Uncovering TET2 chromatin targets at the leukemic onset

Aleksey Lazarenkov

UNIVERSITAT DE BARCELONA

Facultat de Medicina i Ciències de la Salut

Programa de doctorat en Biomedicina

'Uncovering TET2 chromatin targets at the leukemic onset'

Aleksey Lazarenkov

Supervisors de la tesi:

Dr. José Luis Sardina Ortega

Dr. Esteban Ballestar Tarín

Tutor de la tesi:

Dr. Manel Esteller Badosa

Institut de Recerca Contra la Leucèmia Josep Carreras

Table of contents |

List of abbreviations |

5caC: 5-carboxylcytosine

5fC: 5-formylcytosine

5hmC: 5-hydroxymethylcytosine

5mC: 5-methylcytosine

A2: *AGO2*

AML: acute myeloid leukemia

ADD: ATRX–DNMT3–DNMT3L

ALT: alternative splice donors/acceptors

AS: alternative splicing

ATAC: Assay for Transposase-Accessible Chromatin

BAF: Brahma-associated factor

B-ALL: B-cell acute lymphoblastic leukemia

bp: base pair

BP: biological process

cCREs: cis-regulatory elements

CGI: CpG islands

CHD: chromodomain helicase DNA-binding

CHIP: clonal hematopoiesis of indeterminate potential

ChIP-seq: chromatin immunoprecipitation sequencing

CLL: chronic lymphocytic leukemia

CLP: common lymphoid progenitors

CMML: chronic myelomonocytic leukemia

CMP: common myeloid progenitors

Co-IP: co-immunoprecipitation

CpG: CpG site

CRISPR: clustered regularly interspaced short palindromic repeats

DC: dendritic cells

DEG: differential expressed gene

DLBCL: Diffuse large B-cell lymphoma

DMP: differential methylated position

DNAm: DNA methylation

DNMT: DNA methyltransferase

DSG: disuccinimidyl glutarate

E: enhancer

E2: β-estradiol

EPIC: Infinium MethylationEPIC array

eRNA: enhancer-derived RNA

ESC: embryonic stem cells

EX: exon skipping

FA: formaldehyde

FDR: false discovery rate

GMP: granulocyte-monocyte progenitor

GO: gene ontology

GP: granulocyte progenitors

GRE: gene regulatory element

HAT: histone acetylases

HDAC: histone deacetylases

Hi-C: high throughput chromatin conformation capture technique

Hpi: hours post induction

HR: hazard ratio

HSC: hematopoietic stem cells

Hypo-AML: hypomethylated AML patients

Hyper-AML: hypermethylated AML patients

ICR: imprinting control regions

IDR: Irreproducibility Discovery Rate

INO80: inositol requiring 80

IR: intron retention

ISWI: imitation switch

iMac: induce macrophage

IP-MS: immunoprecipitation coupled with mass-spectrometry

iPSC: induced pluripotent stem cells

LAML: TCGA acute myeloid leukemia

MBD: methyl-CpG-binding domain protein

M-CSF: macrophage colony-stimulating factor

MDR: Methylation Dynamic Region

MDS: myelodysplastic syndromes

MEP: megakaryocyte-erythrocyte progenitors

miRNA: microRNA

MK: megakaryocytes

MLP: multi-lymphoid progenitors

MPN: myeloproliferative neoplasms

MPP: multipotent progenitors

MSR: Methylation Steady Regions

MTase: methyltransferase

mtDNA: mitochondrial DNA

NK: natural killer

ncRNA: Non-coding RNA

OS: oxidative stress

P: promoter

PCA: principal component analysis

PGCs: primordial germ cells

Pol II: polymerase II

PRC: polycomb repressive complex

PTM: post translational modifications

PWWP: Pro-Trp-Trp-Pro

RISK: RNA-induced silencing complex

RNA-seq: RNA sequencing

rRNA: ribosomal RNA

RT-qPCR: Reverse transcription polymerase chain reaction

SAM: synergistic activation mediators

shRNA: small hairpin RNAs

siRNA: short interfering RNA

T2: *TET2*

TCGA: The Cancer Genome Atlas Program

TET: ten-eleven translocation

TD: transdifferentiation

TDG: thymine DNA Glycosylase

TF: transcription factor

tRF: tRNA-derived small noncoding RNA fragments

tRNA: transfer RNA

TSS: transcription start sites

UMAP: uniform manifold approximation and projection

UTR: untranslated region

WGBS-seq: whole-genome bisulfite sequencing

XIST: X-inactive specific transcript

Abstract

Epigenetic modifications, including DNA methylation, have been shown to play a prominent role in influencing cell identity in hematopoiesis. However, its role in gene regulation during differentiation remains poorly understood and yet to be fully elucidated. Among the epigenetic regulators, the TET family of proteins is directly responsible for the oxidation of 5mC residues to 5hmC, an essential mechanism for the correct cell fate decisions during both normal and malignant development.

A clinical overview of myeloid malignancies shows that somatic mutations in TET2 are frequently found in a wide cohort of patients, ranging from 10-50%. Several studies have shown that HSCs with mutated TET2 display an aberrant proliferation rate that outcompetes normal HSCs. However, the specific downstream events responsible for this expansion are currently unknown. Additionally, traditional studies focused on defining a gene-centric methylation profile, where modifications on gene bodies and proximal promoters were the main targets. However, it has been recently highlighted that distant gene regulatory elements such as enhancers represent the most highly dynamic methylation regions during cell fate conversion, with TET2 preferentially binding to these sites. Accordingly, examining chromatin activity at those regions and its impact on the cell's transcriptome during blood cancer onset is a compelling enterprise to tackle.

In this study, we extensively profiled TET2 genome occupancy in a highly controllable, rapid, and uniform cellular model of myeloid commitment. We crossed our data with publicly available datasets containing information about chromatin accessibility (by ATAC-seq), configuration (by HiC-seq), and state (by TT-seq and ChIP-seq for histone marks) during myeloid establishment. As a result, we discovered the role of TET2 in activating cell fate commitment programs. We identified subsets of TET2-bound regulatory regions that get demethylated and activated upon myeloid commitment, as well as uncovered novel TET2 implications in long-range chromatin remodeling. Furthermore, we profiled the DNA methylation and expression events affected by TET2 loss of function, and along with the TET2 chromatin occupancy data, we used them to identify TET2 bona fide chromatin targets during myeloid establishment. To gain

insight into the potential involvement of these novel TET2 targets in a leukemic context, we then crossed our data with publicly available methylation data in TET2-mutated AML patients (LAML-TCGA). We observed that 13 genes were both detected as TET2 targets in our system and hypermethylated in AML patients, highlighting their potential involvement in the molecular mechanisms underlying TET2 mutations. Among them, we identified AGO2, which enhancer gets demethylated by TET2 during myeloid commitment, ultimately resulting in transcriptional modulation through direct enhancer-promoter contact.

Altogether, our unbiased chromatin target identification offers a unique opportunity to investigate the relevance of DNA methylation events during myeloid differentiation. We have shown the importance of TET2 in targeting gene regulatory elements of crucial myeloid genes that get abnormally hypermethylated and downregulated upon TET2 loss of function, as seen in TET2-mutated AML patients. This underscores the need to investigate our bona fide TET2 targets, such as AGO2, for potential new therapeutic approaches to address myeloid malignancies in patients.

Chapter 1 |

Introduction

Biologists have long sought to understand how the coordinated action of hundreds of proteins on DNA influences healthy and malignant development. The idea that cells containing the same genetic information had the ability to develop and specialize in completely different ways, hinted at the idea of an "epi" (from Greek ἐπι-: over, upon) regulatory layer in our genome. It was not until 1942 that Conrad Hal Waddington coined the term epi-genetics/epi-phenotype in his Drosophila melanogaster wing study, referring to differences in the phenotype that occurred without changes in the genotype.

Epigenetic studies have drastically evolved, but the initial concept is still maintained. Nowadays, the phenotype is tackled through a multitude of high-throughput omics techniques that interrogate different levels of chromatin modulation. Within the field, several layers of epigenetic regulation have been identified and profiled in various biological systems. These include chromatin remodelers, histone modifications, integral components of 3D chromatin architecture, and accessibility. In addition, DNA modifications, such as cytosine methylation, have been extensively linked to gene regulatory networks. More recently, many epigenetic RNA modifications and non-coding RNAs have also been shown to be crucial for chromatin remodeling and gene activation. Altogether, they constitute a vast regulatory network for essential cell biological processes, including replication, transcriptional regulation or differentiation.

Epigenetic dysregulation is often linked to a great variety of human diseases, including cancer, autoimmune disorders, and neurological pathologies such as Fragile X syndrome or Parkinson's disease. Many of these diseases are associated with aged individuals since epigenetic modifications are frequently altered in the aging process. Collectively, the epigenetic changes associated with cancers and aging constitute a very attractive field of research and have streamlined rapid small molecule-based therapies and reprogramming strategies advancements, which have been further boosted with rapidly evolving epigenome editing tools.

To sum up, the coordinated action of different epigenetic layers, including RNA modifications, DNA methylation, histone marks, 3D chromatin organization, and non-coding elements, are essential for proper mammalian biology, disease development, and aging.

The main work of this thesis has focused on the influence of DNA methylation in hematopoiesis, specifically on myeloid commitment acquisition, and how alterations in DNA methylation might influence the onset and development of myeloid malignancies. DNA methylation has been previously shown to play a prominent role in influencing cell identity in hematopoiesis, as well as directly connected to age-related events such as clonal hematopoiesis of indeterminate potential (CHIP). However, its role in gene regulation during differentiation remains poorly understood. Hence, the study of epigenetic regulators and their contributions to aberrant chromatin structure and cellular transcriptome in the onset and development of hematological malignancies is compelling from both basic and translational perspectives.

§

*In the introduction of this thesis, I aim to provide a comprehensive overview of the various layers of the mammalian epigenome and their respective roles in development and disease (see **Fig. I1**). This is essential not only for defining these mechanisms but also for better understanding the key findings of this work. Our primary objective from the outset was to elucidate the molecular mechanisms underlying DNA methylation within a robust biological framework that takes into account as many regulatory networks as possible. Therefore, it is crucial to thoroughly address these mechanisms, as many of them were utilized in defining an epi-genome map that was rigorously analyzed to support our findings.*

§

(I) Epigenetic modifiers: overview and biological functions

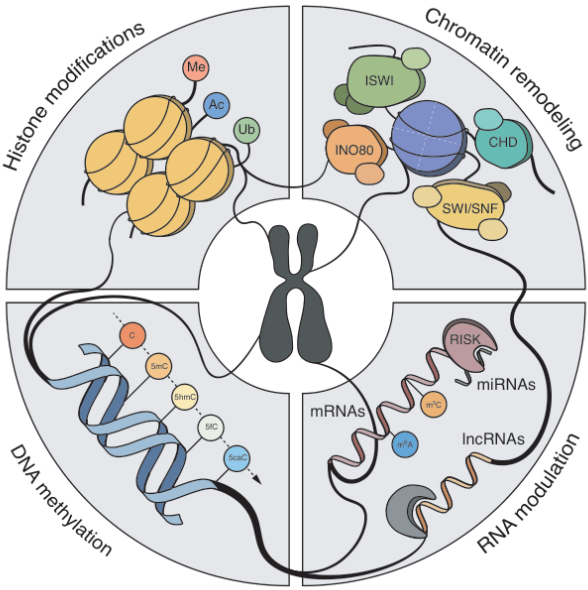

Figure I1. Graphical overview of the most relevant epigenetic layers in mammals

1.1 Histone modifications

Histone proteins are fundamental units responsible for DNA packaging, which forms a wildely known macromolecule structure called chromatin. This process allows DNA storage within a small nuclear volume and, more importantly, maintains a highly complex and coordinated spatio-temporal DNA accessibility.

One of the main histone regulating mechanisms, is the histone decoration by a great variety of post-translational modifications (PTMs). These are maintained by a wide array of writers (acetyltransferases, methyltransferases, kinases, and ubiquitinases), readers (bromodomains, tudor domains, PHD fingers, WD40 domains, CxxC domains, ankyrin repeats) and erasers (deacetylases,

phosphatases, demethylases, deubiquitinases) which dynamically adapt to the requirements of the cell by precisely modulating the PTM-histone network.

Structurally, transcriptional activation and repression are tightly linked to histone tail PTMs. Within this extensive list, the most common ones are lysine (K) and arginine (A) acetylation and methylation at histones 3 and 4. These modifications modulate chromatin compaction and access for transcriptional machinery, where the deposition of one mark or another can lead to transcriptional activation or repression.

This is tightly related to the three-dimensional chromatin organization within the cell nucleus, where two functionally distinct territories are defined: euchromatin and heterochromatin. The euchromatin is formed by low condensed and gene-rich regions with high transcriptional rates, which are marked with histone acetylation and histone H3 lysine 4 methylation (H3K4me) active marks. For instance, the euchromatin typical H3K4me3 decoration has been extensively linked to gene expression. In mammals, it is enriched at transcriptional starting sites (TSS) of well-expressed genes, most likely facilitating the recruitment of transcription factors (TFs) and/or reinforcing transcriptional consistency in a context-dependent manner [1–3]. Similarly, the acetylation of histone H3 on lysine 27 (H3K27ac) is one of the most prominent marks associated with enhancers [4]. Meanwhile, the heterochromatin is a highly condensed territory characterized by low transcription and repressive marks such as histone H3 lysine 9 methylation (H3K9me) mark and hypoacetylation.

Importantly, histone core modifications also play a role in chromatin regulation. For instance, though DNA accessibility modulation by dynamic unwrapping of nucleosomes [5]. PTMs located within the globular domain, such as H3K56ac or H3K64ac, can increase the rate of local DNA unwrapping or destabilize nucleosomes, respectively [6–8].

1.1.1 Histone modifications in development and cancer

Unsurprisingly, many of the histone modifications have been associated with almost all stages of development [9] and are altered in a wide array of cancers [10].

- In development.

Waves of histone modifications occur during primordial germ cells (PGCs) and embryonic development. Early germ cell reprogramming is characterized by H3K9me2 loss and H3K27me3, H3K4me2 and H3K4me3 marks gain. This landscape is shifted around in more advanced stages by the loss of active H3K9ac and repressive H3K9me3 and H3K27me3 marks, among others [11,12]. After fertilization the embryos rely on maternal vs paternal histone marks establishment for proper development. The active H3K4me3 and other marks in the paternal genome get sequentially removed and/or reestablished during pre-implantation stages. Similarly, in the maternal genome, a selective depletion of repressive H3K27me3 also occurs [13–15].

It is common to have histones bivalently marked with both activating H3K4me3 and repressive H3K27me3 marks at promoters and enhancers. This double signature is associated with a poised state, where, although repressed, it can be quickly activated for cell differentiation during development. These can be found in early embryos at promoters of lowly expressed developmental gene families such as SOX, PAX, and POU [16,17].

- In cancer

One of the best examples of the role of chromatin modifiers in cancer is the MLL family of lysine-methyltransferases (KMTs), which are often rearranged in blood malignancies. MLL1 frequently loses its function by translocation in acute myeloid and lymphoid leukemias and drives leukemogenesis through deregulation of Hox genes, among others [18]. Similarly, other histone methyltransferases such as SET1A have been associated with breast cancer metastasis and lung colorectal cancer tumorigenesis through abnormal methylation of TF2 such as YAP1 [19]. Regarding acetylation decoration, histone acetyltransferases (HATs) and deacetylases (HDACs) have been found to correlate with disease cancer progression and survival through epigenetic inhibition of well-known tumor suppressor genes such as *CDKN1A*, *BRCA1*, or ATR. Specifically, downregulation of *HDAC1* has been linked to breast, colon, and lung cancers, in addition to acute promyelocytic leukemia (APL) [20–22].

Based on this knowledge, drugs targeting histone modifier enzymes have been designed and are currently approved for treating hematological or solid cancers.

1.2 Chromatin remodeling

Another layer of epigenetic regulation involves the activity of chromatin remodeling complexes, which can modulate gene expression by exposing binding sites for TFs during differentiation. The best-known complexes are the ATPase domain SWI/SNF (also known as Brahma-associated factor -BAF- in mammals), the chromodomain helicase DNA-binding (CHD), the imitation switch (ISWI) and inositol requiring 80 (INO80). These complexes are crucial for chromatin structure dissociation and assembly through nucleosome eviction and histone exchange.

1.2.1 SWI/SNF | BAF

The BAF complex is the best-known chromatin remodeler and has been described to be crucial for the proper development, functioning as a gene regulator and interacting with a variety of TFs in a tissue-specific manner [23–25].

In embryonic stem cells, specific BAF complexes (esBAF and ncBAF) depend on BRG1 ATPase subunit activity. Their integration and cooperative function with Polycomb repressive complex 2 (PRC2), regulates self-renewal and pluripotency in development. For instance, depletion of Brg1 is lethal in pre-implantation stages and leads to the downregulation of pluripotency genes *Oct4*, *Sox2* and *Nanog*, most likely through promoter regulation [26,27].

1.2.2. ISWI

The ISWI family can be found with the SMARCA family of conserved ATPases in distinct complexes involved in a wide range of cell physiology processes, such as transcriptional regulation, DNA damage response/repair, or DNA replication through heterochromatin and chromosome segregation.

In chromatin maintenance, ISWI complexes have been extensively described to participate in nucleosome maturation and spacing, as well as in enhancing nucleosome sliding during DNA replication and transcription [28–30]. Additionally, the ISWI complex modulates global de-condensation of mitotic chromosomes and impacts interacting genomic region by reducing CTCF binding [31].

1.2.3 CHD

The CHD family is an extensive group of complexes containing distinct chromodomain-containing members. Structurally, they share a similar ATPase domain with the previously described SWI/SNF2 family and function by recognizing histone modifications. This allows chromatin disruption by translocating the nucleosome. The function of these complexes in stem cells has been proven to be essential for cell survival, maintenance, and proliferation [32,33].

