsky, Y., Davuluri, R., Croce, C. M., Guttridge, D. C., Teitell, M. A., Byrd, J. C., Plass, C. (2009). Epigenetic changes during disease progression in a murine model of human chronic lymphocytic leukemia. *Proc Natl Acad Sci U S A, 106*(32), 13433-13438. https://doi.org/10.1073/pnas.0906455106

Cheng, Y., He, C., Wang, M., Ma, X., Mo, F., Yang, S., Han, J., Wei, X. (2019). Targeting epigenetic regulators for cancer therapy: mechanisms and advances in clinical trials. *Signal Transduction and Targeted Therapy, 4*(1). https://doi.org/10.1038/s41392-019-0095-0

Chng, W. J., Huang, G. F., Chung, T. H., Ng, S. B., Gonzalez-Paz, N., Troska-Price, T., Mulligan, G., Chesi, M., Bergsagel, P. L., Fonseca, R. (2011). Clinical and biological implications of MYC activation: a common difference between MGUS and newly diagnosed multiple myeloma. *Leukemia, 25*(6), 1026-1035. https://doi.org/10.1038/leu.2011.53

Cho, S. W., Xu, J., Sun, R., Mumbach, M. R., Carter, A. C., Chen, Y. G., Yost, K. E., Kim, J., He, J., Nevins, S. A., Chin, S. F., Caldas, C., Liu, S. J., Horlbeck, M. A., Lim, D. A., Weissman, J. S., Curtis, C., Chang, H. Y. (2018). Promoter of lncRNA Gene PVT1 Is a Tumor-Suppressor DNA Boundary Element. *Cell, 173*(6), 1398-1412 e1322. https://doi.org/10.1016/j.cell.2018.03.068

Choudhary, C., Kumar, C., Gnad, F., Nielsen, M. L., Rehman, M., Walther, T. C., Olsen, J. V., Mann, M. (2009). Lysine acetylation targets protein complexes and co-regulates major cellular functions. *Science, 325*(5942), 834-840. https://doi.org/10.1126/science.1175371

Cobaleda, C., Jochum, W., Busslinger, M. (2007). Conversion of mature B cells into T cells by dedifferentiation to uncommitted progenitors. *Nature, 449*(7161), 473-477. https://doi.org/10.1038/nature06159

Cobaleda, C. & Sanchez-Garcia, I. (2009). B-cell acute lymphoblastic leukaemia: towards understanding its cellular origin. *Bioessays, 31*(6), 600-609. https://doi.org/10.1002/bies.200800234

Corces, M. R., Buenrostro, J. D., Wu, B., Greenside, P. G., Chan, S. M., Koenig, J. L., Snyder, M. P., Pritchard, J. K., Kundaje, A., Greenleaf, W. J., Majeti, R., Chang, H. Y. (2016). Lineage-specific and single-cell chromatin accessibility charts human hematopoiesis and leukemia evolution. *Nat Genet, 48*(10), 1193-1203. https://doi.org/10.1038/ng.3646

Corces, M. R., Granja, J. M., Shams, S., Louie, B. H., Seoane, J. A., Zhou, W., Silva, T. C., Groeneveld, C., Wong, C. K., Cho, S. W., Satpathy, A. T., Mumbach, M. R., Hoadley, K. A., Robertson, A. G., Sheffield, N. C., Felau, I., Castro, M. A. A., Berman, B. P., Staudt, L. M., Zenklusen, J. C., Laird, P. W., Curtis, C., Greenleaf, W. J., Chang, H. Y. (2018). The chromatin accessibility landscape of primary human cancers. *Science, 362*(6413). https://doi.org/10.1126/science.aav1898

Corces, M. R., Trevino, A. E., Hamilton, E. G., Greenside, P. G., Sinnott-Armstrong, N. A., Vesuna, S., Satpathy, A. T., Rubin, A. J., Montine, K. S., Wu, B., Kathiria, A., Cho, S. W., Mumbach, M. R., Carter, A. C., Kasowski, M., Orloff, L. A., Risca, V. I., Kundaje, A., Khavari, P. A., Montine, T. J., Greenleaf, W. J., Chang, H. Y. (2017). An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. *Nat Methods, 14*(10), 959-962. https://doi.org/10.1038/nmeth.4396

Creyghton, M. P., Cheng, A. W., Welstead, G. G., Kooistra, T., Carey, B. W., Steine, E. J., Hanna, J., Lodato, M. A., Frampton, G. M., Sharp, P. A., Boyer, L. A., Young, R. A., Jaenisch, R. (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc Natl Acad Sci U S A, 107*(50), 21931-21936. https://doi.org/10.1073/pnas.1016071107

Cunningham, F., Achuthan, P., Akanni, W., Allen, J., Amode, M. R., Armean, I. M., Bennett, R., Bhai, J., Billis, K., Boddu, S., Cummins, C., Davidson, C., Dodiya, K. J., Gall, A., Giron, C. G., Gil, L., Grego, T., Haggerty, L., Haskell, E., Hourlier, T., Izuogu, O. G., Janacek, S. H., Juettemann, T., Kay, M., Laird, M. R., Lavidas, I., Liu, Z., Loveland, J. E., Marugan, J. C., Maurel, T., McMahon, A. C., Moore, B., Morales, J., Mudge, J. M., Nuhn, M., Ogeh, D., Parker, A., Parton, A., Patricio, M., Abdul Salam, A. I., Schmitt, B. M., Schuilenburg, H., Sheppard, D., Sparrow, H., Stapleton, E., Szuba, M., Taylor, K., Threadgold, G., Thormann, A., Vullo, A., Walts, B., Winterbottom, A., Zadissa, A., Chakiachvili, M., Frankish, A., Hunt, S. E., Kostadima, M., Langridge, N., Martin, F. J., Muffato, M., Perry, E., Ruffier, M., Staines, D. M., Trevanion, S. J., Aken, B. L., Yates, A. D., Zerbino, D. R., Flicek, P. (2019). Ensembl 2019. *Nucleic Acids Res, 47*(D1), D745-D751. https://doi.org/10.1093/nar/gky1113

Cusanovich, D. A., Daza, R., Adey, A., Pliner, H., Christiansen, L., Gunderson, K. L., Steemers, F. J., Trapnell, C., Shendure, J. (2015). Multiplex single-cell profiling of chromatin accessibility by combinatorial cellular indexing. *Science, 348*(6237), 910-914. https://doi.org/10.1126/science.aab1601

Damle, R. N., Rao, R., Shastry, P. (1999). Antiproliferative activity of conditioned medium from lymphocytes of neuroblastoma (NB) patient and inhibition with NB serum. *Cancer Lett, 147*(1-2), 5-10. https://doi.org/10.1016/s0304-3835(99)00254-2

de Gorter, D. J., Beuling, E. A., Kersseboom, R., Middendorp, S., van Gils, J. M., Hendriks, R. W., Pals, S. T., Spaargaren, M. (2007). Bruton's tyrosine kinase and phospholipase Cgamma2 mediate chemokine-controlled B cell migration and homing. *Immunity, 26*(1), 93-104. https://doi.org/10.1016/j.immuni.2006.11.012

Deans, C. & Maggert, K. A. (2015). What do you mean, "epigenetic"? *Genetics, 199*(4), 887-896. https://doi.org/10.1534/genetics.114.173492

Dekker, J. & Mirny, L. (2016). The 3D Genome as Moderator of Chromosomal Communication. *Cell, 164*(6), 1110-1121. https://doi.org/10.1016/j.cell.2016.02.007

Deng, W., Lee, J., Wang, H., Miller, J., Reik, A., Gregory, P. D., Dean, A., Blobel, G. A. (2012). Controlling long-range genomic interactions at a native locus by targeted tethering of a looping factor. *Cell, 149*(6), 1233-1244. https://doi.org/10.1016/j.cell.2012.03.051

Desai, A. V., El-Bakkar, H., Abdul-Hay, M. (2015). Novel agents in the treatment of chronic lymphocytic leukemia: a review about the future. *Clin Lymphoma Myeloma Leuk, 15*(6), 314-322. https://doi.org/10.1016/j.clml.2014.09.007

Desper, R. & Gascuel, O. (2002). Fast and accurate phylogeny reconstruction algorithms based on the minimum-evolution principle. *J Comput Biol, 9*(5), 687-705. https://doi.org/10.1089/106652702761034136

Dib, A., Gabrea, A., Glebov, O. K., Bergsagel, P. L., Kuehl, W. M. (2008). Characterization of MYC translocations in multiple myeloma cell lines. *J Natl Cancer Inst Monogr*(39), 25-31. https://doi.org/10.1093/jncimonographs/lgn011

Dimri, G. P., Lee, X., Basile, G., Acosta, M., Scott, G., Roskelley, C., Medrano, E. E., Linskens, M., Rubelj, I., Pereira-Smith, O., et al. (1995). A biomarker that identifies senescent human cells in culture and in aging skin in vivo. *Proc Natl Acad Sci U S A, 92*(20), 9363-9367. https://doi.org/10.1073/pnas.92.20.9363

Dixon, J. R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J. S., Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature, 485*(7398), 376-380. https://doi.org/10.1038/nature11082

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., Gingeras, T. R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics, 29*(1), 15-21. https://doi.org/10.1093/bioinformatics/bts635

Dohner, H., Stilgenbauer, S., Benner, A., Leupolt, E., Krober, A., Bullinger, L., Dohner, K., Bentz, M., Lichter, P. (2000). Genomic aberrations and survival in chronic lymphocytic leukemia. *N Engl J Med, 343*(26), 1910-1916. https://doi.org/10.1056/NEJM200012283432602

Dormoy, V., Jacqmin, D., Lang, H., Massfelder, T. (2012). From development to cancer: lessons from the kidney to uncover new therapeutic targets. *Anticancer Res, 32*(9), 3609-3617. https://www.ncbi.nlm.nih.gov/pubmed/22993298

Dreßler, L. (2017). Identifying gene expression targets of the deacetylase inhibitor panobinostat in multiple myeloma cell lines [Master's thesis, Ruperto Carola University Heidelberg, Heidelberg, Germany].

Ellis, B., Haaland, P., Hahne, F., Le Meur, N., Gopalakrishnan, N., Spidlen, J., Jiang, M., Finak, G. (2019). flowCore: Basic structures for flow cytometry data. R package version 1.52.1.