For instance, CHD1 specifically recognizes and binds to active H3K4me3, allowing for post-transcriptional initiation factors and transcriptional activation recruitment in ESCs maintenance. The activity of the mediator complex, a regulator of ESCs, is necessary to recruit CHD1 to its target genes to pre-initiate transcription and ensure pre-mRNA maturation [34]. It maintains an open chromatin state in pre-implantation embryos, allowing for key regulators of cell fate specification, such as Oct4, Nanog, and Cdx2, to be expressed through the Hmgpi pathway. In addition, CHD1 also participates in hematopoietic progenitors' development by mediating endothelial-to-hematopoietic transition [35,36].

1.2.4 INO80

INO80 complexes depend on p400 and SRCAP mammal ATPases. Once again, their nucleosome remodeling abilities have been reported to contribute to transcriptional regulation, DNA replication, and DNA damage repair for proper embryonic development and cell fate commitment [37,38]. Additionally, INO80 is also required for exchanging the histone variant H2A.Z with canonical H2A, which enhances chromatin mobility at promoter regions during ESC differentiation [39–41].

Other findings also link INO80 to specific developmental phases. In the early stages of meiotic prophase in germ cell maturation, the complex regulates double-strand break repair during homologous chromosome recombination [42]. Similarly, to CHDs, the INO80 complex also regulates stem cell pluripotency establishment and maintenance by facilitating Mediator and elongation machinery recruitment to promoter regions of *Oct4*, *Nanog*, *Sox2*, and *Klf4* genes [43].

Overall, chromatin remodelers constitute a crucial regulatory layer of nucleosome dynamics during development and are extensively altered in cancer patients. Actually, around 20% of malignancies present some kind of alterations. For instance, the SWI/SNF subunit ARID1A or PBRM1 is prominently mutated in ovarian cancer and renal carcinomas, respectively, which had been moderately responsive to immune checkpoint blockade therapies [44]. Similarly, inhibiting Chd4 in breast cancer has promising results by arresting cell cycle progression [45].

1.4 Non-coding RNAs

Non-coding RNAs (ncRNAs) are untranslated RNA molecules that modulate gene expression. These include a wide variety of subtypes such as microRNAs (miRNAs), short interfering RNAs (siRNAs), long non-coding RNAs (lncRNAs), as well as others such as small nuclear RNAs (snRNAs), small nucleolar RNAs (snoRNAs), ribosomal RNAs (rRNAs), transfer RNAs (tRNAs), circular RNAs (cRNAs) and piwi-interacting RNAs (piRNAs), which vary on their location and targets.

Focusing on miRNAs and lncRNAs, they have been described to influence gene expression through their interaction with epigenetic modifiers and TFs by several mechanisms described in the following sections.

1.3.1 miRNA

This class of double-stranded RNAs is mostly responsible for mRNA silencing. The main mechanism occurs through the RNA-induced silencing complex (RISC) where, after miRNA processing through the DICER pathway, the

molecule is directed towards the target mRNA (usually 3'UTR regions), ultimately resulting in translation inhibition or direct mRNA degradation. These mechanisms have been extensively described as involved in basic biological processes such as cell proliferation, adhesion, cell death and differentiation [46]. In development, miRNAs regulate gastrulation, neural development and hematopoiesis, among other processes. For instance, miR-302 targets *Lefty1* and *Lefty2* genes, which antagonize the TGF-β signaling pathway responsible for promoting mesendoderm development and suppressing neuroectoderm formation [47]. While miR-181, mainly found in B cells, promotes their specific hematopoietic commitment [48].

A great number of studies have described up/down-regulation of miRNA in solid tumors, which is particularly prevalent in patients with colorectal cancer and glioblastoma multiforme. Several miRNAs (e.g., miR-21, miR-30a, miR-34a, and miR-145) are capable of downregulating tumor suppressors such as PDC4 or ITGB3, which promote cancer metastasis and invasion [49]. Similarly, in ALL miR-125b was been associated to chemoresistance and prognosis, as by repressing ETS1 oncogene [50].

1.3.2 LncRNAs

This class of ncRNAs is transcribed from a wide array of coding or non-coding sequences, which product results in small RNA molecules with specific three-dimensional conformation, allowing them to participate in a variety of biological functions. Although mainly responsible for gene silencing, lncRNAs also participate in 3D chromosome organization, protein sequestration, and scaffolding processes. More specifically, lncRNAs participate in X-chromosome inactivation, imprinting, and the overall remodeling of the chromatin landscape. Similar to miRNA, their interactions with epigenetic modifiers have been shown to be crucial in development and cancer.

- In development

One of the most well-known lncRNAs is XIST, which is responsible for X chromosome inactivation, serving as a protein binding scaffold. During imprinting XIST, and other members, are able to recruit epigenetic regulators

such as HDACs and PRC1/PRC2 complexes resulting in strong gene silencing [51].

- In cancer

lncRNAs have been linked to oncogenesis in solid tumors such as gastrointestinal, bladder, ovarian and colorectal cancers, among many others [52]. For instance, HOTAIR lnRNA alters the repression of a well-known oncogene human epithelial growth factor receptor 2 (HER2) through blocking of miR-331-3p which is responsible of maintaining low HER2 levels in breast cancers [53,54]. Similarly in AML, HOTAIR is upregulated and regulates the c-KIT proto-oncogene by competing with miR-193a [55].

1.5 RNA modifications

RNA modifications are naturally occurring chemical modifications affecting all coding (mRNA) and non-coding RNA types such as ribosomal RNA (rRNA) and transfer RNA (tRNA) (Gebert Nat. Rev. Mol. Cell Biol.2019). More than 100 different modifications have been described, highlighting the importance and complexity of this epigenetic layer. These modifications occur in all four nucleosides. The most common ones are methylation, pseudouridylation, and adenosine-to-inosine (A-to-I) editing [56].

The deposition of these marks occurs during or after transcription, directly influencing the process or during sequential translation, export, and/or RNA processing [57]. The network is finely balanced through RNA writers, readers, and erasers, which, during response to stimuli, can rapidly destabilize and degrade the target RNA pool, affecting cell proteome before transcriptional changes occur. These changes have been described to occur under acute oxidative stress (OS) response, DNA damage, cancer treatments, and during development. An efficient response to modify differentiation, migration or cell cycle is necessary in many cases of cell fate commitment, which are usually associated with microenvironment changes such as occurring in the stem cell niche [58,59]. Unsurprisingly, the same adaptability is hijacked by many cancer cells, which are well known to tolerate a lack of nutrients, oxidative stress, and anti-cancer drugs [60].

The following sub-section will uniquely focus on RNA modifications at mRNAs to provide a concise overview.

1.4.1 mRNAs modifications: role of m6A

mRNAs present a wide variety of modifications, such as base methylation (m6A, m1A, and m5C), ribose sugar methylation (Nm, m6Am), base isomerization to produce pseudouridine (Ψ) and oxidation of m5C to 5-hydroxymethylcytosine (hm5C), as well as classical 5′ cap and 3′ polyadenylation [56,61].

One of the best-known mRNA modifications is m6A, a widespread mark occurring in up to 40% of mammalian coding and non-coding transcripts [62]. m6A is established by the writer protein complex METTL3 and erased mainly by two demethylases, FTO and ALKBH5 [63,64]. Once established, an array of readers, such as the YTHDF or HNRNP family of proteins, are recruited directly or indirectly to modulate the homeostasis of the target transcript [65].

- In development

Several studies highlighted the importance of m6A-mediated transcriptional regulation during embryonic and adult stem cell differentiation. In mice, depletion of the bona fide m6A writer Mettl3 is embryonic lethal, due to an impaired exit from pluripotency, which was related to stable levels of the pluripotency factors Nanog, Smad2 and Smad3 [66,67]. Similarly, in other cell types depletion of Mettl3 leads to other differentiation abnormalities. For instance, it limits naïve T cell proliferation and differentiation by stabilizing the Socs family of proteins [68] or delaying lineage specification neurogenesis in the case of Mettl14 depletion [69]. Furthermore, knock-outs of readers such as Ythdf2, although non-embryonic lethal, have been reported to alter neural stem and progenitor cell differentiation and proliferation, as well as inducing an impairment of the endothelial-to-hematopoietic transition by stabilizing *Notch1a* and *Rhoca* genes [69,70].

- In cancer

In oncogenesis, METTL3 has also been linked to both solid and blood cancers. For instance, recent reports indicate upregulation in colorectal cancers where it enhances the p-STAT3 pathway through JAK1/STAT3 signaling, resulting in increased cancer cell proliferation and metastasis [71]. In AML, high METTL3 levels promote the translation of MYC, BCL2, and PTEN transcripts, leading to enhanced myeloid cell survival and proliferation [72]. In many cases though known TFs recruitment such as CEBPZ [73]. Additionally, YTHDF2 depletion appears to stabilize tumor necrosis factor receptor TNFR2 and other pro-inflammatory genes, which leads to TNF-mediated cell death in leukemic cells [74,75].

§

Being DNA methylation the main focus of this thesis, I would like to provide a dedicated whole section to this epigenetic mark. This section will dive deep into the specifics of this chromatin layer and introduce the essential concepts to understand the results of this work.

§

(II) DNA methylation: biological functions and genome distribution

DNA methylation is the best-known and studied epigenetic mark. Briefly, it consists of the covalent addition of methyl groups (CH3) to the 5' position of cytosine bases, which are mostly found in the context of symmetrical CpG dinucleotides. It is a prevalent mammalian genome mark, with around 75% of total CpGs methylated. Which has been described to greatly influence numerous developmental and cellular commitment processes through methylation-mediated transcriptional modulation [76,77].

2.1 Canonical vs non-canonical DNA methylation

DNA methylation deposition occurs preferentially, although not exclusively, at CpG dinucleotide sites. Methylation of cytosine at non-CpG dinucleotides is much more frequent in plants but has also been reported in mammalian genomes, although its significance has not been fully uncovered [78].

2.1.1 Non-CpG methylation

While most DNA methylation occurs at CpG sites, advances in detection techniques like whole genome bisulfite sequencing (WGBS-seq) have enabled the identification and differentiation of non-CpG methylation from the canonical CpG methylation. The non-CpG methylation occurs in a cell and tissue-specific manner, being most commonly observed in ESCs [79]. It is mainly regulated by DNA methyltransferases (DNMTs), together with additional factors such as Enhancer of zeste homolog 2 (EZH2) and methyl-CpG binding protein 2 (MeCP2) [80].

As for the CpG methylation, recent studies suggest that an imbalance of non-CpG methylation at regulatory regions might also lead to expression changes and, ultimately, developmental issues and disease progression [81]. Although cell type-specific, the non-CpG methylation patterns are non-random and, hence, probably carry biological significance. Additionally, several examples illustrate how non-CpG methylation on gene promoters influences the recruitment of crucial TFs such as SP1 or EBF, leading to its target's deregulation [82,83].

- In development

In ESCs, around 15% of all DNA methylation is distributed across the CpA (12%), CpT (2.6%), CpC (1.2%) sites. These regions can be located at gene bodies, promoters, gene untranslated regions (UTRs), and non-coding distal or proximal regions. Often co-localizing with dense CpG regions or in completely different zones, which most likely reflects a lack of consensus on gene expression influence Current understanding links CpA and CpT dinucleotide methylation to early development in mammals. Genome wide analysis showed that those marks are predominantly found in pluripotent cells or induced

pluripotent stem cells (iPSCs), although rare and highly variable, and disappear upon differentiation [84]. The differential distribution and enrichment of specific motifs between different cell types suggest the mechanisms by which non-CpG regulation might take place. For instance, methylation on the CAG motif mostly occurs in ESCs, while in neurons, it is mostly enriched at CAC regions [85]. Interestingly, the methylation status of these regions does not correlate strongly with gene expression, as hyper-methylation was found in actively transcribed genes in ESCs while hypo-methylated in neurons [86].

Non-CpG methylation dynamics can also be found in human oocyte maturation where promoters of *Peg1/Mest* imprinted genes followed this dynamic pattern, suggesting how this modification occurs transiently but potentially regulates genomic imprinting. Similarly, the influence on genetic inheritance was reported for the *Nf1* gene, which was strongly enriched for non-CpG methylation in maternally derived alleles from oocytes [79,87,88].

- In cancer

Non-CpG methylation has also been associated with alteration in tumor suppressor and oncogenic genes, particularly in carcinomas and breast cancer [89,90]. In lung carcinoma, abnormal clusters of high non-CpG methylation were detected in exonic regions of *TP53*, which were not found in healthy paired samples. Additionally, in prostate cancer cells, the Even-Skipped Homeobox 1 (EVX1) and the Filamin A- Interacting Protein 1-Like (FILIP1L) were up to 80% non-CpG methylated and presented hypermethylated profiles [91].

Other hypothesis points to the importance of non-CpG methylation in mitochondrial DNA (mtDNA), which is predominantly methylated in non-CpG regions [92,93]. Importantly, the methylation patterns appear to be significantly different between healthy and cancer cells, affecting structures such as the D-Loop region responsible for replication, where nearly 50% of DNA methylation is found in non-CpG regions. The modulation of mtDNA levels through abnormal replication might have some implications in cancer, but the functional consequences have not been fully explored yet [94]. In addition to claims suggesting complete absence of non-CpG methylation events in mtDNA [95].

Overall, there is strong evidence of the biological relevance of non-CpG methylation in different cell types, which might be relevant in embryogenesis, aging, and cancer development, although specific events and the difference between CpG and non-CpG methylation patterns that influence these processes are not fully understood.

Although intriguing, our work has been focused on better known CpG methylation. Which functions are broadly explained in the following sections.

2.1.2 CpG methylation

CpG methylation (hereafter DNA methylation or DNAm) is an essential epigenetic control mechanism in mammalian biology. Its widespread genome distribution has led to a much more extensive characterization compared to non-CpG methylation [76,77]. During development and differentiation, cells are directed toward their future lineages, where DNAm poses a fundamental epigenetic barrier that guides and restricts differentiation and prevents regression into an undifferentiated state.

2.1.2.1 DNA methylation biological relevance

The deposition and removal of DNAm occur in a cyclical manner and require the enzymatic activity of multiple proteins (**Fig. I2**). The methylation balance is maintained by the DNMT family of proteins, which deposit the methyl mark on the “naked” cytosine, and the TET family of proteins, which are responsible for the oxidation of 5mC, which can be ultimately removed by additional mechanisms.

Briefly, most CpG sites are symmetrically methylated through *de novo* methylation mechanisms catalyzed by DNMT3 enzymes. To sustain these modifications, the “maintenance” DNA methyltransferase DNMT1 re-establishes the DNA pattern of the template strand onto the newly synthesized one during cell division. Once methylated, the modified cytosine can be the substrate of a stepwise TET-mediated oxidation process that forms 5hmC, 5fC, and 5caC.

- In transcriptional activation

Most studies suggest that 5mC is associated with transcriptional repression, while its removal has been associated with chromatin relaxation and gene activation [96].

Stable unmethylated regions, such as CpG islands (CGIs), are associated with active expression of housekeeping and tumor suppressor genes [97]. However, the majority of CGIs present low methylation independently of whether the gene is transcriptionally active or not, suggesting a much tightly fine-tuned and complex regulatory network. The methyl mark itself does not appear to silence transcription; hence, the exact methylation-dependent transcriptional modulation is not fully understood and most likely occurs in a cell-specific manner with a combination of several epigenome layers.

The general idea is that once the methyl mark is deposited, it can be specifically read by methyl CpG binding domain-containing transcription factors, including activators and enhancers. Extensive studies have identified numerous TFs that preferentially bind methylated sequences, such as important developmental factors including OCT4, KLF4, HOXB13 or C/EBPα [98,99]. Transcription modulation might be directly associated with TFs binding and/or by recruiting additional proteins to facilitate chromatin remodeling. This is the case of C/EBPα and KLF4, which bind to highly methylated enhancers and recruit TET2 to facilitate DNA demethylation and transcriptional activation during cell reprogramming [100]

- In transcriptional repression

Despite well-known repressive role of DNA methylation, some findings indicate an inverse effect. Such as gene upregulation upon DNAm deposition.

For instance, regions decorated with the repressive H3K9me3 or H3K27me3 marks by polycomb repressive complex 2 (PRC2) are usually DNAm exclusive [101,102]. When these get methylated, it leads to transcriptional activation which was reported in healthy and disease conditions, such as breast cancer [103]. Similarly, during neural or endoderm development, DNMT3A recruitment and

proximal promoters' methylation seems to displace the PRC2 complex leading to H3K27me3 loss and activation of relevant genes such as NKX2-2 or FOXA2 [104].

This balance is further supported by the fact that PRC2 may directly interact with DNMTs, and potentially contribute to its recruitment to initiate the methylation cascade that displaces PRC2 in cell-specific developmental conditions [105].

- In chromatin remodeling

DNA methylation can also directly contribute to chromatin remodeling through interactions mediated by DNMTs or by the 5mC mark itself. Extensive studies suggest that DNMT3s and DNMT1 can form catalytically active complexes with chromatin remodeler lymphocyte-specific helicase (LSH), histone methyltransferases (EHMT2) or deacetylases (HDAC1/2). These interactions are essential for gene silencing during lineage commitment and differentiation, where LSH enhances the recruitment of DNMTs and EHMT2 to specific loci [106–108].

Similarly, the remodeling can be directly mediated through 5mC and methyl-CpG-binding domain protein (MBDs), which in mammals include MBD1–MBD4 and methyl-CpG-binding protein 2 (MeCP2). These have been reported to recruit histone deacetylase and nucleosome remodeling complexes necessary for gene silencing [109]. Importantly, in most of the cases, methylation density specifies MBD protein binding. However, there are some cases of recruitment to unmethylated sites through the NuRD-complex, suggesting methylation-dependent and independent mechanisms, which are not fully understood yet [110].