Encode Project Consortium. (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature, 489*(7414), 57-74. https://doi.org/10.1038/nature11247

Erarslan-Uysal, B., Kunz, J. B., Rausch, T., Richter-Pechanska, P., van Belzen, I. A., Frismantas, V., Bornhauser, B., Ordonez-Rueada, D., Paulsen, M., Benes, V., Stanulla, M., Schrappe, M., Cario, G., Escherich, G., Bakharevich, K., Kirschner-Schwabe, R., Eckert, C., Loukanov, T., Gorenflo, M., Waszak, S. M., Bourquin, J. P., Muckenthaler, M. U., Korbel, J. O., Kulozik, A. E. (2020). Chromatin accessibility landscape of pediatric T-lymphoblastic leukemia and human T-cell precursors. *EMBO Mol Med, 12*(9), e12104. https://doi.org/10.15252/emmm.202012104

Ernst, J. & Kellis, M. (2012). ChromHMM: automating chromatin-state discovery and characterization. *Nat Methods, 9*(3), 215-216. https://doi.org/10.1038/nmeth.1906

Ewald, J. A., Desotelle, J. A., Wilding, G., Jarrard, D. F. (2010). Therapy-induced senescence in cancer. *J Natl Cancer Inst, 102*(20), 1536-1546. https://doi.org/10.1093/jnci/djq364

Ewels, P., Magnusson, M., Lundin, S., Kaller, M. (2016). MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics, 32*(19), 3047-3048. https://doi.org/10.1093/bioinformatics/btw354

Fais, F., Ghiotto, F., Hashimoto, S., Sellars, B., Valetto, A., Allen, S. L., Schulman, P., Vinciguerra, V. P., Rai, K., Rassenti, L. Z., Kipps, T. J., Dighiero, G., Schroeder, H. W., Jr., Ferrarini, M., Chiorazzi, N. (1998). Chronic lymphocytic leukemia B cells express restricted sets of mutated and unmutated antigen receptors. *J Clin Invest, 102*(8), 1515-1525. https://doi.org/10.1172/JCI3009

Filtz, T. M., Vogel, W. K., Leid, M. (2014). Regulation of transcription factor activity by interconnected post-translational modifications. *Trends Pharmacol Sci, 35*(2), 76-85. https://doi.org/10.1016/j.tips.2013.11.005

Furlong, E. M. & Levine, M. (2018). Developmental enhancers and chromosome topology. *Science, 361*(6409), 1341-1345. https://doi.org/10.1126/science.aau0320

Gabrea, A., Bergsagel, P. L., Chesi, M., Shou, Y., Kuehl, W. M. (1999). Insertion of excised IgH switch sequences causes overexpression of cyclin D1 in a myeloma tumor cell. *Mol Cell, 3*(1), 119-123. https://doi.org/10.1016/s1097-2765(00)80180-x

Gaidano, G. & Rossi, D. (2017). The mutational landscape of chronic lymphocytic leukemia and its impact on prognosis and treatment. *Hematology Am Soc Hematol Educ Program, 2017*(1), 329-337. https://doi.org/10.1182/asheducation-2017.1.329

Galili, T. (2015). dendextend: an R package for visualizing, adjusting and comparing trees of hierarchical clustering. *Bioinformatics, 31*(22), 3718-3720. https://doi.org/10.1093/bioinformatics/btv428

Garnier, S. (2018). viridis: Default Color Maps from 'matplotlib'. R package version 0.5.1.

Gasperini, M., Tome, J. M., Shendure, J. (2020). Towards a comprehensive catalogue of validated and target-linked human enhancers. *Nat Rev Genet, 21*(5), 292-310. https://doi.org/10.1038/s41576-019-0209-0

George, J., Uyar, A., Young, K., Kuffler, L., Waldron-Francis, K., Marquez, E., Ucar, D., Trowbridge, J. J. (2016). Leukaemia cell of origin identified by chromatin landscape of bulk tumour cells. *Nat Commun, 7*, 12166. https://doi.org/10.1038/ncomms12166

Ghia, P., Stamatopoulos, K., Belessi, C., Moreno, C., Stilgenbauer, S., Stevenson, F., Davi, F., Rosenquist, R., European Research Initiative on, C. L. L. (2007). ERIC recommendations on IGHV gene mutational status analysis in chronic lymphocytic leukemia. *Leukemia, 21*(1), 1-3. https://doi.org/10.1038/sj.leu.2404457

Gillette, T. G. & Hill, J. A. (2015). Readers, writers, and erasers: chromatin as the whiteboard of heart disease. *Circ Res, 116*(7), 1245-1253. https://doi.org/10.1161/CIRCRESAHA.116.303630

Giresi, P. G., Kim, J., McDaniell, R. M., Iyer, V. R., Lieb, J. D. (2007). FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements) isolates active regulatory elements from human chromatin. *Genome Res, 17*(6), 877-885. https://doi.org/10.1101/gr.5533506

Gonzalez, D., van der Burg, M., Garcia-Sanz, R., Fenton, J. A., Langerak, A. W., Gonzalez, M., van Dongen, J. J., San Miguel, J. F., Morgan, G. J. (2007). Immunoglobulin gene rearrangements and the pathogenesis of multiple myeloma. *Blood, 110*(9), 3112-3121. https://doi.org/10.1182/blood-2007-02-069625

Gorkin, D. U., Leung, D., Ren, B. (2014). The 3D genome in transcriptional regulation and pluripotency. *Cell Stem Cell, 14*(6), 762-775. https://doi.org/10.1016/j.stem.2014.05.017

Greenleaf, W. J. (2015). Assaying the epigenome in limited numbers of cells. *Methods, 72*, 51-56. https://doi.org/10.1016/j.ymeth.2014.10.010

Greenstein, S., Krett, N. L., Kurosawa, Y., Ma, C., Chauhan, D., Hideshima, T., Anderson, K. C., Rosen, S. T. (2003). Characterization of the MM.1 human multiple myeloma (MM) cell lines. *Experimental Hematology, 31*(4), 271-282. https://doi.org/10.1016/s0301-472x(03)00023-7

Group, C. T. C. (1999). Chemotherapeutic options in chronic lymphocytic leukemia: a meta-analysis of the randomized trials. CLL Trialists' Collaborative Group. *J Natl Cancer Inst, 91*(10), 861-868. https://doi.org/10.1093/jnci/91.10.861

Gutierrez, C. & Wu, C. J. (2019). Clonal dynamics in chronic lymphocytic leukemia. *Blood Adv, 3*(22), 3759-3769. https://doi.org/10.1182/bloodadvances.2019000367

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

Hahne, F., Gopalakrishnan, N., Khodabakhshi, A. H., Wong, C.-J., Lee, K. (2020). flowStats: Statistical methods for the analysis of flow cytometry data. R package version 4.1.0.

Hallek, M. (2013). Signaling the end of chronic lymphocytic leukemia: new frontline treatment strategies. *Blood, 122*(23), 3723-3734. https://doi.org/10.1182/blood-2013-05-498287

Hallek, M. (2015). Chronic lymphocytic leukemia: 2015 Update on diagnosis, risk stratification, and treatment. *Am J Hematol, 90*(5), 446-460. https://doi.org/10.1002/ajh.23979

Hamblin, T. J., Davis, Z., Gardiner, A., Oscier, D. G., Stevenson, F. K. (1999). Unmutated Ig V(H) genes are associated with a more aggressive form of chronic lymphocytic leukemia. *Blood, 94*(6), 1848-1854. https://www.ncbi.nlm.nih.gov/pubmed/10477713

Hanna, B. S., Llao Cid, L., Iskar, M., Roessner, P. M., Klett, L. C., Ioannou, N., Öztürk, S., Mack, N., Kalter, V., Colomer, D., Campo, E., Bloehdorn, J., Stilgenbauer, S., Dietrich, S., Schmidt, M., Gabriel, R., Rippe, K., Feuerer, M., Ramsay, A., Zapatka, M., Lichter, P., Seiffert, M. IL-10 receptor signaling limits CD8 T-cell activation preventing their terminal exhaustion and tumor immune escape. *Immunity*. (under revision)

Hay, S. B., Ferchen, K., Chetal, K., Grimes, H. L., Salomonis, N. (2018). The Human Cell Atlas bone marrow single-cell interactive web portal. *Exp Hematol, 68*, 51-61. https://doi.org/10.1016/j.exphem.2018.09.004

Heintzman, N. D., Hon, G. C., Hawkins, R. D., Kheradpour, P., Stark, A., Harp, L. F., Ye, Z., Lee, L. K., Stuart, R. K., Ching, C. W., Ching, K. A., Antosiewicz-Bourget, J. E., Liu, H., Zhang, X., Green, R. D., Lobanenkov, V. V., Stewart, R., Thomson, J. A., Crawford, G. E., Kellis, M., Ren, B. (2009). Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature, 459*(7243), 108-112. https://doi.org/10.1038/nature07829

Heintzman, N. D., Stuart, R. K., Hon, G., Fu, Y., Ching, C. W., Hawkins, R. D., Barrera, L. O., Van Calcar, S., Qu, C., Ching, K. A., Wang, W., Weng, Z., Green, R. D., Crawford, G. E., Ren, B. (2007). Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat Genet, 39*(3), 311-318. https://doi.org/10.1038/ng1966

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y. C., Laslo, P., Cheng, J. X., Murre, C., Singh, H., Glass, C. K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell, 38*(4), 576-589. https://doi.org/10.1016/j.molcel.2010.05.004

Hideshima, T., Cottini, F., Ohguchi, H., Jakubikova, J., Gorgun, G., Mimura, N., Tai, Y. T., Munshi, N. C., Richardson, P. G., Anderson, K. C. (2015). Rational combination treatment with histone deacetylase inhibitors and immunomodulatory drugs in multiple myeloma. *Blood Cancer J, 5*, e312. https://doi.org/10.1038/bcj.2015.38

Hideshima, T., Richardson, P. G., Anderson, K. C. (2011). Mechanism of action of proteasome inhibitors and deacetylase inhibitors and the biological basis of synergy in multiple myeloma. *Mol Cancer Ther, 10*(11), 2034-2042. https://doi.org/10.1158/1535-7163.MCT-11-0433

Ibrahim, M. M., Lacadie, S. A., Ohler, U. (2015). JAMM: a peak finder for joint analysis of NGS replicates. *Bioinformatics, 31*(1), 48-55. https://doi.org/10.1093/bioinformatics/btu568

Jacob, F. & Monod, J. (1961). Genetic regulatory mechanisms in the synthesis of proteins. *J Mol Biol, 3*, 318-356. https://doi.org/10.1016/s0022-2836(61)80072-7

Jain, P., Nogueras Gonzalez, G. M., Kanagal-Shamanna, R., Rozovski, U., Sarwari, N., Tam, C., Wierda, W. G., Thompson, P. A., Jain, N., Luthra, R., Quesada, A., Sanchez-Petitto, G., Ferrajoli, A., Burger, J., Kantarjian, H., Cortes, J., O'Brien, S., Keating, M. J., Estrov, Z. (2018). The absolute percent deviation of IGHV mutation rather than a 98% cut-off predicts survival of chronic lymphocytic leukaemia patients treated with fludarabine, cyclophosphamide and rituximab. *Br J Haematol, 180*(1), 33-40. https://doi.org/10.1111/bjh.15018

Jenuwein, T. & Allis, C. D. (2001). Translating the histone code. *Science, 293*(5532), 1074-1080. https://doi.org/10.1126/science.1063127

Jin, C., Kato, K., Chimura, T., Yamasaki, T., Nakade, K., Murata, T., Li, H., Pan, J., Zhao, M., Sun, K., Chiu, R., Ito, T., Nagata, K., Horikoshi, M., Yokoyama, K. K. (2006). Regulation of histone acetylation and nucleosome assembly by transcription factor JDP2. *Nat Struct Mol Biol, 13*(4), 331-338. https://doi.org/10.1038/nsmb1063

Jin, Y., Chen, K., De Paepe, A., Hellqvist, E., Krstic, A. D., Metang, L., Gustafsson, C., Davis, R. E., Levy, Y. M., Surapaneni, R., Wallblom, A., Nahi, H., Mansson, R., Lin, Y. C. (2018). Active enhancer and chromatin accessibility landscapes chart the regulatory network of primary multiple myeloma. *Blood, 131*(19), 2138-2150. https://doi.org/10.1182/blood-2017-09-808063

Jogi, A., Vaapil, M., Johansson, M., Pahlman, S. (2012). Cancer cell differentiation heterogeneity and aggressive behavior in solid tumors. *Ups J Med Sci, 117*(2), 217-224. https://doi.org/10.3109/03009734.2012.659294