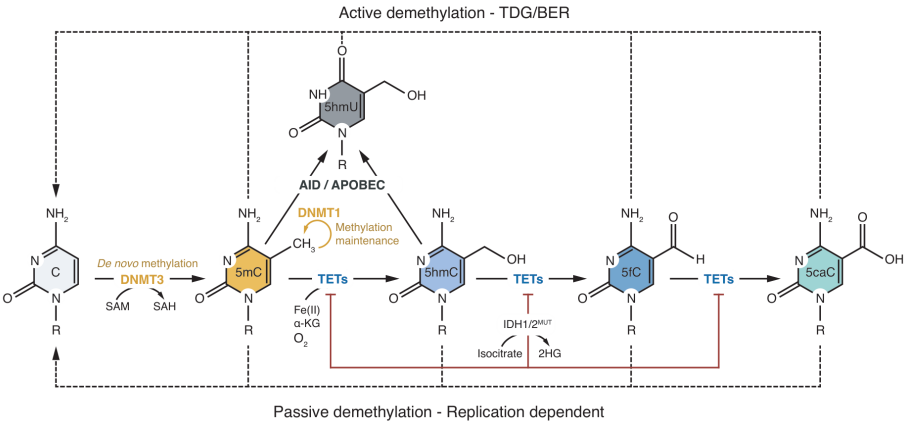

Figure I2. Overview of DNA methylation/demethylation cycle. Main players and their activities are illustrated.

2.1.2.2 DNA Methyltransferases (DNMTs)

DNA Methyl Transferases (DNMTs) are a conserved family of methylases that catalyze the addition of methyl groups from S-adenyl methionine (SAM) to the 5' position of cytosine bases. The mammalian genome encodes five members of the family: the canonical DNMT1, DNMT3A, and DNMT3B forms responsible for the aforementioned cytosine addition and the non-canonical forms DNMT2 and DNMT3L.

The general protein structure of this family consists of a C-terminal methyltransferase catalytic domain and an N-terminal regulatory domain. However, some additional function-specific sub-domains can be found in each member (**Fig. I3**). Interestingly, recent findings in mice identified a novel physiologically relevant DNA methyltransferase Dnmt3C, derived from Dnmt3B duplication [111]

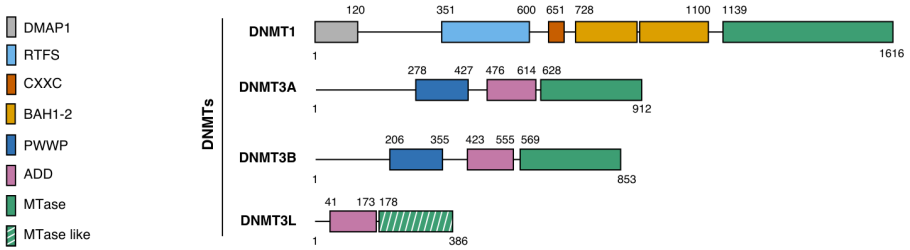

Figure I3. Structural overview of the most known DNMTs.

Structurally, DNMT1 is the largest member, presenting several smaller N-terminal sub-domains that mediate molecular interactions. Such as DNMT1-associated protein 1 (DMAP1) binding domain for interactions with other regulators such as previously mentioned histone deacetylases [107]. In addition to other components such as CXXC zinc-finger DNA binding domain and the catalytic methyltransferase region (MTase).

DNMT3 members, besides the MTase, share Pro-Trp-Trp-Pro (PWWP) and ATRX–DNMT3–DNMT3L (ADD) domains crucial for chromatin interaction [112–114].

Despite the common DNA methyl transferase activity between the active members, two different biological contexts for methylation deposition are known.

- De novo DNA methylation

In mammals it is mainly established by DNMT3A and DNMT3B members. The non-canonical forms, DNMT2 and DNMT3L, although lacking clear DNA methyltransferase catalytic activity share a conserved sequence and maintain some regulatory functions. DNMT2, which solely consists of a catalytic domain, methylates small tRNAs [115], and DNMT3L, which lacks the conserved catalytic domain, functions within the complex of DNMT3A forming heterodimers and influences the methylation during processes such as germline imprinting [116].

The DNA methyltransferase catalytic activity has been tightly linked to histone decoration within different gene regions. CpG-rich promoters of transcriptionally active genes, such as housekeeping genes, usually present the H3K4me3 mark, which is crucial to avoid DNAm deposition. The histone tail K4 is the binding site for the ADD domain of the DNMT3 enzymes, which is typical of methylated-mediated promoter inactivation. Hence, the presence of K4 trimethylation disrupts this interaction. Leading to an auto-inhibitory state where ADD binds the MTase domain to avoid unwanted methylation [117].

Contrarily, DNAm is highly prevalent in gene bodies of transcribed genes. One of the explanations is the presence of H3K36me3 mark deposited by the histone methyltransferase SETD2 following RNA polymerase II (Pol II) activity. The conserved PWWP domain can interact with H3K36me3 and most likely favors the deposition of the 5mC mark at gene bodies [114,118]. Although it has been established that DNMT3s can bind to DNA through the PWWP domain, the exact mechanisms of targeted recruitment are not fully understood and most likely occur by transcription factor interactions in a context-dependent manner.

- Maintenance DNA methylation

Once methylated, cells can copy the symmetrical CpG methylation from the parental DNA strand into the nascent DNA and conserve this epigenetic information across cell division. This is reliant on the DNMT1 enzyme and the E3 ubiquitin-protein ligase UHRF1. Specifically, during the S phase, UHRF1 recognizes and binds newly replicated hemimethylated DNA [119].

The targeting at the replication fork occurs through the SET and RING-associated (SRA) protein, which is guided by the UHRF1 tandem TUDOR-PHD (TTD-PHD) domain capable of recognizing H3K9me2 and H3K9me3 marks. Like DNMT3s, when not recruited, DNMT1 is auto-inhibited through RFTS pocketing in the MTase domain. However, once a proper recognition of the target nucleotides is done, UHRF1 mediated ubiquitination of the histone H3, which is recognized by the RFTS region, freeing the MTase domain for methyl group deposition [120].

2.1.2.3 DNA methylation genomic distribution and its relevance

The DNAm landscape is distributed in a bimodal fashion, with approximately half of the genome being covered by large partially methylated domains (PMDs) and highly methylated domains (HMDs). The other half presents lower methylation levels, mainly encompassing large unmethylated DNAm valleys (DMVs) of several kilobases and shorter unmethylated regions, such as the CGIs [121,122].

Interestingly, about half of the CGIs are not located within gene proximity or associated with any promoter. These have been termed as orphan CGIs (oCGIs) [123]. Independent of their location, they exhibit transcriptional activity and present dynamic patterns of methylation, which have been associated with development, tissue specification, and gene regulation. For instance, through modulating distal gene regulatory elements or acting as alternative promoters [124,125].

Region-specific methylation is crucial to understanding its functional relevance. While, most methylated CpGs are scattered throughout the genome and function as a repressive mark for repetitive elements and retrotransposons, gene regulatory elements (GREs) are the most studied regions. Unsurprisingly, these are the ones that show clear methylation-mediated transcriptional patterns. Despite the clear importance of methylation of promoters and an easier transcriptional correlation, DNAm at intra and intergenic regions has been shown to participate extensively in a variety of biological processes.

- In promoters

Despite widespread genome-wide DNA methylation, dense CGIs frequently lack DNAm and are associated with promoters. Indeed, around 70% of annotated genes have CGIs within their promoter region, encompassing housekeeping, developmental, and tissue-specific genes [126]. Methylation of promoters of CGIs usually leads to transcriptional repression through chromatin closing and impaired transcription factor binding [76]. For instance, crucial silencing in developmental processes such as genome imprinting or chromosome X inactivation are mediated through promoter methylation (see 2.1.2.7 DNA methylation patterns during mammalian development).

However, the precise role of CpG islands in regulating gene expression is still being elucidated on a cell-type basis, as positive or no transcriptional correlation is not uncommon [127]. Nerveless, a major subset of genes, is methylation stable and lifelong silenced in somatic tissue through several mechanisms at different stages of development.

- In enhancers

Distant cis-regulatory elements, such as enhancers, are crucial for fine-tuning cell fate maintenance and acquisition. Chromatin opening and transcription factor binding reshape chromatin configuration, looping genomic regions and influencing transcription through enhancer-promoter contacts [128]. This regulation is relatively insensitive to the distance or position of their targets [4,129].

Active or poised enhancers co-localize with other epigenetic marks, such as histone modifications and can modify the affinity of TF binding through DNAm on their recognition motif region [98,130]. For instance, this is the case for crucial cell lineage and differentiation regulators such as CTCF, NRF1, CEBPA, OCT4 or Myc/Max [100,131,132]. The activation of enhancers usually is accompanied with a loss of methylation being those elements one of the most variable class of differentially methylated regulatory elements in the genome [133].

Importantly, although this is a widespread mechanism that has been detected in numerous cell types and tissues, it is necessary to discern the cause or consequence of methylation loss after transcription factor binding. Active enhancers have a ranging median methylation between 10-60%, hence the need to discern methylation dynamic subset of regions to establish any causality [84,134]. The complexity of this axis in those regions is yet to be elucidated on a cell-to-cell basis. Being also necessary the integration of other informative layers, such as chromatin accessibility or the production of short enhancer-derived RNAs (eRNAs) to establish a solid map of methylation-regulated enhancers.

Of note, the work presented in this thesis extensively addresses this question by establishing a bona fide epigenetic network around TET2 occupancy in a model of experimentally induced cellular conversion. This network allows intersecting DNAm changes with a vast array of datasets, including histone marks, chromatin accessibility, and eRNA production.

- In gene bodies

Another interesting case of DNAm is occurring at gene bodies, a highly evolutionary conserved event that has been linked to increased gene expression, although it is not fully understood yet.

Several reports indicate its importance in transcriptional elongation, splicing and silencing of cryptic intragenic promoters [76]. Specifically, there is an overall enrichment of methylation at exonic vs intronic regions, which might suggest functional implications in splicing and gene expression. Higher CG content appears to epigenetically mark certain exons with high nucleosome occupancy compared to flanking introns, which are often found in proximity to splice sites. In addition, lower methylation levels at alternative exons suggest specific methylation architecture at these regions [135,136]. For instance, gene body methylation alters CD45 hematopoietic marker transcription through Pol II efficiency. In this case, the CCCTC-binding factor (CTCF) is responsible for slowing the elongation rates at specific exons, allowing its inclusion. The presence of DNAm prevents CTCF binding, leading to exon skipping [137].

On the contrary, higher DNAm levels at constitutive exons might promote exon recognition by recruiting MeCP2, which favors histone hypoacetylation and exon inclusion [138]. Some reports also suggest that MeCP2 influences mRNA splicing through interaction with the 5hmC mark [139].

On the other hand, the presence of DNAm at gene bodies also fits with the idea of intragenic promoters silencing, which has been linked to the H3K36me3 mediated methylation machinery recruitment and intragenic transcription silencing [140,141]. Importantly, these alternative promoters are highly conserved, suggesting biological roles in cell-specific contexts, being DNAm a

potential switch for their activity. Silencing of cryptic elements falls into this idea; however, this inhibition is rare and inconsistent among DNMT3s depletion experiments, at least in mouse ESCs [142,143].

2.1.2.7 DNA methylation patterns during mammalian development

DNA methylation plasticity is crucial during early developmental stages. During gametogenesis and embryogenesis, there are multiple waves of DNAm which are crucial to coordinate faithful epigenetic transmission and somatic lineage silencing, such as maternal and paternal imprinting through X chromosome inactivation.

Primordial germ cells undergo sequential demethylation and re-methylation waves to establish sex-specific patterns. In the pre-implantation embryos, once fertilization occurs, the parental epigenetic marks are actively erased except at imprinted genes and other elements, such as retrotransposons, just to be re-methylated again in early blastocyst after implantation to restore the initial embryonic state [76].

Since the specifics of developmental DNA kinetics are highly complex and outside the scope of the thesis, we will highlight only the main DNAm-regulated mechanisms during the process.

- Genomic imprinting

In gametogenesis DNAm is differentially distributed in the two parental germlines creating epigenetically imprinted patterns that are maintained during early embryogenesis. This is known as genomic imprinting. Approximately 20 imprinting control regions (ICRs) have been described, most of which are methylated forcing gene silencing and mono-allelic expression [144].

While in oocytes the majority of ICRs are within CGIs, the paternal imprinting is less prominent and coincides with CpG-poor intergenic regions [145,146]. In oocytes the process is dependent on DNMT3A/L which specifically targets CGIs region including promoters and intragenic region. Importantly, in many cases oocyte transcription starts from alternative promoters which in many cases

coincide with sequences of mammalian apparent long terminal repeat retrotransposons (MaLR), which are usually DNAm silenced [147–149].

Additionally, maternal ICRs are enriched in specific genetic motifs such as TGCCGC, which once methylated can be recognized by the Krüppel-associated box (KRAB)-containing zinc-finger protein 57 (ZFP57). This association allows recruitment of other silencing proteins such as KAP1 and DNMTs, being this crucial for post-fertilization silencing maintenance, allowing the imprinted promoters withstand the early embryonic global DNAm erasure and re-establishment. The ZFP57–KAP1 complex can recruit the SETDB1 methyltransferase and DNMT1/UHRF1 complexes to ensure gene silencing through replication [150,151].

- Silencing of germlime-specific genes

The silencing of CGI promoters also occurs in germline-specific genes during early development before somatic differentiation [152].

These follow a sensitive DNAm silencing mainly through DNMT3B. The enhanced methylation sensitivity has been linked to a non-canonical Polycomb repressive complex 1 (PRC1) known as PRC1.6, which consists of DNA binding proteins such as MAX, MGA, E2F6, and the chromatin remodeler L3MBTL2 (Qin Cell Stem Cell 2012). The latter can interact with the H3K9 methyltransferase complex EHMT2, which is necessary for DNMT3B recruitment and consequential methylation of specific germline genes [153,154].

DNMT3B-mediated DNAm is most prominent in the post-implantation stage, while the PRC1.6-mediated silencing happens during the pre-implantation stages. These findings were confirmed in *Ehmt2* mutant mice embryos, which showed abnormal DNA methylation, although the full mechanism of DNMT3B recruitment, and not DNMT3A, is not fully characterized [155].

- X-chromosome inactivation

In mammalian female development, one X chromosome in each cell is randomly silenced by the long non-coding RNA X-inactive specific transcript (XIST). Temporarily, X-linked CGIs methylation silencing occurs relatively late in the development constituting the final inactivation for already transcriptionally silenced genes [156].

The promoter silencing mainly depends on DNMT3B de novo methylation and the chromatin regulator SMCHD1, which defines compartmentalization on the inactive X chromosome [157,158].. The exact mechanism of DNMT3B recruitment is not yet fully understood as the process is tightly controlled and involves additional layers such as Polycomb-mediated silencing and histone acetylation through XIST transcripts, both of which occur before DNA methylation. The highly regulated heterochromatin in the X chromosome might play a role in facilitating an environment where these late DNA methylations of X-linked CGIs can occur [159].

- Genome integrity

Among other mechanism, DNA methylation can also influence genomic instability through silencing of DNA repair genes or by inhibiting the recruitment of DNA repair [160,161]. DNAm is also considered as a defensive mechanism against transposable elements, which are crucial modulators of genome integrity. In fact, most of the DNA methylation in the genome occupies transposable elements such as retrotransposons. However, it also controls other elements, such as satellite repeats found in telomeric, centromeric, and pericentromeric regions [162].

Transposable elements are highly prevalent, occupying almost half of the genome, and are mainly controlled through CpG-rich promoter methylation, which ensures proper transcriptional silencing. Unsurprisingly, alterations in the DNAm machinery led to strong global retrotransposons reactivation, such as occurring for the intracisternal A particles (IAP) after Dnmt1 depletion, which results in severe developmental defects and prenatal lethality [163].

Besides classical DNMTs, a recent genetic screening identified a specific Dnmt3c in mice that was heavily implicated in retrotransposons regulation [111]. Although lacking the chromatin association PWWP domain Dnmt3c conserves the ADD and MTase regions, hence maintaining its catalytic functions. Importantly, it is found specifically in male germ cells, and when depleted, although not lethal, it leads to meiotic catastrophe, fertility issues, and abnormalities in sex organs, which resembles the DNMT3L knock-out phenotypes [116].

2.1.2.8 Ten-eleven translocation methylcytosine dioxygenases (TETs)

Following the DNA methylation cycle (**Fig. I2**), the conversion of a methylated cytosine back to its un-modified state can occur through passive and active mechanisms, both of which are crucial for correct development and lineage commitment.

- Passive demethylation

Passive demethylation refers to a lack of DNAm maintenance across cell division through replication-dependent dilution. This mainly relies on the recruitment of DNMT1 to hemimethylated DNA (see 2.1.2.2 DNMTs) that, when failing, results in the absence of the methyl groups in the newly replicated DNA strand, hence the term “dilution”.

In addition to direct DNMT1 control, additional mechanisms have been reported to alter this maintenance. For instance, in mice early embryos, DNMT1 is physically excluded from the nucleus following waves of demethylation during pre-implantation [164].

Additionally widespread deposition of demethylation intermediaries seems to directly promote passive mechanisms [165]. In mouse preimplantation development, oxidized methyl marks are dependent on the replication-dependent dilution to un-modified cytosine [166], and synthetic experiments suggest that the presence of the oxidized 5hmC mark can lead to altered

DNMT1 methylation activity [167]. The latter could be mediated through the DNMT1 interactor UHRF1, which was found to bind the 5hmC mark, suggesting a possible mechanism where DNMT1 is recruited to hemi-hydroxymethylated replication sites [168].

- Active demethylation

The active conversion of a methylated cytosine back to its unmodified state does not happen in a single step. Instead, it occurs as a part of a sequential process, involving the modification of a 5mC through deamination or oxidation, followed by the nucleotide replacement (**Fig. 2**). Several mechanisms have been proposed to mediate this process.

Early work focused on finding a direct demethylating enzyme (as those existing in plants), as some of the oxidized products can, in theory, be directly converted to unmodified cytosine. Synthetic studies showed a possible reverse function of DNMT3A/B enzymes, where they could convert 5hmC/5caC to cytosine in the absence of the methyl donor SAM [169]. Although theoretically possible, the absence of SAM in a biological setting is improbable.

A better-known mechanism centers around the activation-induced deaminase (AID) that deaminates cytosines to uracil and 5m-cytosines to thymine as a first step of the process. It is mainly expressed in lymphoid cells in the germinal center and has been associated with cancer and autosomal recessive form of hyper-IgM syndrome (HIGM2) (Mechtcheriakova Cancer Immunol Immunother. 2012, Kumar Nature. 2013, Dominguez Cell Rep 2015). Although also linked to methylation removal in progenitors and myeloid cells [170,171]. Some reports about the involvement of other compensatory deaminases, such as APOBEC3, has been proposed but are poorly supported to be a significant factor during active demethylation [172].

Currently, Ten-eleven-translocation (TET) proteins are considered the main players in active demethylation. Originally, TET1 was discovered in acute myeloid leukemia containing the t(10;11)(q22;q23) translocation as part of TET1-MLL fusion [173].. From there, other family members, including TET2,

TET3, and IDAX, were characterized in mammals and other species. Now we know that TET proteins are large (~180 - 230-kDa) multidomain enzymes responsible for the iterative oxidation of 5mC to 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC).

All of the members contain a conserved catalytic region in the C terminus consisting of a cysteine-rich and a double-stranded β-helix (DSBH) domain with a low-complexity insert and binding sites for the Fe(II) and 2-oxoglutarate (2-OG) cofactors (**Fig. I4**). Additionally, a zinc finger DNA binding domain, CXXC, is also found in TET1 and TET3 members, which is involved in their association with the DNA. However, in TET2 this domain is found on a separated polypeptide called IDAX due to an ancestral chromosomal inversion [174].

As a consequence, all the family members, and especially TET2 which must always be recruited to specific genomic locations through a 'partner' protein, establish complex protein interaction networks. Although TET1 and TET3 are known to have a DNA binding motif, its weak affinity and reports of novel TETs isoforms that lack CXXC [175] suggest the importance of protein interaction and recruitment to specific regions by TFs for all the members of the family [176]. Once recruited, the core catalytic region has a preferential affinity for 5mC at CpG dinucleotides, leading to efficient oxidation without affecting surrounding regions [177].

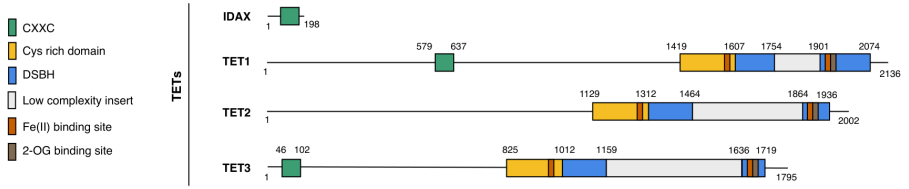

Figure I4. Structural overview of the TET family of proteins.

In the final step, the most well-established player of demethylation is a base excision repair protein, thymine DNA Glycosylase (TDG). It has been extensively shown how TDG can efficiently excise 5-fC and 5-caC marks generated by

TETs, generating an apurinic/apyrimidinic (AP) site that is finally regenerated to un-modified cytosine in both strands [178].

2.1.2.9 Demethylation intermediaries

The intermediate demethylating marks, although chemically necessary, also appear to play a regulatory role by themselves. Notably, 5hmC and its higher oxidized derivatives should be considered not only as mere transient states but as bona fide epigenetic marks as illustrated by their complex network of specialized readers within development and disease [179]

The intermediaries generated by TETs accumulate at different rates at specific genomic regions. For instance, in mouse ESCs, 5hmC is much more prevalent compared to 5fC/5caC. TETs convert approximately 10% of 5mCs to 5hmCs, and only 1-10% of those are further oxidized [180].. These ratios are maintained for several reasons. On one hand, the base excision mechanisms needed for the re-establishment of an unmodified cytosine only act on 5fC/5caC marks. On the other, TETs seem to prefer the 5mC mark over its oxidized forms, indicating substrate preference [177].

All of this, together with other factors such as chromatin accessibility, determine whether a specific CpG maintains one mark or another or gets rapidly converted back to an unmethylated state. For instance, 5hmC-interacting protein Sall4 recruits TET1 to actively demethylate enhancers, but also needs of TET2 to complete the process [181]. These dynamics can be directly related to transcription factor binding as RNA polymerase II has been shown to directly bind with 5caC or 5fC, influencing elongation rates [182]. In addition, the deposition of oxidized marks directly influences the thermodynamic stability of the local DNA, which leads to decreased DNA supercoiling and packaging for better transcription factor chromatin accessibility [183].

2.1.2.10 Demethylation in development

During development, the levels of different TET family members fluctuate, which coincides with waves of demethylation following fertilization and after germline cell specification.

Briefly, *TET3* is highly expressed in oocytes and pre-implantation embryos and rapidly decreases as cells progress towards the blastocyst stage. Conversely, *TET1* and *TET2* present high levels in the blastocyst stage and maintain relatively high levels during germ layer specification.

TETs' importance in embryo development is further highlighted in ESCs with complete TETs knockouts that cannot go through proper embryogenesis, showing defects as early as the gastrulation stage [184,185]. Moreover, when Tet activity is absent, mouse embryonic fibroblasts (MEFs) cannot undergo a proper mesenchymal-to-epithelial transition (MET) during reprogramming into iPSCs [186]. This is likely caused by increased methylation of the promoters of the *Lefty1-2* genes, resulting in their silencing and leading to enhanced Nodal signaling [187]. Loss of either *Tet1* or *Tet2* does not lead to prominent abnormalities at embryonic and postnatal development, while double knockout of *Tet1* and *Tet2* results in mild mid-gestation problems but viable mice, but with reduced fertility in females [188]. Thus, indicating activity redundancy and compensatory mechanisms among the family members.

High levels of TET enzymes are also maintained in some adult tissues, such as neuronal and hematopoietic lineages, particularly in the case of TET2 [189]. Importantly, pluripotent cells express both TETs and DNMTs. Their competition at methylated somatic enhancers is crucial to maintain the pluripotent state, which is lost upon TET demethylation-mediated differentiation [190].. These mechanisms are thoroughly described in the following sections.

(III) TET2 regulation and functions in hematopoiesis and cancer

Among all TET proteins, TET2 enzyme is a critical regulator of DNA methylation in cell proliferation and differentiation. It is aberrantly expressed in various malignancies, such as hematological neoplasms, suggesting its importance during cell differentiation [191]. Understanding the complexity of the transcriptional activation-methylation axis constitutes a key layer of information on how cells differentiate and specialize throughout development. This knowledge is essential for unraveling the initiation and progression of diseases.

3.1 TET2 in blood development

Hematopoiesis is one of the best-known biological processes regulated by DNA methylation. Briefly, hematopoietic stem cells are pluripotent and quiescent cells present in the bone marrow and are responsible for maintaining the homeostasis of the immune system. Within the pool of HSCs, the self-maintenance is designated to the long-term HSCs (LT-HSCs). In contrast, the short-term HSCs (ST-HSCs) differentiate into multipotent progenitors (MPPs) that have the capability of branching towards the myeloid or the lymphoid lineage.

The capacity of the hematopoietic system to respond to stress is essential for bone marrow reconstitution and proper cell fate commitment. Which, when altered, leads to an array of hematological pathologies of varying degrees. Additionally, natural processes such as aging lead to the gradual decline of this capacity, with accumulative alterations through aberrant DNA repair mechanisms, abnormal epigenetics, and skewed myelopoiesis [192].

Proper blood development has been extensively linked to DNAm dynamics conducted by the TET family of proteins [100,193,194]. Although all *TETs* are expressed in the hematopoietic system, *TET2* is the most predominant member and is commonly found to be mutated [195]. Its enzymatic activity is particularly important during myeloid lineage differentiation. Genome-wide profiling shows similar DNAm distribution across all stem and progenitor cell types, which are reduced in differentiated cells of the myeloid lineage, such as monocytes of neutrophils [196]. (**Fig. I5**). Most likely as a result of direct TET2 catalytic activity [197].

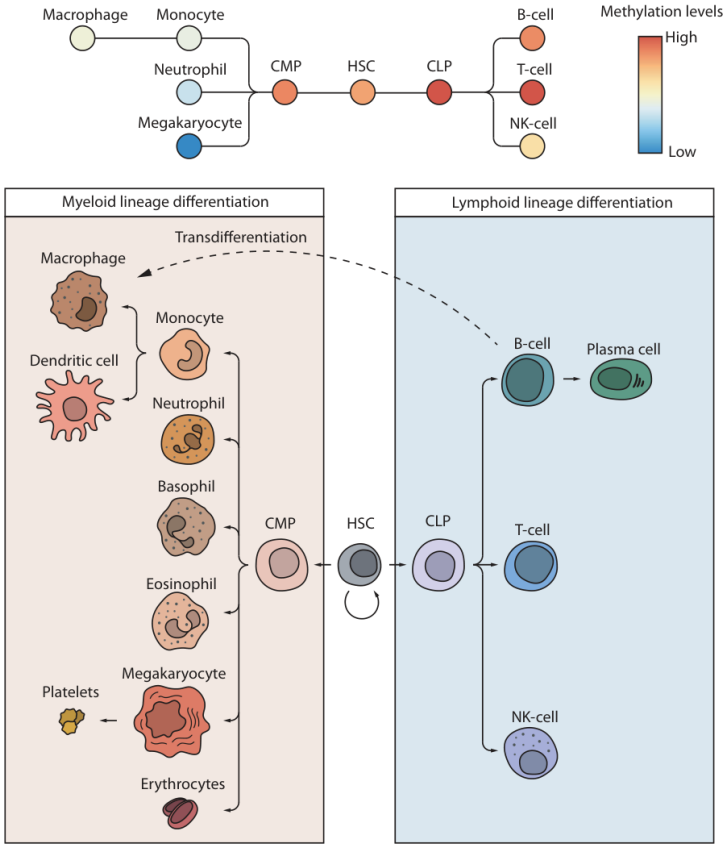

Figure I5. Graphical overview of hematopoiesis, depicting the two main lymphoid and myeloid differentiation branches, and the main cell types found in each type. Top: average DNA methylation levels at cCREs ENCODE regions (https://screen.encodeproject.org/) of selected cell types.

Since the role of TET2 in blood development is the main focus of this thesis, the following sections provide an extensive overview of the intricate molecular mechanisms controlling TET2 gene expression, protein stability and function, and the enzyme's genome recruitment.

3.2 TET2 Protein/Enzymatic Regulation

Several mechanisms regulating gene expression and activity have been elucidated in the last decades, including basal post/transcriptional regulation, direct protein modulation through post-translation modifications, and enzymatic substrate availability, all of which have been described to modulate TET2 in different biological contexts.

3.2.1 Transcriptional and post-transcriptional regulation

Some transcriptional factors are direct regulators of TETs' gene expression. In ESCs, Tet1 and Tet2 are positively regulated by the pluripotency transcription factor Oct4 that binds to conserved non-coding sequences in both genes. Accordingly, upon ESC differentiation, *Tet1* and *Tet2* levels decrease due to Oct4 depletion [198]. A similar regulatory mechanism was described for the CXXC-DNA binding domain protein Rinf (CXXC5), whose depletion leads to decreased *Tet1* and *Tet2* expression [199]. In hematopoiesis, during myeloid cell fate commitment, *Tet2* expression is boosted by the action of the myeloid transcription factor C/EBPα, which binds to *Tet2* enhancer regions [200] in addition to recruiting Tet2 protein itself to *Tet2* gene's distal regulatory regions leading to their demethylation and activation [100]. Furthermore, histone deacetylase 4 (HDAC4) protein has been recently described as a positive regulator of *TET2* expression in the context of MDS and AML [201]

Post-transcriptionally, miRNAs targeting TET2 have been proposed as the primary regulatory mechanism during blood differentiation and in myeloid malignancies. High-throughput screens identified a large subset of *TET2* 3′UTR targeting miRNAs with different efficiencies. Ectopic expression of those led to an array of leukemic traits such as myeloid lineage bias, thus phenocopying a direct *TET2* [202]. For instance, *TET2* targeting with miR-22, which is overexpressed in MDS patients, leads to reduced genome-wide levels of 5hmC, increased self-renewal, and myeloid skewing [203]. Also, in MDS, miR-9 and miR-34a indirectly control TET2 by post-transcriptionally regulating SIRT1 levels, which deacetylates TET2 protein at their catalytic domain, enhancing its enzymatic function [204].

3.2.2 Post-translational regulation

Although TET2 protein levels are mainly regulated via transcriptional mechanisms, TET2 PTMs might be involved in rapidly fine-tuning protein levels in response to external cues. For instance, during ESC differentiation, the CXXC-DNA binding domain protein IDAX (CXXC4) recruits TET2 to DNA, activating caspases that cleave the TET2-IDAX complex, leading to TET2 protein depletion [174]. Similarly, TET proteins have been described as direct substrates of the calpain family of proteases. Calpain1 regulates the degradation of Tet1 and Tet2 in mouse pluripotent ESCs and Calpain 2 of Tet3 during ESCs neural differentiation. These negative regulatory mechanisms ensure correct global 5hmC level maintenance and expression of lineage-specific genes during development [205].

Additionally, Tet2 protein can be largely post-translationally modified through specific residues (de)phosphorylation, (de)acetylation, O-GlcNAcylation, or ubiquitylation, among others [191] (**Fig. I6 top**). The specific output driven by particular PTMs on TET2 activity is cell type and amino acid residue-specific. For instance, cytokine receptor-associated JAK2, in response to FLT3 or EPO/SCF signaling in blood progenitor cells, phosphorylates Tet2 at tyrosine residues 1939 and 1964, leading to enhanced enzymatic activity [206]. Furthermore, the whole N-terminus region of the protein, constituting a large low-complexity domain (**Fig. I6 bottom**), is usually highly phosphorylated in ESCs, suggesting the rapid adaption capabilities of TET2 to modify its localization, activity, or targeting [207]. Interestingly, the O-linked N-acetylglucosaminyltransferase (OGT), a strong TET interactor, adds O-GlcNAcylation groups to serine and threonine residues of Tet2, thereby reducing the number of available phosphorylation sites and their site occupancy [208,209]. Once again, the described phosphorylation vs. O-GlcNAcylation mechanism highlights the fine-tuning TET proteins might undergo for correct localization and activity according to external signals.

TET2 activity can also be regulated through protein (de)acetylases such as NAD-dependent deacetylase sirtuin-1 (SIRT1) that removes acetylation at TET2-specific lysine residues K1472, K1473, and K1478, increasing protein's enzymatic activity. Consequently, reduced SIRT1 activity in human hematopoietic stem progenitor cells (HSPCs) leads to the onset of an MDS-like disease recapitulating the phenotype observed in TET2-mutated MDS patients

[204]. Meanwhile, TET2 global deacetylation mediated by histone deacetylases 1 and 2 (HDAC1 and 2) leads to reduced enzymatic activity, triggering the emergence of abnormal DNAm profiles typically associated with cancer. On the contrary, histone acetyltransferase p300, an enzymatic counterpart of the HDAC1/2 enzymes, was shown to acetylate the TET2 N-terminus region, leading to increased protein activity, stability and partnering with other proteins such as DNMT1. TET2/DNMT1 complex might prevent abnormal promoter methylation typically observed upon exposure to OS [210].

Finally, the ubiquitylation of TET2 has been described as a way to regulate its chromatin association. CRL4 (VprBP) E3 ligase, a member of the ubiquitin ligase complex, was found to interact with the cysteine-rich domain of Tet2 and lead to K1299 monoubiquitylation, which enhanced its chromatin association [211]. On the contrary, USP15-dependent K1299 deubiquitylation leads to decreased TET2 activity which was associated to deficient immune response in solid tumors [212].

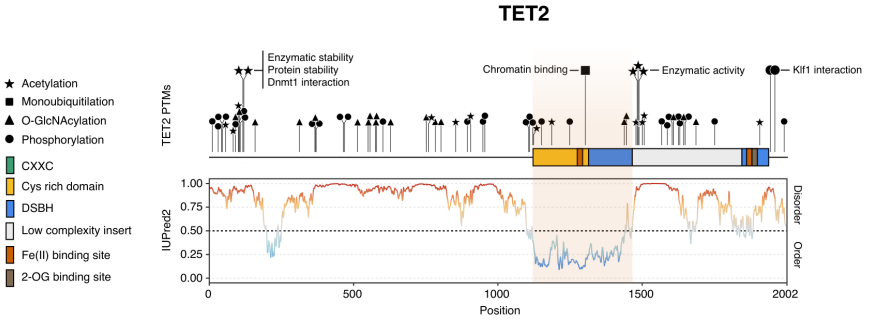

Figure I6. Schematic representation of TET2 post-translational modifications and most frequently found *TET2* mutations. Top: compiled PTM data based on mass spectrometry (MS) and in silico predictions. Bottom: intrinsically disordered protein prediction (IUPred2A, https://iupred2a.elte.hu/). The orange highlight shows the least disordered part of the protein.