Jung, Y. L., Luquette, L. J., Ho, J. W., Ferrari, F., Tolstorukov, M., Minoda, A., Issner, R., Epstein, C. B., Karpen, G. H., Kuroda, M. I., Park, P. J. (2014). Impact of sequencing depth in ChIP-seq experiments. *Nucleic Acids Res, 42*(9), e74. https://doi.org/10.1093/nar/gku178

Kanehisa, M. & Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res, 28*(1), 27-30. https://doi.org/10.1093/nar/28.1.27

Kent, W. J., Zweig, A. S., Barber, G., Hinrichs, A. S., Karolchik, D. (2010). BigWig and BigBed: enabling browsing of large distributed datasets. *Bioinformatics, 26*(17), 2204-2207. https://doi.org/10.1093/bioinformatics/btq351

Kim, T. K., Hemberg, M., Gray, J. M., Costa, A. M., Bear, D. M., Wu, J., Harmin, D. A., Laptewicz, M., Barbara-Haley, K., Kuersten, S., Markenscoff-Papadimitriou, E., Kuhl, D., Bito, H., Worley, P. F., Kreiman, G., Greenberg, M. E. (2010). Widespread transcription at neuronal activity-regulated enhancers. *Nature, 465*(7295), 182-187. https://doi.org/10.1038/nature09033

Klein, U., Rajewsky, K., Kuppers, R. (1998). Human immunoglobulin (Ig)M+IgD+ peripheral blood B cells expressing the CD27 cell surface antigen carry somatically mutated variable region genes: CD27 as a general marker for somatically mutated (memory) B cells. *J Exp Med, 188*(9), 1679-1689. https://doi.org/10.1084/jem.188.9.1679

Klemm, S. L., Shipony, Z., Greenleaf, W. J. (2019). Chromatin accessibility and the regulatory epigenome. *Nat Rev Genet, 20*(4), 207-220. https://doi.org/10.1038/s41576-018-0089-8

Klinke, D. J., 2nd & Brundage, K. M. (2009). Scalable analysis of flow cytometry data using R/Bioconductor. *Cytometry A, 75*(8), 699-706. https://doi.org/10.1002/cyto.a.20746

Kolberg, L. & Raudvere, U. (2020). gprofiler2: Interface to the 'g:Profiler' Toolset. R package version 0.1.9.

Kopylova, E., Noe, L., Touzet, H. (2012). SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. *Bioinformatics, 28*(24), 3211-3217. https://doi.org/10.1093/bioinformatics/bts611

Kouzarides, T. (2000). Acetylation: a regulatory modification to rival phosphorylation? *EMBO J, 19*(6), 1176-1179. https://doi.org/10.1093/emboj/19.6.1176

Kouzarides, T. (2007). Chromatin modifications and their function. *Cell, 128*(4), 693-705. https://doi.org/10.1016/j.cell.2007.02.005

Krause, J. R. (2009). WHO Classification of Tumours of Haematopoietic and Lymphoid Tissues: An Overview. *Critical values, 2*(2), 30-32. https://doi.org/doi:10.1093/criticalvalues/2.2.30

Krivtsov, A. V., Figueroa, M. E., Sinha, A. U., Stubbs, M. C., Feng, Z., Valk, P. J., Delwel, R., Dohner, K., Bullinger, L., Kung, A. L., Melnick, A. M., Armstrong, S. A. (2013). Cell of origin determines clinically relevant subtypes of MLL-rearranged AML. *Leukemia, 27*(4), 852-860. https://doi.org/10.1038/leu.2012.363

Kuilman, T., Michaloglou, C., Mooi, W. J., Peeper, D. S. (2010). The essence of senescence. *Genes Dev, 24*(22), 2463-2479. https://doi.org/10.1101/gad.1971610

Kulis, M., Heath, S., Bibikova, M., Queiros, A. C., Navarro, A., Clot, G., Martinez-Trillos, A., Castellano, G., Brun-Heath, I., Pinyol, M., Barberan-Soler, S., Papasaikas, P., Jares, P., Bea, S., Rico, D., Ecker, S., Rubio, M., Royo, R., Ho, V., Klotzle, B., Hernandez, L., Conde, L., Lopez-Guerra, M., Colomer, D., Villamor, N., Aymerich, M., Rozman, M., Bayes, M., Gut, M., Gelpi, J. L., Orozco, M., Fan, J. B., Quesada, V., Puente, X. S., Pisano, D. G., Valencia, A., Lopez-Guillermo, A., Gut, I., Lopez-Otin, C., Campo, E., Martin-Subero, J. I. (2012). Epigenomic analysis detects widespread gene-body DNA hypomethylation in chronic lymphocytic leukemia. *Nat Genet, 44*(11), 1236-1242. https://doi.org/10.1038/ng.2443

Kulis, M., Merkel, A., Heath, S., Queiros, A. C., Schuyler, R. P., Castellano, G., Beekman, R., Raineri, E., Esteve, A., Clot, G., Verdaguer-Dot, N., Duran-Ferrer, M., Russinol, N., Vilarrasa-Blasi, R., Ecker, S., Pancaldi, V., Rico, D., Agueda, L., Blanc, J., Richardson, D., Clarke, L., Datta, A., Pascual, M., Agirre, X., Prosper, F., Alignani, D., Paiva, B., Caron, G., Fest, T., Muench, M. O., Fomin, M. E., Lee, S. T., Wiemels, J. L., Valencia, A., Gut, M., Flicek, P., Stunnenberg, H. G., Siebert, R., Kuppers, R., Gut, I. G., Campo, E., Martin-Subero, J. I. (2015). Whole-genome fingerprint of the DNA methylome during human B cell differentiation. *Nat Genet, 47*(7), 746-756. https://doi.org/10.1038/ng.3291

Kumar, S. K., Rajkumar, V., Kyle, R. A., van Duin, M., Sonneveld, P., Mateos, M. V., Gay, F., Anderson, K. C. (2017). Multiple myeloma. *Nat Rev Dis Primers, 3*, 17046. https://doi.org/10.1038/nrdp.2017.46

Kundaje, A. (2013). run_spp.R Computer Science Dept., MIT, ENCODE Consortium, http://code.google.com/p/phantompeakqualtools.

Kyle, R. A., Therneau, T. M., Rajkumar, S. V., Larson, D. R., Plevak, M. F., Offord, J. R., Dispenzieri, A., Katzmann, J. A., Melton, L. J., 3rd. (2006). Prevalence of monoclonal gammopathy of undetermined significance. *N Engl J Med, 354*(13), 1362-1369. https://doi.org/10.1056/NEJMoa054494

Ladewig, J., Koch, P., Brustle, O. (2013). Leveling Waddington: the emergence of direct programming and the loss of cell fate hierarchies. *Nat Rev Mol Cell Biol, 14*(4), 225-236. https://doi.org/10.1038/nrm3543

Lambert, S. A., Jolma, A., Campitelli, L. F., Das, P. K., Yin, Y., Albu, M., Chen, X., Taipale, J., Hughes, T. R., Weirauch, M. T. (2018). The Human Transcription Factors. *Cell, 172*(4), 650-665. https://doi.org/10.1016/j.cell.2018.01.029

Landau, D. A., Clement, K., Ziller, M. J., Boyle, P., Fan, J., Gu, H., Stevenson, K., Sougnez, C., Wang, L., Li, S., Kotliar, D., Zhang, W., Ghandi, M., Garraway, L., Fernandes, S. M., Livak, K. J., Gabriel, S., Gnirke, A., Lander, E. S., Brown, J. R., Neuberg, D., Kharchenko, P. V., Hacohen, N., Getz, G., Meissner, A., Wu, C. J. (2014). Locally disordered methylation forms the basis of intratumor methylome variation in chronic lymphocytic leukemia. *Cancer Cell, 26*(6), 813-825. https://doi.org/10.1016/j.ccell.2014.10.012

Landau, D. A., Tausch, E., Taylor-Weiner, A. N., Stewart, C., Reiter, J. G., Bahlo, J., Kluth, S., Bozic, I., Lawrence, M., Bottcher, S., Carter, S. L., Cibulskis, K., Mertens, D., Sougnez, C. L., Rosenberg, M., Hess, J. M., Edelmann, J., Kless, S., Kneba, M., Ritgen, M., Fink, A., Fischer, K., Gabriel, S., Lander, E. S., Nowak, M. A., Dohner, H., Hallek, M., Neuberg, D., Getz, G., Stilgenbauer, S., Wu, C. J. (2015). Mutations driving CLL and their evolution in progression and relapse. *Nature, 526*(7574), 525-530. https://doi.org/10.1038/nature15395

Landgren, O., Kyle, R. A., Pfeiffer, R. M., Katzmann, J. A., Caporaso, N. E., Hayes, R. B., Dispenzieri, A., Kumar, S., Clark, R. J., Baris, D., Hoover, R., Rajkumar, S. V. (2009). Monoclonal gammopathy of undetermined significance (MGUS) consistently precedes multiple myeloma: a prospective study. *Blood, 113*(22), 5412-5417. https://doi.org/10.1182/blood-2008-12-194241

Langmead, B. & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat Methods, 9*(4), 357-359. https://doi.org/10.1038/nmeth.1923

Langmead, B., Trapnell, C., Pop, M., Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol, 10*(3), R25. https://doi.org/10.1186/gb-2009-10-3-r25

Lawrence, M. S., Stojanov, P., Polak, P., Kryukov, G. V., Cibulskis, K., Sivachenko, A., Carter, S. L., Stewart, C., Mermel, C. H., Roberts, S. A., Kiezun, A., Hammerman, P. S., McKenna, A., Drier, Y., Zou, L., Ramos, A. H., Pugh, T. J., Stransky, N., Helman, E., Kim, J., Sougnez, C., Ambrogio, L., Nickerson, E., Shefler, E., Cortes, M. L., Auclair, D., Saksena, G., Voet, D., Noble, M., DiCara, D., Lin, P., Lichtenstein, L., Heiman, D. I., Fennell, T., Imielinski, M., Hernandez, B., Hodis, E., Baca, S., Dulak, A. M., Lohr, J., Landau, D. A., Wu, C. J., Melendez-Zajgla, J., Hidalgo-Miranda, A., Koren, A., McCarroll, S. A., Mora, J., Crompton, B., Onofrio, R., Parkin, M., Winckler, W., Ardlie, K., Gabriel, S. B., Roberts, C. W. M., Biegel, J. A., Stegmaier, K., Bass, A. J., Garraway, L. A., Meyerson, M., Golub, T. R., Gordenin, D. A., Sunyaev, S., Lander, E. S., Getz, G. (2013). Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature, 499*(7457), 214-218. https://doi.org/10.1038/nature12213

Lee, T. I. & Young, R. A. (2013). Transcriptional regulation and its misregulation in disease. *Cell, 152*(6), 1237-1251. https://doi.org/10.1016/j.cell.2013.02.014

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv preprint arXiv:1303. 2013:3997*.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., Genome Project Data Processing, S. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics, 25*(16), 2078-2079. https://doi.org/10.1093/bioinformatics/btp352

Li, P., Xie, L., Gu, Y., Li, J., Xie, J. (2017). Roles of Multifunctional COP9 Signalosome Complex in Cell Fate and Implications for Drug Discovery. *J Cell Physiol, 232*(6), 1246-1253. https://doi.org/10.1002/jcp.25696

Look, A. T. (1997). Oncogenic transcription factors in the human acute leukemias. *Science, 278*(5340), 1059-1064. https://doi.org/10.1126/science.278.5340.1059

Love, M. I., Huber, W., Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol, 15*(12), 1-21. https://doi.org/10.1186/s13059-014-0550-8

Loven, J., Hoke, H. A., Lin, C. Y., Lau, A., Orlando, D. A., Vakoc, C. R., Bradner, J. E., Lee, T. I., Young, R. A. (2013). Selective inhibition of tumor oncogenes by disruption of super-enhancers. *Cell, 153*(2), 320-334. https://doi.org/10.1016/j.cell.2013.03.036

Lozzio, C. B. & Lozzio, B. B. (1975). Human chronic myelogenous leukemia cell-line with positive Philadelphia chromosome. *Blood, 45*(3), 321-334. https://doi.org/10.1182/blood.V45.3.321.321

Lu, C., Ward, P. S., Kapoor, G. S., Rohle, D., Turcan, S., Abdel-Wahab, O., Edwards, C. R., Khanin, R., Figueroa, M. E., Melnick, A., Wellen, K. E., O'Rourke, D. M., Berger, S. L., Chan, T. A., Levine, R. L., Mellinghoff, I. K., Thompson, C. B. (2012). IDH mutation impairs histone demethylation and results in a block to cell differentiation. *Nature, 483*(7390), 474-478. https://doi.org/10.1038/nature10860

Lucas, A. (2019). amap: Another Multidimensional Analysis Package. R package version 0.8-18.

Ludwig, L. S., Lareau, C. A., Ulirsch, J. C., Christian, E., Muus, C., Li, L. H., Pelka, K., Ge, W., Oren, Y., Brack, A., Law, T., Rodman, C., Chen, J. H., Boland, G. M., Hacohen, N., Rozenblatt-Rosen, O., Aryee, M. J., Buenrostro, J. D., Regev, A., Sankaran, V. G. (2019). Lineage Tracing in Humans Enabled by Mitochondrial Mutations and Single-Cell Genomics. *Cell, 176*(6), 1325-1339 e1322. https://doi.org/10.1016/j.cell.2019.01.022

Luger, K., Mader, A. W., Richmond, R. K., Sargent, D. F., Richmond, T. J. (1997). Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature, 389*(6648), 251-260. https://doi.org/10.1038/38444

Luo, W. & Brouwer, C. (2013). Pathview: an R/Bioconductor package for pathway-based data integration and visualization. *Bioinformatics, 29*(14), 1830-1831. https://doi.org/10.1093/bioinformatics/btt285

Macian, F., Lopez-Rodriguez, C., Rao, A. (2001). Partners in transcription: NFAT and AP-1. *Oncogene, 20*(19), 2476-2489. https://doi.org/10.1038/sj.onc.1204386

Maier, H., Ostraat, R., Gao, H., Fields, S., Shinton, S. A., Medina, K. L., Ikawa, T., Murre, C., Singh, H., Hardy, R. R., Hagman, J. (2004). Early B cell factor cooperates with Runx1 and mediates epigenetic changes associated with mb-1 transcription. *Nat Immunol, 5*(10), 1069-1077. https://doi.org/10.1038/ni1119

Maison, C. & Almouzni, G. (2004). HP1 and the dynamics of heterochromatin maintenance. *Nat Rev Mol Cell Biol, 5*(4), 296-304. https://doi.org/10.1038/nrm1355

Malek, S. N. (2013). The biology and clinical significance of acquired genomic copy number aberrations and recurrent gene mutations in chronic lymphocytic leukemia. *Oncogene, 32*(23), 2805-2817. https://doi.org/10.1038/onc.2012.411

Mallm, J. P., Iskar, M., Ishaque, N., Klett, L. C., Kugler, S. J., Muino, J. M., Teif, V. B., Poos, A. M., Grossmann, S., Erdel, F., Tavernari, D., Koser, S. D., Schumacher, S., Brors, B., Konig, R., Remondini, D., Vingron, M., Stilgenbauer, S., Lichter, P., Zapatka, M., Mertens, D., Rippe, K. (2019). Linking aberrant chromatin features in chronic lymphocytic leukemia to transcription factor networks. *Mol Syst Biol, 15*(5), e8339. https://doi.org/10.15252/msb.20188339

Manier, S., Salem, K. Z., Park, J., Landau, D. A., Getz, G., Ghobrial, I. M. (2017). Genomic complexity of multiple myeloma and its clinical implications. *Nat Rev Clin Oncol, 14*(2), 100-113. https://doi.org/10.1038/nrclinonc.2016.122

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal, 17*(1), 10-12. https://doi.org/10.14806/ej.17.1.200

Martinez-Garcia, E., Popovic, R., Min, D. J., Sweet, S. M., Thomas, P. M., Zamdborg, L., Heffner, A., Will, C., Lamy, L., Staudt, L. M., Levens, D. L., Kelleher, N. L., Licht, J. D. (2011). The MMSET histone methyl transferase switches global histone methylation and alters gene expression in t(4;14) multiple myeloma cells. *Blood, 117*(1), 211-220. https://doi.org/10.1182/blood-2010-07-298349

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., DePristo, M. A. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res, 20*(9), 1297-1303. https://doi.org/10.1101/gr.107524.110

McMahon, C. M., Ferng, T., Canaani, J., Wang, E. S., Morrissette, J. J. D., Eastburn, D. J., Pellegrino, M., Durruthy-Durruthy, R., Watt, C. D., Asthana, S., Lasater, E. A., DeFilippis, R., Peretz, C. A. C., McGary, L. H. F., Deihimi, S., Logan, A. C., Luger, S. M., Shah, N. P., Carroll, M., Smith, C. C., Perl, A. E. (2019). Clonal Selection with RAS Pathway Activation Mediates Secondary Clinical Resistance to Selective FLT3 Inhibition in Acute Myeloid Leukemia. *Cancer Discov, 9*(8), 1050-1063. https://doi.org/10.1158/2159-8290.CD-18-1453

Mevik, B.-H., Wehrens, R., Liland, K. H. (2019). pls: Partial Least Squares and Principal Component Regression. R package version 2.7-2.

Mikkilineni, L. & Kochenderfer, J. N. (2021). CAR T cell therapies for patients with multiple myeloma. *Nat Rev Clin Oncol, 18*(2), 71-84. https://doi.org/10.1038/s41571-020-0427-6

Mithraprabhu, S., Kalff, A., Chow, A., Khong, T., Spencer, A. (2014). Dysregulated Class I histone deacetylases are indicators of poor prognosis in multiple myeloma. *Epigenetics, 9*(11), 1511-1520. https://doi.org/10.4161/15592294.2014.983367

Molitor, J., Mallm, J. P., Rippe, K., Erdel, F. (2017). Retrieving Chromatin Patterns from Deep Sequencing Data Using Correlation Functions. *Biophys J, 112*(3), 473-490. https://doi.org/10.1016/j.bpj.2017.01.001

Montefiori, L., Hernandez, L., Zhang, Z., Gilad, Y., Ober, C., Crawford, G., Nobrega, M., Jo Sakabe, N. (2017). Reducing mitochondrial reads in ATAC-seq using CRISPR/Cas9. *Sci Rep, 7*(1), 2451. https://doi.org/10.1038/s41598-017-02547-w

Moreau, P., Hen, R., Wasylyk, B., Everett, R., Gaub, M. P., Chambon, P. (1981). The SV40 72 base repair repeat has a striking effect on gene expression both in SV40 and other chimeric recombinants. *Nucleic Acids Res, 9*(22), 6047-6068. https://doi.org/10.1093/nar/9.22.6047

Mraz, M., Chen, L., Rassenti, L. Z., Ghia, E. M., Li, H., Jepsen, K., Smith, E. N., Messer, K., Frazer, K. A., Kipps, T. J. (2014). miR-150 influences B-cell receptor signaling in chronic lymphocytic leukemia by regulating expression of GAB1 and FOXP1. *Blood, 124*(1), 84-95. https://doi.org/10.1182/blood-2013-09-527234

Muckenhuber, M., Lander, I., Müller-Ott, K., Erdel, F., Mallm, J.-P., Klett, L. C., Bauer, C., Hechler, J., Kepper, N., Rippe, K. Epigenetic signals that direct cell type specific interferon beta response in mouse cells. (submitted)

Nagel, S., Ehrentraut, S., Meyer, C., Kaufmann, M., Drexler, H. G., MacLeod, R. A. (2015). Aberrantly Expressed OTX Homeobox Genes Deregulate B-Cell Differentiation in Hodgkin Lymphoma. *PLoS One, 10*(9), e0138416. https://doi.org/10.1371/journal.pone.0138416

Namba, M., Ohtsuki, T., Mori, M., Togawa, A., Wada, H., Sugihara, T., Yawata, Y., Kimoto, T. (1989). Establishment of five human myeloma cell lines. *In Vitro Cell Dev Biol, 25*(8), 723-729. https://doi.org/10.1007/BF02623725

Nilsson, K., Bennich, H., Johansson, S. G., Ponten, J. (1970). Established immunoglobulin producing myeloma (IgE) and lymphoblastoid (IgG) cell lines from an IgE myeloma patient. *Clin Exp Immunol, 7*(4), 477-489. https://www.ncbi.nlm.nih.gov/pubmed/4097745

O'Leary, N. A., Wright, M. W., Brister, J. R., Ciufo, S., Haddad, D., McVeigh, R., Rajput, B., Robbertse, B., Smith-White, B., Ako-Adjei, D., Astashyn, A., Badretdin, A., Bao, Y., Blinkova, O., Brover, V., Chetvernin, V., Choi, J., Cox, E., Ermolaeva, O., Farrell, C. M., Goldfarb, T., Gupta, T., Haft, D., Hatcher, E., Hlavina, W., Joardar, V. S., Kodali, V. K., Li, W., Maglott, D., Masterson, P., McGarvey, K. M., Murphy, M. R., O'Neill, K., Pujar, S., Rangwala, S. H., Rausch, D., Riddick, L. D., Schoch, C., Shkeda, A., Storz, S. S., Sun, H., Thibaud-Nissen, F., Tolstoy, I., Tully, R. E., Vatsan, A. R., Wallin, C., Webb, D., Wu, W., Landrum, M. J., Kimchi, A., Tatusova, T., DiCuccio, M., Kitts, P., Murphy, T. D., Pruitt, K. D. (2016). Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Research, 44*(D1), D733-D745. https://doi.org/10.1093/nar/gkv1189

Oakes, C. C., Claus, R., Gu, L., Assenov, Y., Hullein, J., Zucknick, M., Bieg, M., Brocks, D., Bogatyrova, O., Schmidt, C. R., Rassenti, L., Kipps, T. J., Mertens, D., Lichter, P., Dohner, H., Stilgenbauer, S., Byrd, J. C., Zenz, T., Plass, C. (2014). Evolution of DNA methylation is linked to genetic aberrations in chronic lymphocytic leukemia. *Cancer Discov, 4*(3), 348-361. https://doi.org/10.1158/2159-8290.CD-13-0349

Oakes, C. C., Seifert, M., Assenov, Y., Gu, L., Przekopowitz, M., Ruppert, A. S., Wang, Q., Imbusch, C. D., Serva, A., Koser, S. D., Brocks, D., Lipka, D. B., Bogatyrova, O., Weichenhan, D., Brors, B., Rassenti, L., Kipps, T. J., Mertens, D., Zapatka, M., Lichter, P., Dohner, H., Kuppers, R., Zenz, T., Stilgenbauer, S., Byrd, J. C., Plass, C. (2016). DNA methylation dynamics during B cell maturation underlie a continuum of disease phenotypes in chronic lymphocytic leukemia. *Nat Genet, 48*(3), 253-264. https://doi.org/10.1038/ng.3488