3.2.3 Enzymatic regulation

As previously mentioned, TET enzymes are O2, Fe2+ and α-KG dependent dioxygenases. Metabolite and cofactor availability constitute another relevant

layer of protein regulation potentially influencing hematopoiesis and leukemic development. For instance, experimentally-induced ascorbate (VitC) depletion leads to increased HSC function and compartment expansion, resembling the aberrant self-renewal phenotype typically observed upon Tet2 depletion in HSCs [213–215]. Contrarily, related studies show how VitC treatment can rescue an aberrant self-renewal phenotype initiated upon Tet2 in vivo depletion, as well as enhance DNA demethylation and gene activation during dendritic cell reprogramming [216,217]. Alternatively, 2-Hydroxyglutarate (2-HG), an oncometabolite produced in *IDH1/2* mutated patient cells, competitively inhibits TET2 catalytic activity, resulting in genome-wide DNA hypermethylation and impaired myeloid differentiation [218,219].

Similarly, mutations in other metabolic players, such as fumarate hydratase (FH) and succinate dehydrogenase (SDH), lead to fumarate and succinate accumulation that inhibit Tet enzymatic activity even with stable α-KG levels [220]. The interplay between these metabolic intermediaries and TET2 might also be directly relevant in clinics as mutations in iron and 2-oxoglutarate-binding sites have been reported in AML patients [221,222].

Finally, although essential for TET catalytic activity, oxygen has been described as a minor direct regulator in physiological settings. However, low oxygen levels might indirectly regulate *TET2* expression and activity in leukemic cells through a mechanism involving the activation of the hypoxia-inducible factor 1α (HIF1α), establishing a complex cross-play of hypoxia metabolism and epigenetic regulation in AML [223,224].

3.3 Partner-Instructed TET2 Genomic Recruitment

Regulation of TET activity by controlling enzymes' genomic distribution allows surgically modifying DNAm at particular genomic loci and only in specific cell types. Since TET2 lacks the low-affinity (CXXC) DNA binding domain, which is present in TET1 and TET3, the enzyme must always be recruited to specific genomic locations through a partner protein, such as transcription factors. Here, we recapitulate the role of Tet2 recruitment to specific genomic regions in different lineages (**Fig. I7**)

3.3.1 During development

Molecular mechanisms underlying the TETs pluripotency-related functions have recently been partially uncovered by systematically identifying and characterizing critical Tet2 interactors in ESCs.

For instance, the naïve pluripotency TF Nanog was the first pluripotency-related protein identified to interact with Tet2 and Tet1, by mediating their recruitment to reprogramming target genes [225,226]. Similarly, Prdm14-Tet2 partnering drives active demethylation at pluripotency and germline-associated genes such as *Tcl1*, *Tcfap2c*, and *Spo11*, *Sycp3*, respectively [227]. The process was even further dissected by analyzing DNA 5hmC dynamics in iPS reprogramming systems where Tet2 is recruited by the Klf4 and Tfcp2lf1 TFs to drive active enhancer demethylation and activation at pluripotency-related genes [100].

Additionally, Tet2 has been described to interplay with a handful of non-canonical pluripotency TFs equally essential in regulating DNAm in ESCs. The latter includes the CxxC-DNA binding domain proteins Idax (CXXC4) and Rinf (CXXC5), potentially influencing Tet2 functions in differentiation and pluripotency, respectively [174,199]. Rinf co-occupies with Tet2 and Nanog distal gene regulatory regions of relevant pluripotency genes such as *Oct3/4*, Sox2, and Nanog itself, positively regulating their transcription.

Recent studies also indicate that the RNA-binding protein PSPC1 associates with Tet2 and targets the enzyme to the MERVL endogenous retroviral elements in ESCs. Once recruited to MERVL elements, Tet2 contributes to their transcriptional and epitranscriptomic regulation by recruiting HDACs to chromatin and oxidizing MERVL transcripts, respectively [228].

3.3.2 During myeloid cell fate commitment

Tet2-mediated epigenetic gene regulation is crucial during myeloid cell development. C/EBPα is an essential factor in the differentiation process from HSPCs to GMPs [229]. C/EBPα alone, through its pioneer activity, or in concert with PU.1, binds regulatory regions of myeloid genes to establish the myeloid cell fate [230]. To this end, C/EBPα directly activates Tet2 expression and, through direct interaction with the enzyme, targets it to regulatory regions of

myeloid genes such as Klf4 or Chd7, driving their active DNA demethylation and subsequent enhancer activation [100].

RUNX1, another key hematopoietic transcription factor, has also been described to interact with TET2, thus potentially leading to the DNA demethylation observed at RUNX1 binding sites, including promoters of PTPN22, RUNX1, and RUNX3, among others, during early hematopoiesis [231].

Wilm's tumor (WT1) gene encodes a sequence-specific transcription factor found mutated in a mutually exclusive manner with TET2 in AML patients and directly associates with TET2 to regulate WT1-target genes, including Wnt and MAPK signaling-related genes such as BTRC, DACT1, and TBL1X, which prevents AML onset [232,233].

Furthermore, TET2 activity is crucial in myeloid terminal differentiation [234]. For instance, during monocyte-to-osteoclast differentiation, the master myeloid TF PU.1 recruits Tet2 to promoters of key osteoclast-genes such as ACP5, CTSK, and TM7SF4, leading to their demethylation and cell fate transition [235]. Tolerogenic dendritic cells (tolDCs) are also terminally differentiated myeloid cells with potent immunosuppressive properties. Mechanistically, the tolerogenic phenotype is acquired through a synergistic interplay between the glucocorticoid receptor (GR) and the specific myeloid transcription factor MAFB. Both TFs target Tet2 at genomic loci exclusively demethylated in tolDCs [236]. In addition, EGR2, an essential transcription factor during IL4/GM-CSF-driven monocyte (MO) to monocyte-derived DCs (moDCs) differentiation, has also been described to interact with Tet2 and initiate DNA demethylation at both EGR2 stable and transient binding sites [237].

Additionally, HSPCs commitment towards erythroid lineage also correlates with 5hmC accumulation and increased expression of erythroid-specific genes such as EPOR, GATA1, and HBB [238]. 5hmC accumulation might be mediated by Klf1-dependent Tet2 recruitment at erythroid-specific genes which can be enhanced upon Jak2-driven Tet2 phosphorylation [206].

Importantly, TET2 genomic recruitment is not always associated with positive regulation of gene expression as some findings indicate that IκBζ-dependent

Tet2 targeting to the *Il6* locus leads to Hdac2 recruitment in dendritic cells. Resulting in Il6 gene repression and inflammation resolution [239].

3.3.3 During lymphoid cell fate commitment.

B cell differentiation is also tightly regulated at the epigenetic level. For instance, B cell maturation is characterized by extensive reshaping of the cellular methylome, where Tet2 might contribute to the process by interacting with B-cell master regulators such as EBF1, IRF4/8, E2A, and PU.1 or BATF. This results in focal demethylation at regulatory regions of key B-cell loci, including *IgK* or *Aicda* [240–242].

Similarly, 5hmC is also accumulated at genes encoding key regulators of T cell identity, development, and differentiation [243]. However, what factors recruit Tet enzymes to the T-cell key regulatory regions is poorly known. Findings in Tregs indicate that Tet1 and Tet2 are upregulated in response to the sulfhydrating nuclear transcription factor Y subunit beta (NFYB) in addition to being targeted by the activated forms of Smad3 and Stat5 to the Foxp3 promoter favoring its hypomethylation and stable gene expression [244].

3.3.4 In Response to External Stimuli

Previous findings have suggested that upon exposure to OS, protein repair complexes containing DNMTs [245,246] and TET2 can be recruited to the damaged DNA regions to ensure genome stability [247].

As a result, several epigenetic regulatory mechanisms that tie Tet2 to protection against abnormal DNAm during stress have been proposed. For instance, during OS, Tet2 was found to interact with the thymidine glycosylase (TDG), which resulted in both protein's recruitment to chromatin in a DNMT1-dependent manner. Suggesting to protect against the acquisition of abnormal DNAm at typically unmethylated gene regulatory regions [210].

Moreover, during DNA damage response Tet2 interacts with the SMAD nuclear interacting protein 1 (SNIP1) that regulates the expression of relevant c-MYC target genes involved in apoptosis. Therefore, reduced SNIP1 levels lead to

diminished 5hmC levels and gene expression at c-MYC target genes [248]. The latter might be relevant during chemotherapeutic drug response, which can trigger hydroxymethylation changes in a mechanism mediated by the interaction between the promyelocytic leukemia protein (PML) and TET2, linking the protein to direct clinical applications [249].

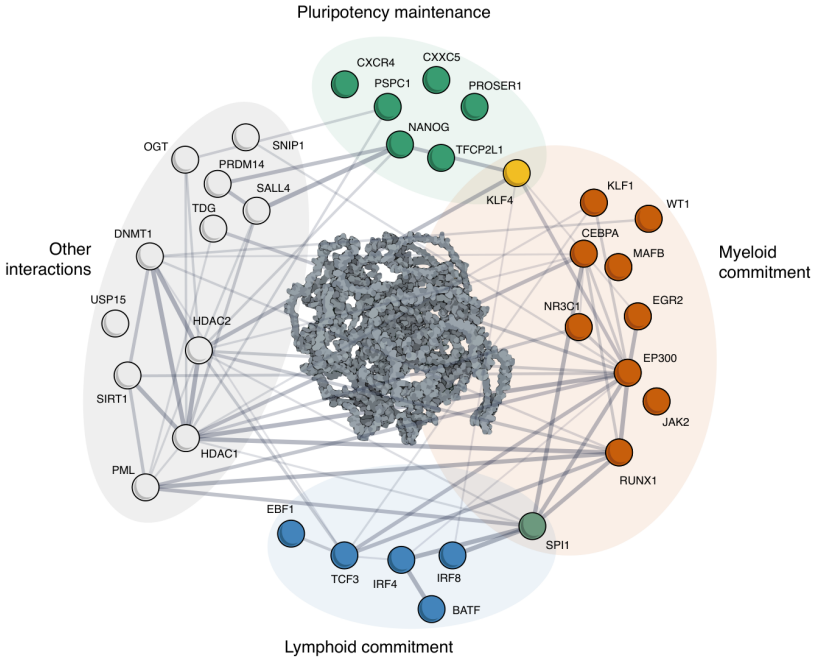

Figure I7. Graphical overview of TET2 protein partners/interactors. In the center, the human TET2 AlphaFold Database predicted protein structure (AlphaFold v2.3) [250,251]. Encompassing the TET2 protein structure, color-coded bubbles cluster of TET2 interactors based on previously described cell fate implications and physiological responses. Solid lines represent the physical subnetwork of the TET2 interactors. Thicker lines represent higher interaction confidence (STRING v12.0) [252].

3.4 TET2 in hematologic malignancies

DNA methylation aberrations are considered hallmarks of cancer onset and progression and are often associated with *TET2* mutations. Although frequently found in patients suffering from myeloid malignancies, these are also commonly observed in other types, including B and T lymphomas, such as the angioimmunoblastic T-cell lymphoma (AITL), which shows the highest incidence of *TET2* mutations (roughly 80%) among all blood cancer types. The broad *TET2* mutational profile observed in the hematopoietic system has awakened great interest in the field to unravel the molecular mechanisms underlying TET2 involvement in the onset of preleukemic and leukemic diseases.

3.4.1 Pre-leukemic conditions: Clonal Hematopoiesis

Over the lifespan of an individual some somatic mutations accumulate in the stem cells. Although the majority have no impact, some can result in an enhanced clonal expansion in certain conditions.

Epigenetic regulators, including *TET2* and *DNMT3A,* that enhance self-renewal and hinder proper differentiation are among the most commonly altered genes. This might be considered a pre-leukemic state where, according to the mutational burden, susceptible individuals have a higher predisposition to develop malignant neoplasms, both myeloid and non-hematological ones, such as cardiovascular diseases or solid cancers [253].

In hematopoiesis, loss of function mutations in *TET2* lead to an expected decrease in 5hmC [254]. It has been well reported how conditional *TET2* knockouts in mice showed an expansion of the early hematopoietic stem cell compartment, along with a hypermethylated phenotype. This same phenotype has also been observed in healthy humans with CH and it correlates with variant allelic frequencies (VAF) levels [214,255]. Mechanistically *TET2* mutant HSCs show an enhanced resistance to inflammatory-stress-induced apoptosis, such as that induced by IL-6 and/or TNF-α [256]. Ultimately presenting a competitive advantage in a pro-inflammatory environment. Moreover, in myeloid cells, TET2 appears to directly mediate active repression of IL-6 during inflammation,

suggesting the presence of pro-inflammatory conditions in TET2-depleted niches [239].

Regardless, HSCs with altered TET2 or DNMT3A activity acquire a competitive clonal advantage due to an altered DNAm landscape that finally allows aberrant gene expression. However, the precise order of events leading from CHIP to the development of myeloid malignancies is not yet fully uncovered, nor is the CHIP potential as a clinical predictor.

3.4.2 Leukemic conditions

Myeloid malignancies constitute a heterogeneous group of diseases that encompass myelodysplastic syndromes (MDSs), myeloproliferative neoplasms (MPNs), and acute myeloid leukemia (AML), all of which have been described to be influenced by epigenetics in varying degrees.

As previously mentioned, *TET2* loss of function mutations are frequently found in patients with myeloid malignancies. Inactivating mutations occur along the whole gene coding region and are not restricted to the 3′ catalytic domain region. Although unsurprisingly, the well-conserved catalytic domain is still a hotspot for the mutation (**Fig. I8 middle**). Pathogenicity AlphaMissense exploration (Cheng Science 2023) of TET2 protein reveals a high pathogenicity probability at the well-structured catalytic domain, although not exclusively (**Fig. I8 bottom**).

Some hotspots for pathogenic alterations such as catalytic domain proximal region (900-1128 aa), which coincide with exon 3 that has been reported as a hotspot for mutations in myeloid malignancies, including nonsense, insertion and deletion and premature stop codons which results in loss of function truncating proteins [215,218,257]. Additionally, the region encompassing TET2 functionally defined acetylated lysines K110/K111 [210] appear to be highly pathogenic. A closer look shows average scores of 0.59 and 0.48, which are much higher compared to proximal amino acids I109 and L112 which scored 0.16 and 0.18 respectively. The latter suggesting potential relevance of post-translational modification in TET2 abnormal activity in disease.

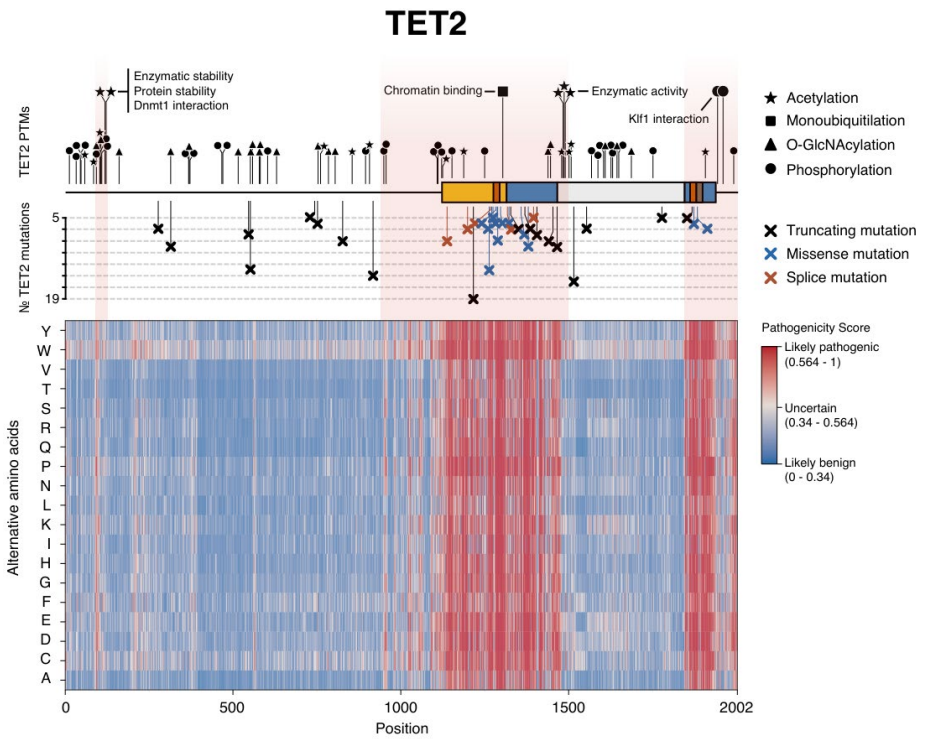

Figure I8. Schematic representation of TET2 post-translational modifications and most frequently found *TET2* mutations. Top: compiled PTM data based on mass spectrometry (MS) and in silico predictions. Middle: a compilation of *TET2*'s most frequent mutations and their type. Filtering was done from a combined 12,845 samples from 35 studies where the mutation was found in >5 different patients. Data from the 'Myeloid' dataset from cBioPortal (https://www.cbioportal.org/). Bottom: AlphaMissense Pathogenicity Heatmap. Red highlight shows the manually selected pathogenic hotspot.

Regardless, *TET2* mutations are typically considered an early event in myeloid cancer development and are characterized by presenting aberrant hypermethylation profiles due to incorrect TET2 function [221,258,259]. This is the case, for instance, of chronic myelomonocytic leukemia (CMML), a mixed MDS/MPN neoplasm where *TET2* mutations are present in up to 60% of patients. In CMML cells, aberrant methylation was observed at promoters of genes involved in neoplastic transformation, WNT and PDGF signaling pathways, inflammation and apoptosis [260,261]. Similar gene sets were observed aberrantly methylated and expressed in *TET2*-mutated AML patients,

including genes coding for tumor suppressors, transcriptional modulators, nuclear import receptors, and myeloid cytokines, such as *PDZD2*, *CDK8*, *IPO8*, and *CSDA*, respectively [218].