Oscier, D. G., Gardiner, A. C., Mould, S. J., Glide, S., Davis, Z. A., Ibbotson, R. E., Corcoran, M. M., Chapman, R. M., Thomas, P. W., Copplestone, J. A., Orchard, J. A., Hamblin, T. J. (2002). Multivariate analysis of prognostic factors in CLL: clinical stage, IGVH gene mutational status, and loss or mutation of the p53 gene are independent prognostic factors. *Blood, 100*(4), 1177-1184. https://doi.org/10.1182/blood.V100.4.1177.h81602001177_1177_1184

Palma, M., Krstic, A., Peña Perez, L., Berglöf, A., Meinke, S., Wang, Q., Blomberg, K. E. M., Kamali-Moghaddam, M., Shen, Q., Jaremko, G., Lundin, J., De Paepe, A., Höglund, P., Kimby, E., Österborg, A., Månsson, R., Smith, C. I. E. (2018). Ibrutinib induces rapid down-regulation of inflammatory markers and altered transcription of chronic lymphocytic leukaemia-related genes in blood and lymph nodes. *British Journal of Haematology, 183*(2), 212-224. https://doi.org/10.1111/bjh.15516

Paradis, E. & Schliep, K. (2019). ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics, 35*(3), 526-528. https://doi.org/10.1093/bioinformatics/bty633

Park, P. J. (2009). ChIP-seq: advantages and challenges of a maturing technology. *Nat Rev Genet, 10*(10), 669-680. https://doi.org/10.1038/nrg2641

Pei, L., Choi, J. H., Liu, J., Lee, E. J., McCarthy, B., Wilson, J. M., Speir, E., Awan, F., Tae, H., Arthur, G., Schnabel, J. L., Taylor, K. H., Wang, X., Xu, D., Ding, H. F., Munn, D. H., Caldwell, C., Shi, H. (2012). Genome-wide DNA methylation analysis reveals novel epigenetic changes in chronic lymphocytic leukemia. *Epigenetics, 7*(6), 567-578. https://doi.org/10.4161/epi.20237

Plass, C., Pfister, S. M., Lindroth, A. M., Bogatyrova, O., Claus, R., Lichter, P. (2013). Mutations in regulators of the epigenome and their connections to global chromatin patterns in cancer. *Nat Rev Genet, 14*(11), 765-780. https://doi.org/10.1038/nrg3554

Ponts, N., Harris, E. Y., Prudhomme, J., Wick, I., Eckhardt-Ludka, C., Hicks, G. R., Hardiman, G., Lonardi, S., Le Roch, K. G. (2010). Nucleosome landscape and control of transcription in the human malaria parasite. *Genome Res, 20*(2), 228-238. https://doi.org/10.1101/gr.101063.109

Puente, X. S., Bea, S., Valdes-Mas, R., Villamor, N., Gutierrez-Abril, J., Martin-Subero, J. I., Munar, M., Rubio-Perez, C., Jares, P., Aymerich, M., Baumann, T., Beekman, R., Belver, L., Carrio, A., Castellano, G., Clot, G., Colado, E., Colomer, D., Costa, D., Delgado, J., Enjuanes, A., Estivill, X., Ferrando, A. A., Gelpi, J. L., Gonzalez, B., Gonzalez, S., Gonzalez, M., Gut, M., Hernandez-Rivas, J. M., Lopez-Guerra, M., Martin-Garcia, D., Navarro, A., Nicolas, P., Orozco, M., Payer, A. R., Pinyol, M., Pisano, D. G., Puente, D. A., Queiros, A. C., Quesada, V., Romeo-Casabona, C. M., Royo, C., Royo, R., Rozman, M., Russinol, N., Salaverria, I., Stamatopoulos, K., Stunnenberg, H. G., Tamborero, D., Terol, M. J., Valencia, A., Lopez-Bigas, N., Torrents, D., Gut, I., Lopez-Guillermo, A., Lopez-Otin, C., Campo, E. (2015). Non-coding recurrent mutations in chronic lymphocytic leukaemia. *Nature, 526*(7574), 519-524. https://doi.org/10.1038/nature14666

Pula, B., Golos, A., Gorniak, P., Jamroziak, K. (2019). Overcoming Ibrutinib Resistance in Chronic Lymphocytic Leukemia. *Cancers (Basel), 11*(12). https://doi.org/10.3390/cancers11121834

Qu, K., Zaba, L. C., Satpathy, A. T., Giresi, P. G., Li, R., Jin, Y., Armstrong, R., Jin, C., Schmitt, N., Rahbar, Z., Ueno, H., Greenleaf, W. J., Kim, Y. H., Chang, H. Y. (2017). Chromatin Accessibility Landscape of Cutaneous T Cell Lymphoma and Dynamic Response to HDAC Inhibitors. *Cancer Cell, 32*(1), 27-41 e24. https://doi.org/10.1016/j.ccell.2017.05.008

Queiros, A. C., Villamor, N., Clot, G., Martinez-Trillos, A., Kulis, M., Navarro, A., Penas, E. M., Jayne, S., Majid, A., Richter, J., Bergmann, A. K., Kolarova, J., Royo, C., Russinol, N., Castellano, G., Pinyol, M., Bea, S., Salaverria, I., Lopez-Guerra, M., Colomer, D., Aymerich, M., Rozman, M., Delgado, J., Gine, E., Gonzalez-Diaz, M., Puente, X. S., Siebert, R., Dyer, M. J., Lopez-Otin, C., Rozman, C., Campo, E., Lopez-Guillermo, A., Martin-Subero, J. I. (2015). A B-cell epigenetic signature defines three biologic subgroups of chronic lymphocytic leukemia with clinical impact. *Leukemia, 29*(3), 598-605. https://doi.org/10.1038/leu.2014.252

Quinlan, A. R. & Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics, 26*(6), 841-842. https://doi.org/10.1093/bioinformatics/btq033

Rada-Iglesias, A., Bajpai, R., Swigut, T., Brugmann, S. A., Flynn, R. A., Wysocka, J. (2011). A unique chromatin signature uncovers early developmental enhancers in humans. *Nature, 470*(7333), 279-283. https://doi.org/10.1038/nature09692

Rai KR, S. A., Cronkite EP, Chanana AD, Levy RN, Pasternack BS. (1975). Clinical staging of chronic lymphocytic leukemia. *Blood, 46*(2), 219-234. https://doi.org/10.1182/blood-2016-08-737650

Rajkumar, S. V., Dimopoulos, M. A., Palumbo, A., Blade, J., Merlini, G., Mateos, M.-V., Kumar, S., Hillengass, J., Kastritis, E., Richardson, P., Landgren, O., Paiva, B., Dispenzieri, A., Weiss, B., LeLeu, X., Zweegman, S., Lonial, S., Rosinol, L., Zamagni, E., Jagannath, S., Sezer, O., Kristinsson, S. Y., Caers, J., Usmani, S. Z., Lahuerta, J. J., Johnsen, H. E., Beksac, M., Cavo, M., Goldschmidt, H., Terpos, E., Kyle, R. A., Anderson, K. C., Durie, B. G. M., Miguel, J. F. S. (2014). International Myeloma Working Group updated criteria for the diagnosis of multiple myeloma. *The Lancet Oncology, 15*(12), e538-e548. https://doi.org/10.1016/s1470-2045(14)70442-5

Rakheja, D., Medeiros, L. J., Bevan, S., Chen, W. (2013). The emerging role of d-2-hydroxyglutarate as an oncometabolite in hematolymphoid and central nervous system neoplasms. *Front Oncol, 3*, 169. https://doi.org/10.3389/fonc.2013.00169

Ramamoorthy, S., Kometani, K., Herman, J. S., Bayer, M., Boller, S., Edwards-Hicks, J., Ramachandran, H., Li, R., Klein-Geltink, R., Pearce, E. L., Grun, D., Grosschedl, R. (2020). EBF1 and Pax5 safeguard leukemic transformation by limiting IL-7 signaling, Myc expression, and folate metabolism. *Genes Dev*. https://doi.org/10.1101/gad.340216.120

Rendeiro, A. F., Krausgruber, T., Fortelny, N., Zhao, F., Penz, T., Farlik, M., Schuster, L. C., Nemc, A., Tasnady, S., Reti, M., Matrai, Z., Alpar, D., Bodor, C., Schmidl, C., Bock, C. (2020). Chromatin mapping and single-cell immune profiling define the temporal dynamics of ibrutinib response in CLL. *Nat Commun, 11*(1), 577. https://doi.org/10.1038/s41467-019-14081-6

Rendeiro, A. F., Schmidl, C., Strefford, J. C., Walewska, R., Davis, Z., Farlik, M., Oscier, D., Bock, C. (2016). Chromatin accessibility maps of chronic lymphocytic leukaemia identify subtype-specific epigenome signatures and transcription regulatory networks. *Nat Commun, 7*, 11938. https://doi.org/10.1038/ncomms11938

Reznik, E., Wang, Q., La, K., Schultz, N., Sander, C. (2017). Mitochondrial respiratory gene expression is suppressed in many cancers. *Elife, 6*. https://doi.org/10.7554/eLife.21592

Reznikoff, W. S. (2003). Tn5 as a model for understanding DNA transposition. *Molecular Microbiology, 47*(5), 1199-1206. https://doi.org/10.1046/j.1365-2958.2003.03382.x

Reznikoff, W. S. (2008). Transposon Tn5. *Annu Rev Genet, 42*, 269-286. https://doi.org/10.1146/annurev.genet.42.110807.091656

Richelda, R., Ronchetti, D., Baldini, L., Cro, L., Viggiano, L., Marzella, R., Rocchi, M., Otsuki, T., Lombardi, L., Maiolo, A. T., Neri, A. (1997). A Novel Chromosomal Translocation t(4; 14)(p16.3; q32) in Multiple Myeloma Involves the Fibroblast Growth-Factor Receptor 3 Gene. *Blood, 90*(10), 4062-4070. https://doi.org/10.1182/blood.V90.10.4062

Rickert, R. C. (2013). New insights into pre-BCR and BCR signalling with relevance to B cell malignancies. *Nat Rev Immunol, 13*(8), 578-591. https://doi.org/10.1038/nri3487

Rieger, M. A. & Schroeder, T. (2012). Hematopoiesis. *Cold Spring Harb Perspect Biol, 4*(12). https://doi.org/10.1101/cshperspect.a008250

Roadmap Epigenomics, C., Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., Yen, A., Heravi-Moussavi, A., Kheradpour, P., Zhang, Z., Wang, J., Ziller, M. J., Amin, V., Whitaker, J. W., Schultz, M. D., Ward, L. D., Sarkar, A., Quon, G., Sandstrom, R. S., Eaton, M. L., Wu, Y. C., Pfenning, A. R., Wang, X., Claussnitzer, M., Liu, Y., Coarfa, C., Harris, R. A., Shoresh, N., Epstein, C. B., Gjoneska, E., Leung, D., Xie, W., Hawkins, R. D., Lister, R., Hong, C., Gascard, P., Mungall, A. J., Moore, R., Chuah, E., Tam, A., Canfield, T. K., Hansen, R. S., Kaul, R., Sabo, P. J., Bansal, M. S., Carles, A., Dixon, J. R., Farh, K. H., Feizi, S., Karlic, R., Kim, A. R., Kulkarni, A., Li, D., Lowdon, R., Elliott, G., Mercer, T. R., Neph, S. J., Onuchic, V., Polak, P., Rajagopal, N., Ray, P., Sallari, R. C., Siebenthall, K. T., Sinnott-Armstrong, N. A., Stevens, M., Thurman, R. E., Wu, J., Zhang, B., Zhou, X., Beaudet, A. E., Boyer, L. A., De Jager, P. L., Farnham, P. J., Fisher, S. J., Haussler, D., Jones, S. J., Li, W., Marra, M. A., McManus, M. T., Sunyaev, S., Thomson, J. A., Tlsty, T. D., Tsai, L. H., Wang, W., Waterland, R. A., Zhang, M. Q., Chadwick, L. H., Bernstein, B. E., Costello, J. F., Ecker, J. R., Hirst, M., Meissner, A., Milosavljevic, A., Ren, B., Stamatoyannopoulos, J. A., Wang, T., Kellis, M. (2015). Integrative analysis of 111 reference human epigenomes. *Nature, 518*(7539), 317-330. https://doi.org/10.1038/nature14248

Robinson, J. T., Thorvaldsdottir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G., Mesirov, J. P. (2011). Integrative genomics viewer. *Nat Biotechnol, 29*(1), 24-26. https://doi.org/10.1038/nbt.1754

Ropero, S., Fraga, M. F., Ballestar, E., Hamelin, R., Yamamoto, H., Boix-Chornet, M., Caballero, R., Alaminos, M., Setien, F., Paz, M. F., Herranz, M., Palacios, J., Arango, D., Orntoft, T. F., Aaltonen, L. A., Schwartz, S., Jr., Esteller, M. (2006). A truncating mutation of HDAC2 in human cancers confers resistance to histone deacetylase inhibition. *Nat Genet, 38*(5), 566-569. https://doi.org/10.1038/ng1773

Ross-Innes, C. S., Stark, R., Teschendorff, A. E., Holmes, K. A., Ali, H. R., Dunning, M. J., Brown, G. D., Gojis, O., Ellis, I. O., Green, A. R., Ali, S., Chin, S. F., Palmieri, C., Caldas, C., Carroll, J. S. (2012). Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. *Nature, 481*(7381), 389-393. https://doi.org/10.1038/nature10730

Royston, I., Majda, J. A., Baird, S. M., Meserve, B. L., Griffiths, J. C. (1980). Human T cell antigens defined by monoclonal antibodies: the 65,000-dalton antigen of T cells (T65) is also found on chronic lymphocytic leukemia cells bearing surface immunoglobulin. *J Immunol, 125*(2), 725-731. https://www.ncbi.nlm.nih.gov/pubmed/6993560

Sanchez, G. J., Richmond, P. A., Bunker, E. N., Karman, S. S., Azofeifa, J., Garnett, A. T., Xu, Q., Wheeler, G. E., Toomey, C. M., Zhang, Q., Dowell, R. D., Liu, X. (2018). Genome-wide dose-dependent inhibition of histone deacetylases studies reveal their roles in enhancer remodeling and suppression of oncogenic super-enhancers. *Nucleic Acids Res, 46*(4), 1756-1776. https://doi.org/10.1093/nar/gkx1225

Sarkar, D., Le Meur, N., Gentleman, R. (2008). Using flowViz to visualize flow cytometry data. *Bioinformatics, 24*(6), 878-879. https://doi.org/10.1093/bioinformatics/btn021

Sasaki, M., Knobbe, C. B., Munger, J. C., Lind, E. F., Brenner, D., Brustle, A., Harris, I. S., Holmes, R., Wakeham, A., Haight, J., You-Ten, A., Li, W. Y., Schalm, S., Su, S. M., Virtanen, C., Reifenberger, G., Ohashi, P. S., Barber, D. L., Figueroa, M. E., Melnick, A., Zuniga-Pflucker, J. C., Mak, T. W. (2012). IDH1(R132H) mutation increases murine haematopoietic progenitors and alters epigenetics. *Nature, 488*(7413), 656-659. https://doi.org/10.1038/nature11323

Satpathy, A. T., Granja, J. M., Yost, K. E., Qi, Y., Meschi, F., McDermott, G. P., Olsen, B. N., Mumbach, M. R., Pierce, S. E., Corces, M. R., Shah, P., Bell, J. C., Jhutty, D., Nemec, C. M., Wang, J., Wang, L., Yin, Y., Giresi, P. G., Chang, A. L. S., Zheng, G. X. Y., Greenleaf, W. J., Chang, H. Y. (2019). Massively parallel single-cell chromatin landscapes of human immune cell development and intratumoral T cell exhaustion. *Nat Biotechnol, 37*(8), 925-936. https://doi.org/10.1038/s41587-019-0206-z

Schmidl, C., Rendeiro, A. F., Sheffield, N. C., Bock, C. (2015). ChIPmentation: fast, robust, low-input ChIP-seq for histones and transcription factors. *Nat Methods, 12*(10), 963-965. https://doi.org/10.1038/nmeth.3542

Schoenfelder, S. & Fraser, P. (2019). Long-range enhancer-promoter contacts in gene expression control. *Nat Rev Genet, 20*(8), 437-455. https://doi.org/10.1038/s41576-019-0128-0

Seifert, M., Sellmann, L., Bloehdorn, J., Wein, F., Stilgenbauer, S., Durig, J., Kuppers, R. (2012). Cellular origin and pathophysiology of chronic lymphocytic leukemia. *J Exp Med, 209*(12), 2183-2198. https://doi.org/10.1084/jem.20120833

Shechter, D., Dormann, H. L., Allis, C. D., Hake, S. B. (2007). Extraction, purification and analysis of histones. *Nat Protoc, 2*(6), 1445-1457. https://doi.org/10.1038/nprot.2007.202

Shen, H. & Laird, P. W. (2013). Interplay between the cancer genome and epigenome. *Cell, 153*(1), 38-55. https://doi.org/10.1016/j.cell.2013.03.008

Shen, L., Shao, N., Liu, X., Nestler, E. (2014). ngs.plot: Quick mining and visualization of next-generation sequencing data by integrating genomic databases. *BMC Genomics, 15*, 284. https://doi.org/10.1186/1471-2164-15-284

Shlyueva, D., Stampfel, G., Stark, A. (2014). Transcriptional enhancers: from properties to genome-wide predictions. *Nat Rev Genet, 15*(4), 272-286. https://doi.org/10.1038/nrg3682

Singh, S. P., Dammeijer, F., Hendriks, R. W. (2018). Role of Bruton's tyrosine kinase in B cells and malignancies. *Mol Cancer, 17*(1), 57. https://doi.org/10.1186/s12943-018-0779-z

Slowikowski, K. (2020). ggrepel: Automatically Position Non-Overlapping Text Labels with 'ggplot2'. R package version 0.8.2.

Smirnov, A., Panatta, E., Lena, A., Castiglia, D., Di Daniele, N., Melino, G., Candi, E. (2016). FOXM1 regulates proliferation, senescence and oxidative stress in keratinocytes and cancer cells. *Aging (Albany NY), 8*(7), 1384-1397. https://doi.org/10.18632/aging.100988

Smith, A., Howell, D., Patmore, R., Jack, A., Roman, E. (2011). Incidence of haematological malignancy by sub-type: a report from the Haematological Malignancy Research Network. *Br J Cancer, 105*(11), 1684-1692. https://doi.org/10.1038/bjc.2011.450

Smith, J. P. & Sheffield, N. C. (2020). Analytical Approaches for ATAC-seq Data Analysis. *Curr Protoc Hum Genet, 106*(1), e101. https://doi.org/10.1002/cphg.101

Smith, Z. D. & Meissner, A. (2013). DNA methylation: roles in mammalian development. *Nat Rev Genet, 14*(3), 204-220. https://doi.org/10.1038/nrg3354

Song, L. & Crawford, G. E. (2010). DNase-seq: a high-resolution technique for mapping active gene regulatory elements across the genome from mammalian cells. *Cold Spring Harb Protoc, 2010*(2), pdb prot5384. https://doi.org/10.1101/pdb.prot5384

Spitz, F. & Furlong, E. M. (2012). Transcription factors: from enhancer binding to developmental control. *Nat Rev Genet, 13*(9), 613-626. https://doi.org/10.1038/nrg3207

Stevenson, F. K., Krysov, S., Davies, A. J., Steele, A. J., Packham, G. (2011). B-cell receptor signaling in chronic lymphocytic leukemia. *Blood, 118*(16), 4313-4320. https://doi.org/10.1182/blood-2011-06-338855

Stilgenbauer, S. (2015). Prognostic markers and standard management of chronic lymphocytic leukemia. *Hematology Am Soc Hematol Educ Program, 2015*, 368-377. https://doi.org/10.1182/asheducation-2015.1.368

Strahl, B. D. & Allis, C. D. (2000). The language of covalent histone modifications. *Nature, 403*(6765), 41-45. https://doi.org/10.1038/47412

Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W. M., 3rd, Hao, Y., Stoeckius, M., Smibert, P., Satija, R. (2019). Comprehensive Integration of Single-Cell Data. *Cell, 177*(7), 1888-1902 e1821. https://doi.org/10.1016/j.cell.2019.05.031

Tandon, N., Ramakrishnan, V., Kumar, S. K. (2016). Clinical use and applications of histone deacetylase inhibitors in multiple myeloma. *Clin Pharmacol, 8*, 35-44. https://doi.org/10.2147/CPAA.S94021

Thal, M. A., Carvalho, T. L., He, T., Kim, H. G., Gao, H., Hagman, J., Klug, C. A. (2009). Ebf1-mediated down-regulation of Id2 and Id3 is essential for specification of the B cell lineage. *Proc Natl Acad Sci U S A, 106*(2), 552-557. https://doi.org/10.1073/pnas.0802550106

The 1000 Genomes Project Consortium, Auton, A., Brooks, L. D., Durbin, R. M., Garrison, E. P., Kang, H. M., Korbel, J. O., Marchini, J. L., McCarthy, S., McVean, G. A., Abecasis, G. R. (2015). A global reference for human genetic variation. *Nature, 526*(7571), 68-74. https://doi.org/10.1038/nature15393

Tsompana, M. & Buck, M. J. (2014). Chromatin accessibility: a window into the genome. *Epigenetics Chromatin, 7*(1), 33. https://doi.org/10.1186/1756-8935-7-33

Turcan, S., Rohle, D., Goenka, A., Walsh, L. A., Fang, F., Yilmaz, E., Campos, C., Fabius, A. W., Lu, C., Ward, P. S., Thompson, C. B., Kaufman, A., Guryanova, O., Levine, R., Heguy, A., Viale, A., Morris, L. G., Huse, J. T., Mellinghoff, I. K., Chan, T. A. (2012). IDH1 mutation is sufficient to establish the glioma hypermethylator phenotype. *Nature, 483*(7390), 479-483. https://doi.org/10.1038/nature10866

Van Damme, M., Crompot, E., Meuleman, N., Mineur, P., Dessars, B., El Housni, H., Bron, D., Lagneaux, L., Stamatopoulos, B. (2014). Global histone deacetylase enzymatic activity is an independent prognostic marker associated with a shorter overall survival in chronic lymphocytic leukemia patients. *Epigenetics, 9*(10), 1374-1381. https://doi.org/10.4161/15592294.2014.969628

Versteege, I., Sevenet, N., Lange, J., Rousseau-Merck, M. F., Ambros, P., Handgretinger, R., Aurias, A., Delattre, O. (1998). Truncating mutations of hSNF5/INI1 in aggressive paediatric cancer. *Nature, 394*(6689), 203-206. https://doi.org/10.1038/28212

Ververis, K., Hiong, A., Karagiannis, T. C., Licciardi, P. V. (2013). Histone deacetylase inhibitors (HDACIs): multitargeted anticancer agents. *Biologics, 7*, 47-60. https://doi.org/10.2147/BTT.S29965

Visvader, J. E. (2011). Cells of origin in cancer. *Nature, 469*(7330), 314-322. https://doi.org/10.1038/nature09781

Waddington, C. H. (1942). The epigenotype. *Endeavour, 1*, 18-20. https://ci.nii.ac.jp/naid/10030653753/en/

Waddington, C. H. (1957). *The strategy of the genes; a discussion of some aspects of theoretical biology*. Allen & Unwin.