Global methylation analysis also revealed a shared hypermethylation signature in a subset of AML patients carrying *IDH1/2* mutations, a well-known TET2 inhibitor and mutually exclusive mutated genes. Mutations in the IDH1/2-TET2-WT1 network, which collectively appear in 30-50% of AML cases, also present an apparent hypermethylated phenotype as a consequence of deficient TET2 activity [232,233,258,262]. Several ongoing clinical trials aim to restore TET2 activity within the IDH1/2-TET2-WT1 mutated network by combining hypomethylating agents with ascorbate treatment [263,264] (ClinicalTrials.gov identifier: NCT03999723, NCT03682029).

In the lymphoid lineage, aberrant B cell differentiation and chronic lymphocytic leukemia (CLL) have been associated with DNA hypomethylation. Hypomethylation at gene bodies and enhancer regions correlates with gene expression differences in CLL samples compared to normal B cells [265–267]. However, available data do not support a primary role of TET enzymes in CLL, but perhaps an accessory role in establishing leukemia-specific patterns of 3D chromatin conformation. The latter might be accomplished, for instance, by modulating CTCF binding to the chromatin, a DNA methylation-sensitive mechanism [268].

TET2 has also been studied in the context of lymphoproliferative diseases. Diffuse large B-cell lymphoma (DLBCL) patients have shown specific hypermethylation signatures on promoters of tumor suppressor genes involved in cell fate and cell cycle changes. Germinal center analysis of B cells in *Tet2* deficient mice showed promoter hypermethylation and defective transcription factor binding at essential B-cell pathway genes. Tet2 knockout mice partially phenocopied DLBCL patients, characterized by downregulation of antigen presentation genes/interferon pathway, lymphoma-like transcriptional profiles. Which phenocopied patients with frequently mutated CREBBP acetyltransferase and a failure at the germinal center exit. Therefore, suggesting a TET2 role in B cell lymphoma development and how acquiring mutations in HSCs might influence B-cell maturation and cancer development [269–271].

(IV) Epigenetic layers axis overview

Through this introduction, we have established the importance of different epigenetic mechanisms, with a particular focus on DNA methylation. However, it is important to keep in mind that these very different layers are heavily interconnected and function in concert with each other to provide a highly adaptable and controlled gene regulatory network (**Fig. I9).**

Indeed, DNAm changes influence the recruitment of transcriptional repressor complexes, among many other functions. When the DNA is methylated, MBDs can recruit gene-silencing machinery such as HDACs, polycomb, and chromatin remodeling complexes, typically associated with genes that can go through methylation-mediated transcription silencing. For instance, the recruitment of MeCP1/MeCP2 proteins that contain a transcriptional repressor domain allows the association with Sin3 and HDAC1/2 complexes, which leads to significant gene silencing [272].

Similarly, reports suggest an inverse dependence where histone modifications might be necessary for proper DNA methylation. Histone lysine methylation mark H3K9 has been described to be at least partially necessary for the recruitment of DNA methyltransferase machinery, which can directly interact with SUV39H1 and SETDB1 H3K9-histone methyltransferases [273]. Drug studies have also highlighted the DNAm and histone modification dependencies. Such as demethylating agents 5-aza-2′-deoxy-cytidine (5-aza-dC) or HDAC inhibitors Trichostatin A (TSA), which influenced gene expression by altering the axis between both epigenetic layers [274–276].

In disease, DNAm alteration is a well-established hallmark of cancer, which in many cases is accompanied by specific histone profiles. For instance, abnormal promoter CpG island hypermethylation in cancer is usually associated with a repressive H3K27me3 mark. The evaluation of histone profiles at those regions has revealed co-enrichment with activating H3K4me3, which might define a specific subset of bivalent target genes that are more susceptible to DNA hypermethylation in cancer [277–279].

Besides clear DNAm implications in hematological malignancies, numerous solid cancers present deregulated epigenetics. In squamous cell carcinomas,

alterations in the methyltransferase NSD1 lead to abnormal H3K36me2 profiles, which impact DNMT3A recruitment, resulting in global DNA hypomethylation. Which also occurs in histone H3 lysine 36 to methionine mutation cases, which inhibit the catalytic activities of H3K36 methyltransferases, leading to the same hypomethylating phenotypes [280,281].

Similarly, in renal cell carcinomas, mutations in other methyltransferases, such as SETD2 mutations, lead to an abnormal gain of H3K36me3, especially in intergenic regions such as poised enhancers of developmental genes. Resulting in DNA hypermethylation and activation of genes related to an undifferentiated state [282].

Additional regulatory elements such as EZHIP, a PRC2 repressor, have also been reported to get activated with promoter hypomethylation, highlighting the complexity of the interplay, not only through direct site recruitment but also by regulating gene expression of network regulatory players [283]

To sum up, the pluripotent potential of stem cells is tightly related to their highly dynamic chromatin state, which is fine-tuned through epigenetic modulation. While a permissive chromatin landscape is found in the pluripotent state, it gets compacted and concentrated at specific loci once cells start to differentiate into different lineages. A combination of previously described epigenetic regulators prevents chromatin compaction and gene expression in self-renewal stages. The signals for multilineage cell fate commitment induce chromatin closure, forming tight condensed heterochromatin regions that promote silencing of pluripotency genes such as *Oct4*, *Nanog*, and *Sox2*. This is further illustrated during cell reprogramming, where the induced pluripotent state is accompanied by chromatin changes to facilitate enhanced transcription factor binding to pluripotency promoters and enhancers.

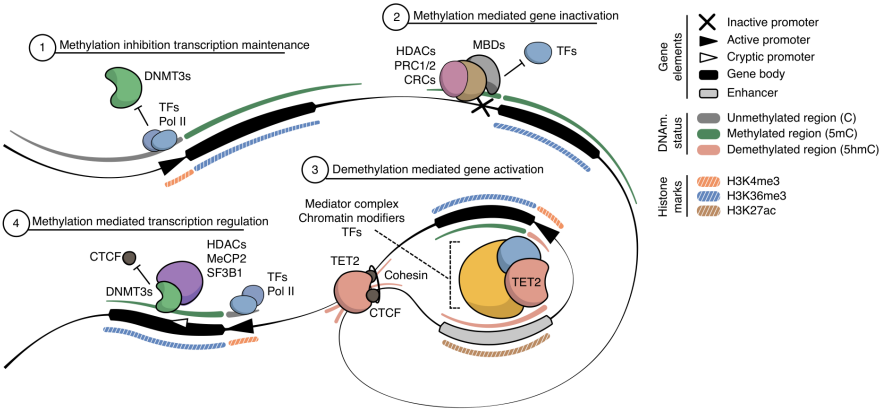

Figure I9. Graphical overview of potential transcriptional modulation mediated by DNA methylation. 1) Inhibition of DNAm deposition, characteristic of developmental genes. To avoid abnormal silencing, deposition of H3K4me3 ensures that DNMT3A/B stays in an inactive autoinhibitory conformation. 2) Canonical DNAm gene silencing. The dense deposition of the methyl mark at promoter regions allows the direct recruitment of MBDs, which facilitate the recruitment of chromatin repressive complexes (HDACs, PRCs, and chromatin remodeling complexes). 3) De-methylation mediated gene activation. Typical of enhancers, which get demethylated during cell fate commitment, allowing for TF binding and chromatin looping to establish direct enhancer-promoter contacts for strong gene modulation. 4) DNAm transcriptional modulation through alternative mechanisms, such as alternative/cryptic promoters or splicing modulation. CRCs: chromatin remodeling complexes

Hypothesis and objectives

In healthy hematopoietic development, cell fate decisions are considered a unidirectional and irreversible process. The stability of differentiated cells ensures the proper function of the system, avoiding any deviation that might lead to malignant propagation. However, the plasticity of these differentiated cells has been proven to be malleable. A number of studies have shown how transcription factor transduction, nuclear transfer, or cell fusion might lead to cell fate transitions. On one hand, a somatic cell can be reprogrammed back into a pluripotent state, but on the other, it can be directly converted into another differentiated cell type without passing through a progenitor state. The later conversion is coined as transdifferentiation (TD) [284–286].

Based on previous findings in mouse iPS reprogramming systems and the characterization of several terminal differentiation models, we hypothesize that the methylome plays a significant role in cell fate conversion processes and can be used to evaluate disease-related events [100,200,235].

Specifically, we hypothesize that TET2 is a crucial player in regulating chromatin and transcriptional dynamics during the C/EBPα-driven human myeloid differentiation. More importantly, evaluating TET2-mediated demethylation during the process could provide bona fide TET2 targets in a loss-of-tumorigenicity system, which most likely overlaps to some extent with *in vivo* leukemogenesis.

Hence, the main goal of the project is the:

> *'Identification of the major events that lead to the reversion of the tumorigenic capacity of leukemic cells upon CEBPA-dependent TET2 activation.'*

To achieve that, we took advantage of a C/EBPα-mediated, highly efficient leukemic human B-cell to non-tumorigenic macrophage conversion model (BlaER1) (Rapino Cell Rep 2013). Briefly, the system was generated by retrovirally transducing a lymphoblastic leukemia cell line with a construct of *CEBPA* fused with estrogen receptor hormone binding domain (CEBPA-ER). This fusion allows a rapid translocation of the protein from the cytosol into the cell nucleus after estradiol (E2) treatment, which initiates a loss of tumorigenicity

TD process along 6-7 days. This efficiently converts leukemic B-cells into induced nontumorigenic macrophages (iMac) with an epi/transcriptome resembling healthy cells [287]. These are phagocytic, quiescent, and highly migratory in addition to showing inflammatory responsiveness, highlighting not only changes at the omics level but also clear physiological implications [288–290].

Preliminary data to support the working hypothesis

We have compiled and analyzed several publicly available datasets as well as data from the BlaER transdifferentiation model, to reinforce the understanding of the proposed objectives.

1. Healthy hematopoiesis expression and methylation evaluation

Evaluation of normal hematopoiesis has shown that TETs have a prominent role in cell fate acquisition, particularly TET2, which plays a major role in myeloid commitment [195]. To reinforce this, we analyzed single-cell RNA-seq datasets of healthy individuals and observed, with great resolution, how TET2 is specifically upregulated in the myeloid lineage (**Fig. O1**), suggesting its importance in the process.

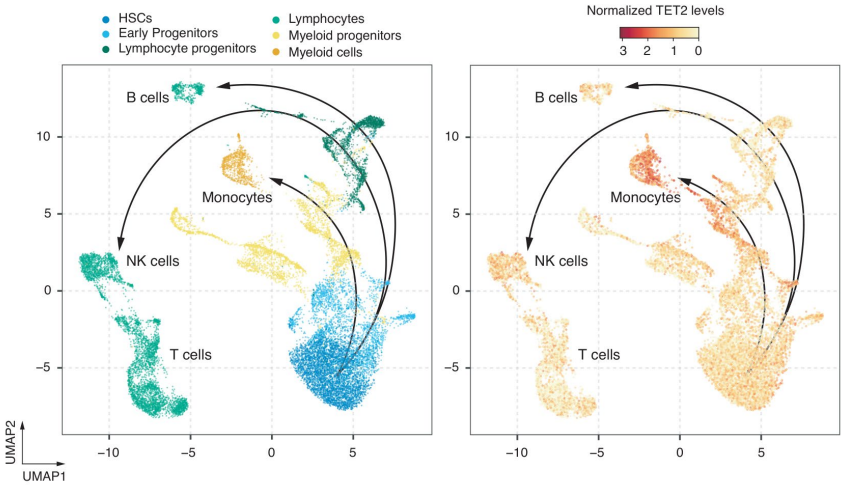

Figure O1. Single-cell RNA-seq UMAP dimensional reduction from a young, healthy hematopoiesis donor ('Young1.All.T1', [291]. Left: shows the annotated blood population. Right: shows normalized TET2 expression levels within each cell. Arrows represent theoretical hematopoiesis differentiation trajectories.

A closer look at DNAm levels in normal hematopoiesis further provides a better understanding of TET2 role in myeloid differentiation. Evaluation of DNAm levels at ENCODE human candidate cis-regulatory elements (cCREs) (https://screen.encodeproject.org/) revealed distinct methylomes among the cell types analyzed (**Fig. O2**). Notably, lymphoid lineage progenitors displayed higher methylation levels than myeloid progenitors maintained them through their maturation towards T-cells. While other lymphoid cells, such as B cells and natural killer (NK) cells displayed lower methylation than their respective progenitors, in line with the potential involvement of TET2 in B-cell germinal center maturation [270,271] and the observed hypomethylation profiles found in active NK cells [292,293]. Contrarily, in the myeloid lineage, a clear methylation loss is observed in mature populations of monocytes, neutrophils, and megakaryocytes, which is most likely directly mediated by *TET2* upregulation.

Importantly, these are genome-wide analyses that consider every type of regulatory element, including promoters, enhancers, CTCF-bound regions, and chromatin-interacting sites. The latter provides a solid overview of the global methylome levels in different populations. However, as previously mentioned, DNA demethylation events are highly cell and region-specific. Hence, a thorough evaluation is necessary to determine subsets of TET2 chromatin targets that lose methylation, which was extensively explored in myeloid cell fate commitment during the work of this thesis.

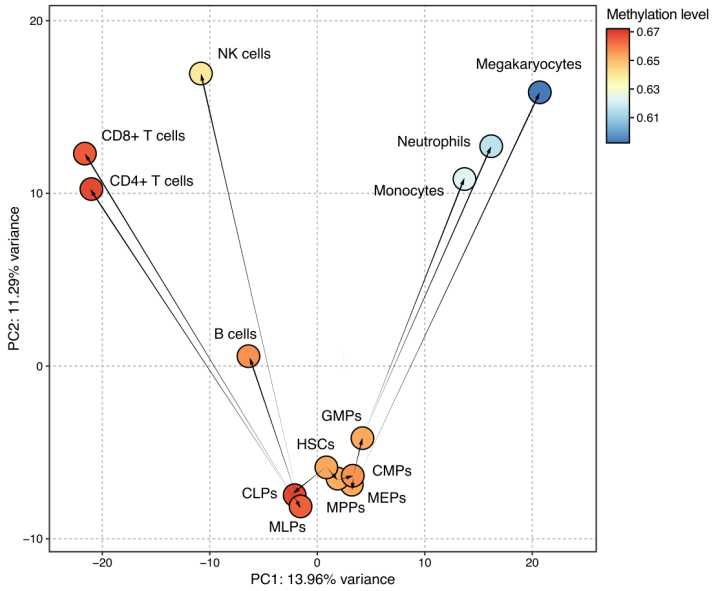

Figure O2. Principal component analysis (PCA) of the average DNAm levels at human ENCODE cCREs, in different hematopoietic populations (Farlik Cell Stem Cell 2016). Only the regions covered at least x3 were considered for the analysis (n=60,818).

2. TET2 regulation during BlaER transdifferentiation

A transcriptional evaluation during the process indicates a rapid silencing of B-cell programs, and a progressive activation of myeloid related genes. Which coincides with a strong TET2 upregulation, but not TET1 or TET3. Suggesting potential TET2 implication in myeloid cell fate acquisition (**Fig. O3 top**). Additionally, previous findings in CEBPA-mediated TD of mouse pre-B cells into macrophages showed TET2 playing a role in myeloid gene activation [200].

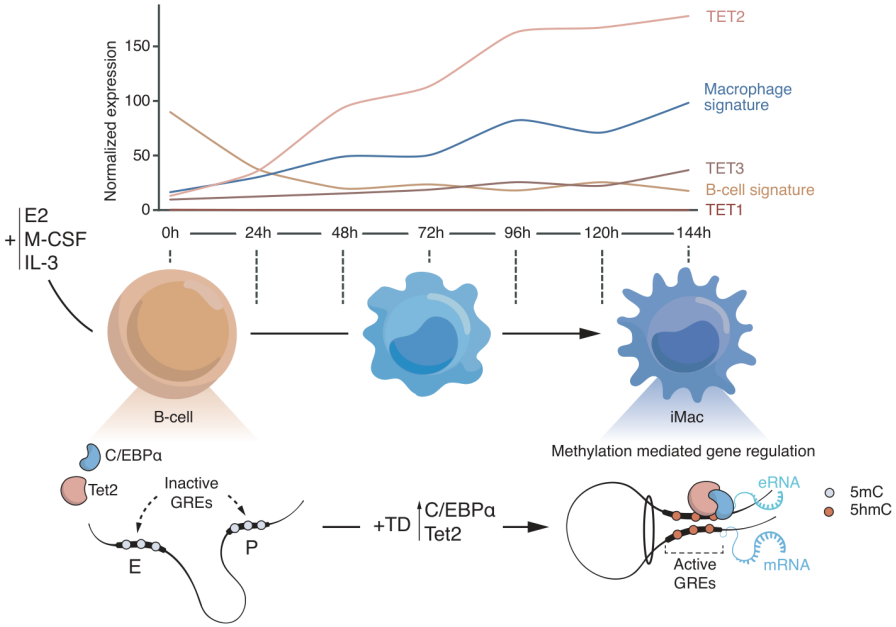

Figure O3. Graphical overview of BlaER TD process. Top: publicly available transcription data (Stik Nat Gen 2020) depicting TETs expression dynamics and the average expression B-cell and Macrophage related gene signatures. Bottom: hypothesized DNAm gene regulatory events that mediate myeloid programs activation.

Importantly we detected early binding of C/EBPα to distal regulatory TET2 enhancers, most likely favoring its rapid activation after induction (**Fig. O4**). Accordingly, hindering TET2 accumulation during the cellular conversion should deregulate the chromatin and transcriptional dynamics activated during myeloid commitment and potentially mirror, to an extent, abnormal leukemogenic hypermethylation detected in TET2 mutated AML patients.