Wagner, E. J. & Carpenter, P. B. (2012). Understanding the language of Lys36 methylation at histone H3. *Nat Rev Mol Cell Biol, 13*(2), 115-126. https://doi.org/10.1038/nrm3274

Wahlfors, J., Hiltunen, H., Heinonen, K., Hämäläinen, E., Alhonen, L., Janne, J. (1992). Genomic Hypomethylation in Human Chronic Lymphocytic-Leukemia. *Blood, 80*(8), 2074-2080. https://doi.org/10.1182/blood.V80.8.2074.2074.

Walker, B. A., Boyle, E. M., Wardell, C. P., Murison, A., Begum, D. B., Dahir, N. M., Proszek, P. Z., Johnson, D. C., Kaiser, M. F., Melchor, L., Aronson, L. I., Scales, M., Pawlyn, C., Mirabella, F., Jones, J. R., Brioli, A., Mikulasova, A., Cairns, D. A., Gregory, W. M., Quartilho, A., Drayson, M. T., Russell, N., Cook, G., Jackson, G. H., Leleu, X., Davies, F. E., Morgan, G. J. (2015). Mutational Spectrum, Copy Number Changes, and Outcome: Results of a Sequencing Study of Patients With Newly Diagnosed Myeloma. *J Clin Oncol, 33*(33), 3911-3920. https://doi.org/10.1200/JCO.2014.59.1503

Wang, L., Wang, S., Li, W. (2012). RSeQC: quality control of RNA-seq experiments. *Bioinformatics, 28*(16), 2184-2185. https://doi.org/10.1093/bioinformatics/bts356

Warnes, G. R., Bolker, B., Bonebakker, L., Gentlema, R., Huber, W., Liaw, A., Lumley, T., Maechler, M., Magnusson, A., Moeller, S., Schwartz, M., Venables, B. (2020). gplots: Various R Programming Tools for Plotting Data. R package version 3.0.4.

Weirauch, M. T. & Hughes, T. R. (2011). A Catalogue of Eukaryotic Transcription Factor Types, Their Evolutionary Origin, and Species Distribution. In T. R. Hughes (Ed.), *A Handbook of Transcription Factors* (pp. 25-73). Springer Netherlands. https://doi.org/10.1007/978-90-481-9069-0_3

West, A. C. & Johnstone, R. W. (2014). New and emerging HDAC inhibitors for cancer treatment. *J Clin Invest, 124*(1), 30-39. https://doi.org/10.1172/JCI69738

Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis. . *Springer-Verlag New York*.

Wierzbinska, J. A., Toth, R., Ishaque, N., Rippe, K., Mallm, J. P., Klett, L. C., Mertens, D., Zenz, T., Hielscher, T., Seifert, M., Kuppers, R., Assenov, Y., Lutsik, P., Stilgenbauer, S., Roessner, P. M., Seiffert, M., Byrd, J., Oakes, C. C., Plass, C., Lipka, D. B. (2020). Methylome-based cell-of-origin modeling (Methyl-COOM) identifies aberrant expression of immune regulatory molecules in CLL. *Genome Med, 12*(1), 29. https://doi.org/10.1186/s13073-020-00724-7

Wu, C. & Morris, J. R. (2001). Genes, genetics, and epigenetics: a correspondence. *Science, 293*(5532), 1103-1105. https://doi.org/10.1126/science.293.5532.1103

Wu, H. & Sun, Y. E. (2006). Epigenetic regulation of stem cell differentiation. *Pediatr Res, 59*(4 Pt 2), 21R-25R. https://doi.org/10.1203/01.pdr.0000203565.76028.2a

Wunderlich, Z. & Mirny, L. A. (2009). Different gene regulation strategies revealed by analysis of binding motifs. *Trends Genet, 25*(10), 434-440. https://doi.org/10.1016/j.tig.2009.08.003

Xanthopoulos, C. & Kostareli, E. (2019). Advances in Epigenetics and Epigenomics in Chronic Lymphocytic Leukemia. *Current Genetic Medicine Reports, 7*(4), 214-226. https://doi.org/10.1007/s40142-019-00178-3

Xu, X., Xiong, X., Sun, Y. (2016). The role of ribosomal proteins in the regulation of cell proliferation, tumorigenesis, and genomic integrity. *Sci China Life Sci, 59*(7), 656-672. https://doi.org/10.1007/s11427-016-0018-0

Yan, F., Powell, D. R., Curtis, D. J., Wong, N. C. (2020). From reads to insight: a hitchhiker's guide to ATAC-seq data analysis. *Genome Biol, 21*(1), 22. https://doi.org/10.1186/s13059-020-1929-3

Yu, G., Wang, L. G., He, Q. Y. (2015). ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics, 31*(14), 2382-2383. https://doi.org/10.1093/bioinformatics/btv145

Yuniati, L., Scheijen, B., van der Meer, L. T., van Leeuwen, F. N. (2019). Tumor suppressors BTG1 and BTG2: Beyond growth control. *J Cell Physiol, 234*(5), 5379-5389. https://doi.org/10.1002/jcp.27407

Zenz, T., Mertens, D., Kuppers, R., Dohner, H., Stilgenbauer, S. (2010). From pathogenesis to treatment of chronic lymphocytic leukaemia. *Nat Rev Cancer, 10*(1), 37-50. https://doi.org/10.1038/nrc2764

Zhang, Y., Liu, T., Meyer, C. A., Eeckhoute, J., Johnson, D. S., Bernstein, B. E., Nusbaum, C., Myers, R. M., Brown, M., Li, W., Liu, X. S. (2008). Model-based analysis of ChIP-Seq (MACS). *Genome Biol, 9*(9), R137. https://doi.org/10.1186/gb-2008-9-9-r137

Zirkel, A., Nikolic, M., Sofiadis, K., Mallm, J. P., Brackley, C. A., Gothe, H., Drechsel, O., Becker, C., Altmuller, J., Josipovic, N., Georgomanolis, T., Brant, L., Franzen, J., Koker, M., Gusmao, E. G., Costa, I. G., Ullrich, R. T., Wagner, W., Roukos, V., Nurnberg, P., Marenduzzo, D., Rippe, K., Papantonis, A. (2018). HMGB2 Loss upon Senescence Entry Disrupts Genomic Organization and Induces CTCF Clustering across Cell Types. *Mol Cell, 70*(4), 730-744 e736. https://doi.org/10.1016/j.molcel.2018.03.030

Appendix

Suppl. table S 1. Quality metrics of H3K27ac ChIP libraries of MM cell lines.

	H3K27ac peaks	Normalized strand cross-correlation coefficient (NSC)	Relative strand cross-correlation coefficient (RSC)	FRiP-score
KMS-11 0 nM panobinostat	42507	1.09	1.00	0.12
KMS-11 8 nM panobinostat	73170	1.09	1.03	0.24
MM1.S 0 nM panobinostat	27664	1.11	1.04	0.08
MM1.S 8 nM panobinostat	57795	1.05	1.01	0.12
U266 0 nM panobinostat	53852	1.14	1.05	0.19
U266 8 nM panobinostat	70665	1.07	1.01	0.19

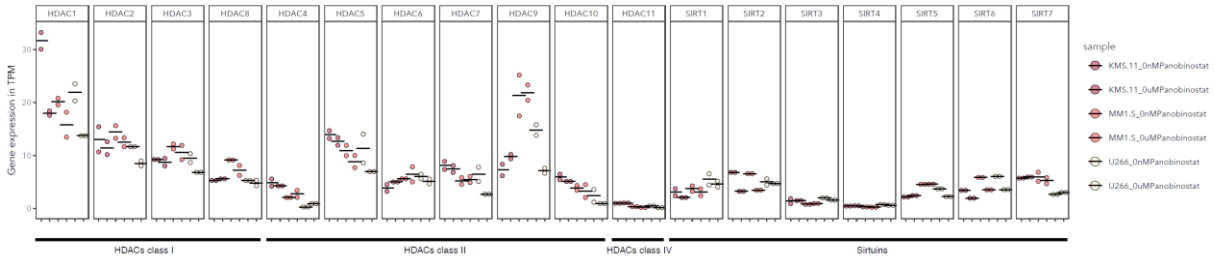
Suppl. table S 2. Quality metrics of targeted single cell DNA-seq of CLL patients.

Sample during ibrutinib treatment	# Of cells	Average reads per cell	Average reads per cell per amplicon	Panel uniformity	Read pairs	Valid barcode reads	% Reads Valid Cells
CLL24 START	5.503	53.021,39	185	97,55%	359.694.617	291.776.707	81,12
CLL24 REMISSION	1.981	159.428,71	557	98,95%	404.896.656	315.828.276	78,00
CLL24 RELAPSE	4.483	56.185,48	196	97,90%	319.850.456	251.879.491	78,75
CLL25 START	282	458.566,85	1.603	98,95%	148.395.236	129.315.853	87,14
CLL25 1 MONTH	6.583	34.884,23	122	97,20%	307.169.226	229.642.912	74,76
CLL25 7 MONTH	5.353	39.847,03	139	97,20%	274.874.939	213.301.161	77,60

Suppl. table S 3. Genes used for senescence score calculation from scRNA-seq.