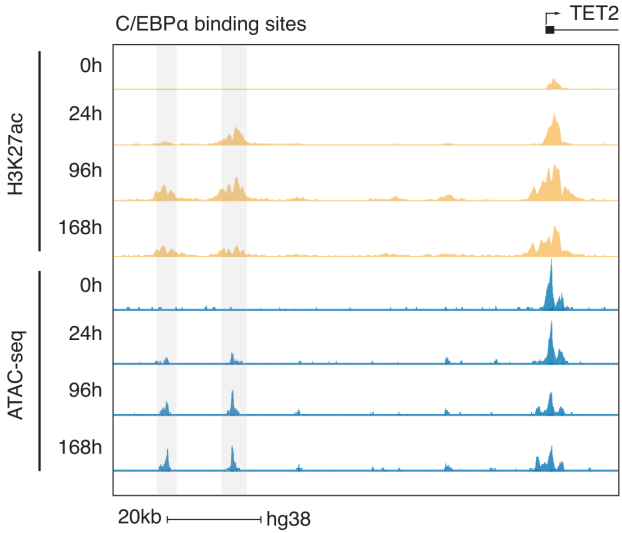

Figure O4. Snapshot of distal TET2 GREs, showing the chromatin accessibility (ATAC-seq) and H3K27ac enhancer mark deposition during TD. Grey highlight indicates C/EBPα binding sites.

Although several TET2 depletion studies have reported stem and progenitor cell compartment expansion during stress and inflammation [294] and a prominent myeloid lineage skewing [197], the ultimate events related to myeloid terminal differentiation are not fully understood, particularly at distal gene regulatory regions where TET2 has been reported to preferentially bind [295].

Project overview

To address the main goal of this work, we aim to collect and integrate data on TET2 genome-wide activity during transdifferentiation. The specific tasks of the project are the following:

T1. Identification of TET2 chromatin targets and partner proteins during induced B to macrophage cell fate conversion

Efficient TET2 immunoprecipitation has been extremely challenging, mainly limited by the lack of robust and commercially available IP-grade antibodies. Although the depletion of *TET2* in hematopoietic cells results in broad DNAm changes [296], it is still unknown which regions are directly targeted by the protein and what partner proteins participate in the process. In addition, little is known about the potential catalytic vs non-catalytic functions at these sites, which might be associated with the onset of myeloid and lymphoid lineages diseases, respectively [297]. To fully understand transcriptional changes occurring in altered TET conditions, it is crucial to investigate the TET2 interacting chromatin network instead of relying on shallow promoter-centric DNAm approaches.

To circumvent the previously mentioned technical limitations, we have inserted a 3xFLAG-tag at the endogenous *TET2* locus using CRISPR knock-in. This provided us with an efficient protein pulldown platform and led to the generation of high-quality TET2 immunoprecipitation data. This approach was applied to the following sub-objectives:

T1.1 Uncovering of TET2-associated partners/recruiters during TD

TET2 immunoprecipitation coupled with mass-spectrometry (IP-MS) was used to provide a list of high-confidence interactors, which were further validated by co-immunoprecipitation (Co-IP) experiments.

T1.2 Uncovering of TET2 chromatin targets

A dual cross-linking ChIP protocol was used to identify TET2 genome-wide binding sites, which were extensively integrated with DNAm and chromatin status data to better understand its biological role.

T.2 TET2 molecular profiling during induced B to macrophage cell fate conversion

To establish a confident correlation between methylome and transcriptome dynamics, we used an inducible system to deplete *TET2* during TD. Specifically, we have generated stable BlaER cell lines with doxycycline-inducible shRNA expression, which allowed us to interrogate the aforementioned interplay. This could be done in a reproducible manner and without limitations of the constitutive knock-down system, such as cell adaptation and activation of compensatory pathways. This approach was applied to the following sub-objectives:

T.2.1 Methylome evaluation in TET2 depletion during TD

We have used Infinium MethylationEPIC arrays to interrogate changes in the methylation landscape during TET2-depleted TD. The arrays' findings were supported by a complete genome-wide DNAm dataset (by WGBS-seq) during normal TD, which allowed for an improved understanding of the affected regions.

T.2.2 Transcriptome evaluation in TET2 depletion during TD

Total RNA-seq was used to evaluate the impact of TET2 depletion on gene expression. This dataset was assessed in relation to the global DNAm landscape to uncover the interplay between both genomic layers.

T3. Establishment of TET2 chromatin network during TD and target identification

The generated datasets (from T1. and T2.) were combined with additionally available data in the TD system to evaluate highly specific TET2 targets. The following epigenetic datasets were used: chromatin accessibility (ATAC-seq), histone marks (ChIP-seq H3K4me1, H3K27ac, others), chromatin interactions (Hi-C), and transcription factor occupancy (*CEBPA*, *BRD4/CTCF*).

T4. Evaluation of confident TET2 targets in normal and malignant hematopoiesis

The most confident TET2 targets (identified in T3.) were explored in relation to publicly available AML patients' methylation and expression datasets, including the Cancer Genome Atlas Program (TCGA) and the 'Blueprint' consortium. The latter provided a better *in vivo* understanding of TET2 activity in normal and malignant hematopoiesis and its influence on cancer epigenetics and transcriptome.

Materials and Methods

Cell culture

BlaER1 (BlaER) lymphoblastic leukemia cell line was derived from RCH-ACV cells after transduction of CEBPA fused with the estrogen receptor (ER) hormone-binding domain and a GFP fluorescence reporter (MSCV-C/EBPaER-IRES-GFP). BlaER cells and derived lines, were grown in suspension in RPMI 1640 (GIBCO) supplemented with 10% heat-inactivated FBS (GIBCO), 1X Penicillin-Streptomycin (GIBCO), 1X L-glutamine (GIBCO) and 0,1X β-mercaptoethanol (GIBCO). Culture medium was replaced every 2-3 days upon counting in a hemocytometer and using trypan blue exclusion dye to discriminate between live and dead cells. Then, cells were seeded at 2×10^5 cells/ml into an appropriate tissue culture flask. Culture density was maintained between 0.2–2×10^6 cells/ml.

HEK-293T cells were grown in Dulbecco's Modified Eagle's Medium (DMEM) (+) D-glucose supplemented with 10% heat-inactivated FBS, 1X L-glutamine, and 1X Penicillin-Streptomycin.
For optimal growth, all cell lines were kept in a 5% CO2 humidified atmosphere at 37 ºC.

Cells were checked for mycoplasma infection every month and tested negative.

Transdifferentiation of human B cells into macrophages

To induce transdifferentiation of human leukemic B cells (B cells) into induced macrophages (iMac), 2-5×10^5 BLaER cells were seeded in an appropriate volume. To activate CEBPA, 100 nM 17-beta estradiol (E2) (Sigma Aldrich), human IL-3 (10 ng/ml) (PeproTech) and human M-CSF (10 ng/ml) (PeproTech) were added to the medium to favor the conversion.

Doxycycline treatment

To reduce TET2 mRNA levels, shRNA expression was induced in the shTET2 pLKO-Tet-On transduced stable cell lines by adding 1 µg/ml doxycycline (Sigma) dissolved in deionized water. Fresh working stock was maintained no more than 1 month at -20°C, to avoid thawing and freezing cycles. Doxycycline was re-added to the medium every 3 days during transdifferentiation. Two days pre-treatment was performed to ensure protein depletion prior transdifferentiation induction.

Surface markers profiling by flow cytometry

Transdifferentiation efficiency was evaluated through cell surface antigens flow cytometry analysis. Briefly, after two PBS washes, 0.05–0.1 × 10^6 cells were blocked with human Fc Receptor binding inhibitor (eBiosciences Cat #16916173) for 10 minutes, followed by addition of directly conjugated CD11b (CD11b-APC BD #550019) and CD19 (CD19-PE BD #340364). Antibodies staining was done for 20 minutes at room temperature and protected from light, before a final PBS wash and DAPI (4',6-diamidino-2-phenylindole) staining as a viability marker. Sample acquisition was performed using BD FACSCanto™ II Flow Cytometry System (BD Biosciences, San Diego) and data analysis with FlowJo™ software.

Vectors construction

All of the shRNA oligonucleotides were custom synthesized (Integrated DNA Technologies), annealed in annealing buffer (10 mM Tris pH 7.5, 50 mM NaCl, 1 mM EDTA) at 95°C for 4 min before slow cooling to room temperature. Prior ligation, oligonucleotides were phosphorylated (T4 Polynucleotide Kinase New England Biolabs) according to the manufacturer's recommendations. Oligonucleotide's sequences for vector construction are indicated in **Table 1.**

- pLKO-Tet-On-shTET2

  Generation of TET2 pLKO-Tet-On plasmids was done according to Wiederschain et al 2009. 5 µg of pLKO-Tet-On (Addgene #21915) was double digested with BshTI (AgeI) (Thermo Scientific # FD1464) and EcoRI (Thermo Scientific #FD0274) for 10 minutes at

37ºC. The plasmid backbone fragment was agarose gel isolated (PureLink™ Quick Gel Extraction Kit Invitrogen). The resulting product was de-phosphorilated (Antarctic Phosphatase New England Biolabs) to enhance ligation efficiency. Ligation reaction with corresponding shRNA inserts was performed overnight at 16°C with T4 DNA Ligase (New England Biolabs). Resulting reaction was transformed in DH5α competent cells and plated to grow overnight. Positive bacterial colonies were identified by PCR assay screening and confirmed by Sanger sequencing.

TET2 shRNA target specific sequences were extracted from The RNAi Consortium shRNA Library (Broad Institute, https://www.broadinstitute.org/rnai-consortium/rnai-consortium-shrna-library) based on the highest specificity score and targeting different exons. Scrambled control sequences were duplicated from Wiederschain et al 2009.

- pSicoR-AGO2

pSicoR PGK puro plasmids (Addgene #12084) for constitutive knock-down of AGO2 were cloned following previously described strategy. For backbone isolation the plasmid was double digested with KspAI (HpaI) (Thermo Scientific #FD1034) – XhoI (Thermo Scientific #FD0694) for 15 minutes at 37C.

pSicoOligomaker (v1.5) software (https://web.mit.edu/jacks-lab/protocols/pSico.html) was used for shRNA design following the recommended parameters.

- pMK-TET2-3xFlag

The TET2-3xFlag (DYKDDDDKDYKDDDDK) targeting vectors were cloned by serial modification of base vectors pMK292 (Addgene #72830), pMK293 (Addgene #72831) and pFETCh Donor (Addgene #63934). The designed sequences aim to eliminated the stop codon at the end of the last exon of TET2, through homology-directed repair

(HDR), in order to generate the TET2-3xFlag-P2A-NeoR/HygR tagged protein. Two different plasmids with Neomycin and Hygromycin resistance were generated to allow the selection of cells where both *TET2* alleles are tagged.

Gibson assembly (Gibson Nat Methods 2009) strategy was used to generate the final vector. All components were PCR amplified (Phire Hot Start II DNA Polymerase Thermo Scientific) with overlapping tails of approximately 15 bp. 800 bp homology arms (5'HA and 3'HA) at the last exon of TET2 were amplified from BlaER genomic DNA extracted with Wizard Genomic DNA Purification kit (Promega). pMK backbone, NeoR/HygR and 3xFLAG-P2A elements, were obtained by PCR amplification from pMK293, pMK292 and pFETCh donor plasmids respectively. All fragments were purified with PureLink™ Quick Gel Extraction Kit (Invitrogen). Gibson assembly was done for 1 hour at 50 °C before transforming in DH5α competent cells and plated to grow overnight. Positive bacterial colonies were identified by PCR assay screening and confirmed by Sanger sequencing.

Guide RNAs (gRNAs) targeting TET2 last coding exon were cloned onto px330-mcherry (Addgene #98750) empty backbone following Zhang lab guidelines (https://www.zlab.bio/resources). The design of the gRNAs was done using CRISPR Guide RNA Design Tool (https://www.benchling.com). Alt-R Genome Editing Detection Kit (Integrated DNA Technologies) was used to confirm on-target TET2 effect.

Lentiviral production and cellular transduction

Plasmids for transfection or transduction were isolated using NucleoBond Xtra Midi (MACHEREY-NAGEL) kit according to the manufacturers protocol.

- <u>TET2 pLKO-Tet-On / pSicoR-AGO2</u>

Lentiviral transduction was performed to generate the stable BlaER cell lines. For each plasmid, low passaged HEK293T cells that were seeded at 0.3–0.5 × 10^6 cells/ml and co-transfecting with VSV-G and psPAX2 lentiviral plasmids. Calcium phosphate-mediated transfection protocol was followed. Medium change was performed 12 hours post-transfection and supernatants containing lentiviral particles were harvested and filtered (45 µM) at 48 hours post-transfection. Viral particles were concentrated at 70.000×G for 2 hours at 10°C and resuspended by agitation for 3 hours at 4°C. Then, cells were spin-infected (1000×G) at 32ºC for 90 minutes, to facilitate virus-cell interaction before seeding in an appropriate volume. Twelve hours post-infections cells were washed to remove debris and remaining viral particles.

Two days post-infection, puromycin at 1 µg /ml was added to the medium and maintained for 7 days to select the transduced cells. Medium with puromycin was changed every 3 days.

- pMK-TET2-3xFlag

To generate TET2-3xFlag knock-in, two sequential rounds of transfection were performed to target both alleles. Transfection was carried out by electroporation (Amaxa Nucleofector KitC, Lonza) according to the manufacturer's instructions. One microgram of TET2-3xFlag-P2A-NeoR plasmid was used per 100 µl of Nucleofector solution mix which was used to electroporate 1×10^6 BlaER cells (Nucleofector® II Device, Program: X-001). Next day cells were washed two times with PBS to remove any remaining nucleofection solution. Two days after transfection, Neomycin at 500 µg/ml (Sigma-Aldrich) was added to the medium to select edited cells. Selection medium was changed every 2–3 days until the growth rate of the cells gradually increased to normal rates (every 30-50 hours). Neomycin resistant cells were subjected to the same procedure with TET2-3xFlag-P2A-HygR plasmid, to generate double resistant cell line which were single cell sorted in Flow Cytometer LSRFortessa (BD Bioscience). After 2 weeks of expansion, PCR screening was performed to identify bi-allelically tagged cloned. Briefly, 0.1–0.5 × 10^5 cells were washed with PBS and

incubated in 25 µl of alkaline extraction buffer (0.1% NaOH, 0.1% EDTA) at 95°C for 30 min followed by 4°C for 15min. The resulting reaction was quenched with Tris-HCl and 1-3 µl were used for PCR screening reaction. Primers sequences for screening are indicated in **Table 2**.

Western blots

Up to 10 × 10^6 cells, were washed twice with PBS and incubated with a cell lysis buffer (50mM ph7.5 Tris-HCl, 150mM NaCl, 1mM EDTA, 1mM EGTA, 5mM Mg2Cl, 0.5% Triton X-100) on ice for 1 hour, with tube inversion every 20 min. Whole cell lysate was then centrifuge at 12.000×G for 15 min to remove any cell debris. The clean lysate was quantified by Bradford protein assay. Up to 100 µg of protein were mixed with Laemmli Buffer (2% SDS, 10% Glycine, 5% 2-Mercaptoethanol, 0.0002% Bromophenol blue, 62.5mM pH6.8 Tris-HCl) and boiled at 95 °C for 10 min before loading.

Proteins were separated by electrophoresis in an 8% polyacrylamide gel and transferred to a nitrocellulose blotting membrane at 400mA for 4 hours. Blocking was done with 5% nonfat milk in TBS–Tween (50 mM Tris, 150 mM NaCl, 0.1% Tween 20) for at least 30 min shaking at room temperature. Primary antibody incubation was done a 4°C shaking overnight (anti-TET2, anti-EFTUD2, anti-AGO2) or 2 hours at room temperature (anti-FLAG M2, anti-α/β Tubulin). Three 10 min washes with TBS-Tween were done before proceeding with dye-labeled secondary antibody (REF) incubation in 2.5% nonfat milk TBS–Tween for 1 hour at room temperature and protected from light. Afterwards, membranes were washed five times for 10 min with TBS-Tween and revealed using Odyssey CLx instrument (LI-COR). Primary and secondary antibodies references are listed in **Table 3**.

RNA-seq

Approximately 1–3×10^6 cells were collected at 0, 72 and 168 hours of transdifferention. RNA was extracted with the RNeasy Mini Kit (QIAGEN Cat#74104) according to manufacturer's instructions and quantified with NanoDrop spectrophotometer. Briefly, quality control with Agilent 2100 was performed prior library preparation to ensure RNA integrity (RIN>7). 0.2-0.5 ug of RNA was used for library preparation. rRNA removal was done using the

RNase H. Then, libraries were sequenced in a DNABSEQ-G400 sequencer using a pair-end 150-bp protocol. More than 50 million reads were obtained for each sequenced sample.

Real-time quantitative Polymerase Chain Reaction (RT-qPCR)

Total RNA was extracted using Trizol (eBioscences). 500 ng of total RNA were converted into cDNA using the RNA to cDNA kit (Applied Biosystems) following the manufacturer's instructions. Real-time quantitative PCR reactions were performed using SYBR Green reagent and analyzed using QuantStudio 5 System (Applied Biosystems). *HPRT1* was used as housekeeping genes. Unpaired student's t-test was used to determine statistical differences in gene expression among the different samples tested (t-test, ***p<0.001). Normality and homogeneity in variance were assumed for RT-qPCR experiments with biological triplicates. Primer sequences are listed in **Table 4**.

ChIP-seq

Dual cross-linking ChIP protocol was followed as previously described [299]. Approximately 50×10^6 cells were collected at 72 and 168 hours of transdifferention and washed twice in ice-cold PBS. Fresh stock solution of 0.25M disuccinimidyl glutarate (DSG) (Fisher Scientific) in DMSO was prepared and added to a final concentration of 2mM in ice-cold PBS. Cell pellets were resuspended in 10 ml of DSG crosslinking solution and incubated rotating at room temperature for 30 min. Next, formaldehyde (Sigma) was added to a final concentration of 1% and samples were rotated for additional 10 minutes at room temperature. To stop fixation, freshly prepared glycine was added to a final concentration of 0.125M and mixed well before proceeding.