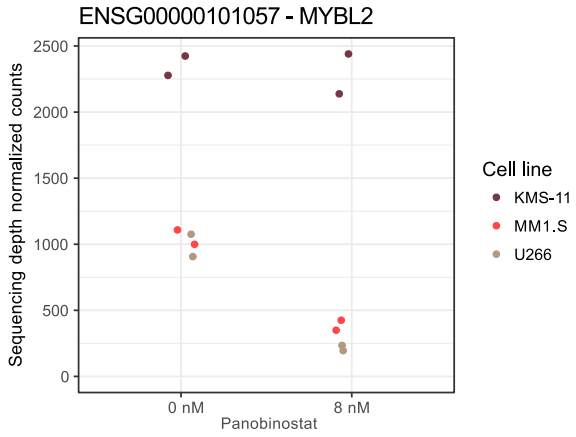
	Gene names
Early senescence expression score (upregulated genes)	*FHOD3, TMEM200A, ANGPTL4, CDKN2A, CCND2, PCDHB14, CHRNA1, FAT1, PIK3IP1, GAS6, ZMAT3, IGFBP7, LRP1, DUSP8, LZTS3, THSD4, MGARP, IL32, YPEL5, PRSS23, PDCD1LG2, MR1, FANK1, MALAT1, NCKAP5, IDUA, DCBLD2, KRT15, SERPINE2, PRRG1, SERPINE1, PPP1R3C, PLXNA3, ANGPT1, PLAT, MAN2B2, GABARAPL1, MAP1LC3A, QSOX1, CDKN1A, GDF15, MYEOV, , , FNDC3B, TMEM159, FAM214B, SAMD4A, TIMP2, DAB2, PALLD, DMD, PSAP, TOB1, EPB41L5, CERCAM, DST, FMN1, FILIP1L, TRPM4, USP53, PLOD2, RRAS, ITGAV, LOXL2, PVR, TBC1D2, MXRA7, PLA2G4C, ABLIM1, POFUT2, PRNP, ANO4, ADAM23, HGSNAT, SPARC, CCDC80, PLBD2, EVI5L, MAP1LC3B, DNASE1L1, P4HA2, HSD17B14, GPR137B, LRP10, CRYL1, NCSTN, MGAT5, CD44, CPQ, TNFAIP1, SORT1, IL17D, NEAT1, AHNAK2, MAN1B1, CLN8, SLC9A7, PPP2R5B, SVIL, BPGM, PLK3, SYNGR1, MFGE8, SLC12A4, SDC4, NFAT5, FBXL2, APP, CLIP4, TNFRSF10D, FADS3, CYP26B1, RRM2B, FHL1, PAM, IDS, EDN1, MANBA, ITGB5, PDE5A, STS, DCUN1D3, MAFK, HSD11B1L, BSDC1, OPTN, GADD45A*
Early senescence expression score (downregulated genes)	*EMP2, AUTS2, NEURL1B, PBK, BLM, , MND1, RAD51AP1, FBXO43, SNCAIP, CDKN3, CDCA7, E2F2, ZNF93, BUB1, PBX1, TMEM97, CDC25C, KNTC1, ESCO2, CDK1, PLK4, ATAD2, DEPDC1, TTK, NEK2, WDR76, NUSAP1, PSRC1, HMGB2, CDKN2C, LRP4, KIF24, STMN3, CCNA2, CENPI, PHGDH, ERCC6L, BUB1B, PRIM1, AMOT, NEIL3, MCM10, MXD3, BORA, CENPH, RBL1, FANCD2, ASF1B, GINS3, POLA1, TMPO, TICRR, SPC25, CCNB2, BARD1, MELK, DHFR, APOBEC3B, FIGNL1, TROAP, SPAG5, C1orf112, C21orf58, GMNN, FANCB, CCDC150, SPC24, KIF22, MTBP, DEPDC1B, DSCC1, EXO1, GINS1, MCM5, CENPU, ANP32E, KIF15, TPX2, CDC6, STIL, CCNE2, VRK1, CDCA2, MCM7, TOP2A, MCM6, EBP, CDC7, LMNB1, CENPM, MAD2L1, DLGAP5, FBXO5, POLQ, KIF20A, FANCA, NUF2, POLE2, NCAPG, OIP5, KIFC1, FANCI, NCAPH, HMMR, CENPA, CALCRL, AURKB, RFC4, MCM2, BRCA1, UBE2T, SKA1, NDC80, CDC45, ZWINT, DTL, METTL7A, SLC9A9, KIF11, POC1A, SHCBP1, SKA3, CHTF18, EXOSC8, SUV39H2, DNA2, TFAM, ZWILCH, E2F1, NCAPD2, LIN9, KIF4A, KIF2C, CCNB1, MRPL3, TIPIN, CSE1L, NDC1, EZH2, TMSB15A, TRAIP, GTSE1, LBR, CCNF, RACGAP1, OAS3, SHMT1, CENPN, CRISPLD1, KIF23, HMGB1, ORC1, UBE2C, WDR62, GINS2, MCM4, RFC3, MTFR2, RAD54L, MNS1, KIF18B, E2F8, FOXM1, PLK1, PARPBP, FAM83D, CDCA5, HELLS, NCAPD3, AUNIP, HAT1, CDCA8, CENPF, CDCA3, NUP107, SFXN2, GINS4, CDC20, PKMYT1, RNF138, MKI67, RAD51, PCNA, KIF14, CLSPN, CCDC77, CMSS1, LRRCC1, CENPW, MSH2*

Suppl. figure S 1 Gene expression of HDACs in untreated MM cell lines.



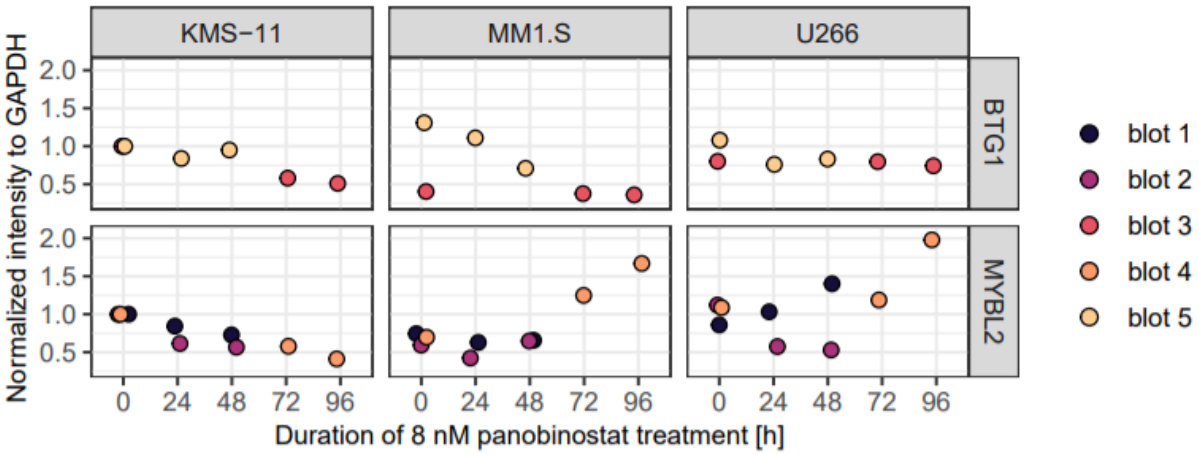
Suppl. figure S 1 Gene expression of HDACs in MM cell lines.
Dotplots show gene expression levels (TPM) of all HDACs in the MM cell lines KMS-11, MM1.S and U266. RNA-seq experiments were conducted twice with technical replicates of library preparation.

Suppl. figure S 2 Gene expression of MYBL2 in MM cell lines.



Suppl. figure S 2 Gene expression of MYBL2 in panobinostat treated MM cell lines.
Dotplot shows gene expression levels of *MYBL2* in the MM cell lines KMS-11, MM1.S and U266, which were treated with 8 nM panobinostat or DMSO as control for 24 hours before bulk RNA-seq was conducted.

Suppl. figure S 3 Protein abundance of BTG1 and MYBL2 in panobinostat treated MM cell lines.



Suppl. figure S 3 Quantification of protein abundance of BTG1 and MYBL2 in MM cell lines under panobinostat treatment by immunoblotting.
Normalized intensities were calculated as ratio to untreated KMS-11 cells of the same immunoblot.

Acknowledgements

I would like to take this opportunity to thank Prof. Dr. Karsten Rippe for being my “Doktorvater” and giving me the opportunity to conduct my thesis in his group. Thank you, Karsten, for your constant support and advice, your enthusiasm for all details of the project, for giving me the opportunity to develop new skills and all the discussions and time you spent. I am very grateful that I could conduct my thesis in your group.

Further, I would like to thank Prof. Dr. Benedikt Brors for helpful advice and suggestions during all my TAC meetings and for being my second referee and chairing my PhD defense.

Thanks to Dr. Julien Bethune and Dr. Simon Raffel for your willingness to be part of my TAC committee and fruitful discussions.

I wish to thank Prof. Dr. Ursula Kummer and Dr. Sevin Turcan for being referees of my PhD defense committee.

Thanks to the Genomics & Proteomics Core Facility and the Omics IT and Data Management Core Facility of the DKFZ for conducting the sequencing and handling of the data storage as well as all collaboration partners for the good cooperation in particular Prof. Dr. Daniel Mertens, Dr. Sabrina Kugler, Prof. Dr. Stephan Stilgenbauer, Dr. Deyan Yosifov, Dr. Martina Seiffert, Laura Lao Cid, Dr. Murat Iskar, Dr. Naveed Ishaque, Dr. Marc Zapatka, Prof. Dr. Marc Raab, Dr. Marc Bärtsch, Prof. Dr. Christoph Plass, Dr. Justyna Wierzbinska, Dr. Roma Kurilov, Dr. Charles Imbusch and Prof Dr. Alwin Krämer.

I thank the German Federal Ministry of Education and Research (BMBF) and the Helmholtz International Graduate School for the funding.

I would also like to thank Dr. William Mueller for proof-reading and for opening doors for me.

I wish to address special thanks to Sabrina Schumacher. Liebe Sabrina, ich kann dir nicht genug danken für all deine Hilfe und Unterstützung, immer verlässlich und flink. Danke für deine Hilfsbereitschaft und die tolle Zusammenarbeit. Außerdem möchte ich Caroline Bauer für ihre Hilfe bei allen kleineren und größeren Dingen im Labor danken. I would like to thank Dr. Jan-Philipp Mallm for sharing all his expertise, for patiently answering all my questions and for enabling me to be part of exciting developments. I further thank Dr. Nikolaus Kepper and Simon Steiger for the support with

computational and cluster-related issues of all kind. Thanks for all the time you spent to keep everything running. Thanks also to Sofie for the help with all organizational matters.

I especially thank Dr. Markus Muckenhuber for being my “coding buddy” and companion during the PhD. Thanks for discussing all small and big issues and your constant help, which made this a lot easier.

To all current and former members of the Rippe group, who accompanied me on this journey: Thank you for making this a wonderful time both in- and outside of the lab. It is or was a great pleasure to work with you in such a positive and friendly atmosphere. Thank you so much for all your support, your feedback, your proof-reading and never-ending readiness to help. Thanks also for all the coffee corner meetings, mensa lunch breaks, ice cream sessions, birthday cakes, costume parties, cocktail evenings, Christmas parties, … . Thank you, Dr. Alexandra Poos, Dr. Delia Braun, Jorge Trojanowski, Dr. Stephan Tirier, Dr. Fabian Erdel, Dr. Anne Rademacher, Lukas Frank, Dr. Verena Beier, Katharina Bauer, Armin Hadzic, Isabelle Lander, Robin Weinmann, Linda Schuster, Norbert Mücke, Lisa Dreßler, Dr. Inn Chung, Dr. Katharina Deeg, Dr. Teresa Pankert and Dr. Michelle Liberio.

My special thanks go to Anting Grams for his emotional support and care. Thanks for cheering me up with all these small things especially over the past few months. Thank you for making my life outside the lab so wonderful and exciting. I am very happy to have you by my side.

Last and most importantly I would like to express my deepest gratitude to my family to whom I dedicate this thesis. Without you this would never have been possible. I am eternally grateful to my parents Dr. Regine Peschina-Klett and Dr. Alfred Klett for their precious advice and loving support on so many levels during my whole life. I am thankful that you encouraged my curiosity especially about natural sciences from an early age, paved my way and gave me the confidence that I can make it to this point. Thank you for your everlasting encouragement and for always believing in me, especially in moments when I did not. Just as much support is my beloved little sister and at the same time best friend Sophie Klett. Thank you for your lifelong companionship and for always being there for me. I do not know what I would do without you. I am grateful for all the countless moments we shared and will share. Thank you for being the best sister anyone could imagine.