Crosslinked cells were washed twice in ice-cold PBS and resuspended in cold IP buffer (1 volume SDS buffer (100mM NaCl, 50mM pH8.1 Tris-HCl, 5mM pH8 EDTA, 0.2% NaN3, 0.5% SDS) : 0.5 volume Triton dilution buffer (100mM NaCl, 100mM pH8.6 Tris-HCl, 5mM pH8 EDTA, 0.2% NaN3, 5% Triton X-100) supplemented with proteinase inhibitors (Roche Cat #118733580001). Chromatin was sheared to 100-300bp fragments with Bioruptor Pico Sonicator (Diagenode) at 4°C for 13 cycles of 30s on / 30s off in 15 ml Bioruptor Pico

Tubes (Diagenode Cat #C30010017). The sonicated chromatin was spun down at 20.000×G for 20 min at 4ºC to remove any insoluble portion and the supernatant was pre-cleared with 20 ul of Dynabead A/G mix (Invitrogen). Chromatin was conjugated with 5 µg of Monoclonal anti-FLAG M2 antibody (Sigma) rotating overnight at 4ºC. Next day, 50 µl of Dynabead A/G mix were blocked with BSA (5 mg/ml) rotating for 2h at 4ºC and added to the sample for immunoprecipitation for 2h at 4ºC in rotation.

Chromatin-antibody-bead complexes were washed three times with ice-cold low salt buffer (50mM pH7.5 HEPES, 140mM NaCl, 1% Trito X-100) and one time with ice-cold high salt buffer (50mM pH7.5 HEPES, 500mM NaCl, 1% Triton X-100). After the last wash, the complexes were de-crosslinked in elution buffer (1% SDS, 0.1M NaHCO3) by overnight incubation at 65ºC with shaking at 1300 rpm. Next day, the eluted portion was treated with RNase A (20 ng/ml) for 1 hour at 37ºC followed by proteinase K (200 ng/ml) for 2 hours at 65ºC. Finally, DNA was purified by phenol:chloroform:isoamyl alcohol (25:24:1) extraction.

For ChIP-qPCR analysis, DNA was diluted 1:20 and relative enrichment was calculated with the following formula (100*2^(Adjusted input - Ct (IP)). Oligonucleotide sequences are indicated in **Table 4**.

For ChIP-seq, samples were quantified with Agilent 2100 before library preparation. Library preparation and sequencing were performed by the sequencing service provider using a DNABSEQ-G400 sequencer and a SE50 protocol.

Co-Immunoprecipitation (Co-IP) / Mass spectrophotometry (IP-MS)

Whole protein extracts from 20×10^6 cells were used for each immunoprecipitation. Fresh cells were washed twice in PBS and incubated with cell lysis buffer (50mM ph7.5 Tris-HCl, 150mM NaCl, 1mM EDTA, 1m EGTA, 5mM Mg2Cl, 0.5% Triton X-100) for 1 hour on ice with tube inversion every 20 min. Whole cell lysate was then centrifuge at 12.000×G for 15 min to remove any cell debris. Proteins were immunoprecipitated with 50 µl of ANTI-FLAG M2 Magnetic Beads M8823 (Sigma) rotating overnight at 4ºC. Next day, beads were washed six times with ice-cold cell lysis buffer and proceeded with western blot

co-immunoprecipitation analysis or mass spectrometry pipeline. For western blots, beads were resuspended in Laemmli Buffer (2% SDS, 10% Glycine, 5% 2-Mercaptoethanol, 0.0002% Bromophenol blue, 62.5mM pH6.8 Tris-HCl) and boiled at 95ºC for 10 min. Beads were magnetically separated from the eluted protein extract which was loaded directly into gel (see: Western blots).

For mass spectrometry analysis, after the last wash, beads were resuspended with Tris-Urea Buffer (Tris-HCl 6M, Urea 100mM) followed by a reduction step with dithiothreitol (DTT) and alkylation with chloroacetamide (CAA). The digestion of the proteins was done with trypsin at 30°C overnight. Next, the samples were purified with C18 columns before injecting into Orbitrap Fusion Lumos Tribrid Mass Spectrometer (Thermo Scientific). Raw data was analyzed with MaxQuant software. The Contaminant Repository for Affinity Purification (CRAPome) database was used to eliminate false positive and non-specific binding proteins. Statistically significant interactors were identified using the Significance Analysis of INTeractome (SAINT) statistical method (Teo J Proteomics 2014). Raw and processed files were provided by the Proteomic Unit of the Josep Carreras Research Institute (Barcelona).

Whole genome bisulfite sequencing (WGBS)

Genomic DNA was extracted from 1 million cells at 0, 24, 96, and 168 hours of transdifferentiation using the DNeasy Blood & Tissue kit (QIAGEN Cat #69504) following the manufacturer's instructions and quantified using Qubit dsDNA (Invitrogen Cat #Q32851). Cytosine conversion, library preparation and sequencing were done by the provider of the sequencing services. Briefly, genomic DNA was fragmented to 200-400 bp and bisulfite treated. For library construction, sequencing adapters were ligated, followed by double-strand DNA synthesis and PCR amplification. Next, libraries were sequenced on Illumina HiSeqTM2500 using a pair-end 150-bp protocol rendering >70Gb/sample. Raw data quality assessment was performed, and low-quality reads were trimmed.

Bisulfite conversion (BS) and pyrosequencing

Sample DNA methylation status was assessed by bisulfite pyrosequencing. Briefly, 1 ug of genomic DNA was bisulfite (BS)-converted using the EZ DNA Methylation Gold Kit (Zymo Research Cat# D5006) following the manufacturer's

instructions. BS-treated DNA was PCR-amplified using the IMMOLASE DNA polymerase Kit (Bioline). Primers used for the PCR were designed with PyroMark Assay Design 2.0 software (QIAGEN) and listed in **Table 5**.

PCR products were pyrosequenced with the Pyromark Q48 system (QIAGEN), according to the manufacturer's protocol and analyzed with PyroMark Q48 Autoprep (QIAGEN).

DNA methylation arrays

For DNA methylation assessment during TET2 depletion Infinium MethylationEPIC v1.0 Bead-Chip arrays were used. This platform allows the interrogation of around 935,000 CpG sites per sample at single-nucleotide resolution, covering 99% of the reference sequence (RefSeq) genes. Briefly, 1–3 × 10^6 cells were collected at 0, 72 and 168 hours of transdifferention with 3 biological replicates for each group. Genomic DNA was extracted with Wizard Genomic DNA Purification kit (Promega) according to manufacturer's protocol and quantified with NanoDrop spectrophotometer. Bisulfite conversion was done as previously described. 100-500 ng of bisulphite converted DNA was hybridized according to Illumina protocol. Raw files (IDAT files) were provided by the Genomics Unit of the Josep Carreras Research Institute (Barcelona).

Bioinformatic analyses

All sequencing data obtained were mapped onto the human genome assembly hg38 (Ensembl GRCh38) and analyzed with R (4.2.1) using packages from the Bioconductor suite (v3.0) [357]. For peak calling, regions overlapping the 'Encode blacklist' regions were removed (The ENCODE Project Consortium Nature 2012), as well as mitochondrial reads. Peaks were annotated to genomic features in R with the package ChIPseeker (v1.32.1) [358], using Benjamini-Hochberg (BH) FDR corrections. All Gene Ontology (GO) enrichment analyses were performed using the clusterProfiler package (v4.4.4) [359] and, when integrated with expression data, plotted with plotGODESeq package (https://ascistance.co.uk/plotGODESeq/README.html). Bigwig tracks were generated using deepTools BamCoverage (v3.3.1) [360]. The integration of different datasets was carried out using the bedtools package (v2.31.1) [361].

Heatmaps and clustering analyses were performed using the ComplexHeatmap (v2.10.0) [362] and deepTools computeMatrix (v3.3.1) [363] packages. The Multidimensional Scaling (MDS) and Principal Component Analysis (PCA) were performed using the limma (v3.52.2) [364], and factoextra packages (v1.0.7), respectively. Upset plots were generated using the upsetR package (v1.4.0) [365]. The RNA-seq MA plots were generated using ggmaplot function from ggpubr R package (v0.6.0) (https://rpkgs.datanovia.com/ggpubr/). The remaining plots were generated using the R package ggplot2 (v3.4.2).

- **Bisulfite-seq analysis**

  Raw sequence reads from WBGS libraries were trimmed to remove poor-quality reads and adapter contamination using the package Trimmomatic (v0.36) [366]. The remaining sequences were mapped using Bismark (v0.16.3) [367], to the human reference genome GRCh38 in paired-end mode. Reads were then deduplicated, and CpG methylation calls were extracted from the deduplicated mapping output using the Bismark methylation extractor in paired-end mode. The estimation of DNA methylation levels was calculated by dividing the count of reads identifying a cytosine (C) by the sum of reads identifying either a cytosine (C) or a thymine (T). Only CpGs with coverage equal to or higher than 5X in all tested samples were used for downstream analyses. N = 24,298,869 CpGs analyzed.

- **Analysis of DNA methylation arrays.**

  IDAT raw data from the MethylationEPIC BeadChip 850k v2.0 microarrays was loaded using R to perform all the analyses, QC, and preprocessing steps using the package minfi (v1.42.0) [368]. Probes with low detection p-value (< 0.1), probes with a known SNP at the CpG site, and known cross-reactive probes were removed. For the resulting CpGs, beta values were calculated using minfi functions. DMPs were calculated using the limma package (v3.52.2) [364] in R, adjusting by the Benjamini-Hochberg method. Only DMPs with p-values < 0.05 and a differential of DNA methylation equal to or greater than 10% ($\Delta\beta \geq 0.1$) were selected for further analyses.

- **RNA-seq analysis**

  Reads were mapped using STAR (v2.7.6) [369]. Gene expression was quantified using the function featureCounts from the package Subread (v2.0.3) [370]. Differentially expressed genes were detected using the R package DESeq2 (v1.36.0) [371], applying q-value < 0.05 filtering.

- **ChIP-seq analysis**

  For the ChIP-seq analyses, reads were trimmed using the TrimGalore package (v0.6.6) to remove adaptors and mapped using Bowtie2 (v2.4.4.1) [372]. Duplicated reads were removed with the tool MarkDuplicates from the package Picard (v3.1.0) (Picard Tools https://broadinstitute.github.io/picard). Peaks were called using MACS2 (v2.2.5), parameter (-q 0.05) [373]. Biological replicates were handled with Irreproducibility Discovery Rate (IDR) method, using the package idr (v1.3) [374]. Peaks with an IDR < 0.05 were selected for the downstream analysis.

- **Splicing analysis**

  To evaluate the influence of TET2 depletion in splicing events the Vertebrate Alternative Splicing and Transcription Tools (VAST-TOOLS) was used [305] (vast-tools v2.5.1). Briefly, raw RNA-seq data was aligned to Human (hg38, Hs2) VastDB annotation. The replicates output was merged and combined according to package manual for the downstream analysis. Differential splicing analysis was performed using the 'compare' option in order to contrast PSIs (percent-splice-in) between samples. PSIs with adjusted p-values < 0.05 were used for downstream analysis. Splicing events were grouped intro 3 categories: exon skipping - EX (MIC, S, C1, C2, C3, ANN), alternative splice donor/acceptor – ALT (Alt3, Alt5) and intron retention – IR.

- **Identification of dynamic regulatory regions**

  Confident TET2 chromatin binding sites (IDR<0.05) coordinates were used for evaluation of the epigenomic context in the downstream analysis.

  Intersecting of ATAC-seq peaks with H3K4me1 and transcription start sites (TSS) was done as previously described [289]. The evaluation of H3K27ac signal was done using the Diffbind R package (v.2.2.12). H3K27ac loss, gain or stability was defined comparing 168 hpi vs 0 hpi, where differences were calculated using the DBA_DESEQ2 method and -filter for fold change > 2 and FDR < 0.05 significance. To define transient gain regions 0 hpi vs 72 hpi vs 168 hpi timepoints were considered. Same strategy as before was used with filtering for fold change > 1 in 72 hpi vs 0 hpi comparison and fold change < 1 in 168 hpi vs 72 hpi. Expression of the associated genes in each category was evaluated and duplicated genes were excluded from the analysis.

  Methylation steady (MSRs) and dynamic regions (MDRs) were defined by comparing 168 hpi vs 0 hpi average methylation levels (Python numpy.nanmean function) (v2.1) within the defined TET2 binding sites . Only confident CpGs methylation (5X coverage) were considered (see Bisulfite-seq analysis) and changes in methylation equal to or greater than 20% ($\Delta \geq 0.2$) were classified as dynamic. Same quantification method was used to evaluate the methylation levels within the classified TET2 binding sites GREs.

- **DNA motif analysis**

  DNA motif analysis of the MSR vs MDR peaks were analyzed using MEME Suite - motif-based sequence analysis tool [375]. Simple Enrichment Analysis (v5.5.7) tool was used (https://meme-suite.org/meme/tools/sea) with shuffled input sequences as the background. Motif enrichment was filtered for absolute log2FC enrichment >1 and qvalue<0.05.

- **Hi-C analysis**

  Raw Hi-C interactions at 0, 72 and 168 hours of transdifferention were processed with the HiC-Pro (v3.1.0) pipeline following the recommended parameters with ICE normalization [376]. Normalized contact maps at 5kb were used for data visualization in (https://epigenomegateway.wustl.edu/). Contacts with a score ≥1 were considered significant.

  Long-range interactions between TET2 binding sites were computed using the HiCExplorer tool hicAggregateContacts. Hi-C matrices were generated at 10-kb resolution using HiCExplorer and long-range interactions (5–10 Mb) between the provided coordinates [377].

Chapter 2 |

Results

(V) Generation of experimental cell lines models

To tackle the project's main objective, we have generated several new BlaER-derived cell lines. As previously described, this mainly encompassed the establishment of inducible *TET2* depletion models and bi-allelic knock-in tagging of endogenous *TET2*. The establishment of these lines has ensured a proper ground for evaluating the proposed questions in a reproducible manner without intrinsically perturbing the cell's biology. This is of upmost importance since we aimed to interrogate a dynamic cell system where a complete chromatin rewiring occurs within a relatively short time (6-7 days). Hence, any perturbations, such as transfection and/or transduction procedures and antibiotic selection, prior TD, can lead to impaired physiology unrelated to the experimental question.

Regardless, this approach has allowed us to interrogate our hypothesis in a biologically reproducible manner and validate our findings in a homogeneous way. The following sections describe in detail the strategies used to generate and validate the aforementioned cell lines.

5.1 TET2 depletion strategies, validations and troubleshooting

5.1.1 Inducible TET2 depletion model generation

To precisely modulate *TET2* levels, we took advantage of a pLKO-Tet-On system, a doxycycline-inducible lentiviral RNAi depletion model. This system enabled us to generate stable and inducible cell lines, allowing us to deplete *TET2* in a timely and reproducible manner [298]. We have generated 3 doxycycline-inducible pLKO-Tet-On shRNA cell lines targeting *TET2* 3'UTR and

CDS regions (shT2.1, shT2.2, shT2.3). Scrambled sequence (shCtrl) shRNA was used as a control in all the experiments. The shRNA targeting sequences were designed according to Broad Institute RNAi Consortium guidelines and targeted a previously described region at the exon 3 of *TET2* (targeted by shT2.1 and shT2.3). This exon is one of the largest on the gene, encoding the structurally disordered N-terminal TET2 region (1-1128aa) (**Fig. I4**) present in all *TET2* isoforms (TET2-a/b/c). Importantly, the region has been reported as a hotspot for mutations in myeloid malignancies, including nonsense, insertion and deletion, and premature stop codons, which results in loss of function truncating proteins [215,218,257].

Functional validation was done after 48 hours of doxycycline (Dox) treatment in un-induced B-cells. The treated cells showed no phenotypic abnormalities or cell growth alterations, suggesting stable working conditions and a lack of TET2 influence in the B-cell state (data not shown). Consequently, *TET2* levels were determined by RT-qPCR and showed variable mRNA depletion among the cell lines tested. The values were normalized to shCtrl which was also subjected to the doxycycline treatment in parallel. While shT2.1 and shT2.3 showed approximately 50% *TET2 mRNA* decrease, shT2.2 only led to a 40% reduction (**Fig. R1a**). Although we have achieved a consistent downregulation among the generated cell lines, we aimed to enhance the downregulation effect.

To improve the efficiency, we have attempted several strategies targeting different system components:

S1. First, we have attempted to generate novel cell lines following increased multiplicity of infection (MOI) protocols, which, in theory, should provide an increased number of siRNA targeting molecules. However, RT-qPCR validations did not show any improvements, suggesting limitations in other components of the system.

S2. Similarly, we attempted to use commercially available MISSION shRNA Vectors (Sigma) inducible with isopropyl-ß-D-thio-galactoside (IPTG). The newly generated cell lines showed a scarce 10% *TET2* reduction in the 10-100 µM IPTG treatment range. The system was abandoned after several unsuccessful experiments.

S3. Additionally, we have attempted to modify the doxycycline treatment since activation with dosages as low as 10 ng/ml was

previously reported [298]. Our evaluation showed a consistently similar reduction in the 0.2-5 µg/ml doxycycline range, with no apparent deviations from the previously observed 50% reduction in shT2.1. Higher dosages (10 µg/ml) led to cell death and completely abnormal *TET2* levels (**Fig. R1b**). Consequentially, the working concentrations were maintained at 1 µg/ml for future experiments.

S4. Finally, we have engineered a shT2.1-HygR plasmid by replacing the antibiotic resistance from the original backbone. The idea was to perform cross-infection of already established lines (shT2.1-PuroR + shT2.1-HygR) and faithfully select a double-resistant population. Although the line was successfully established after two weeks of hygromycin selection, the functional validations, once again, have not shown any significant improvement in downregulating efficiencies.

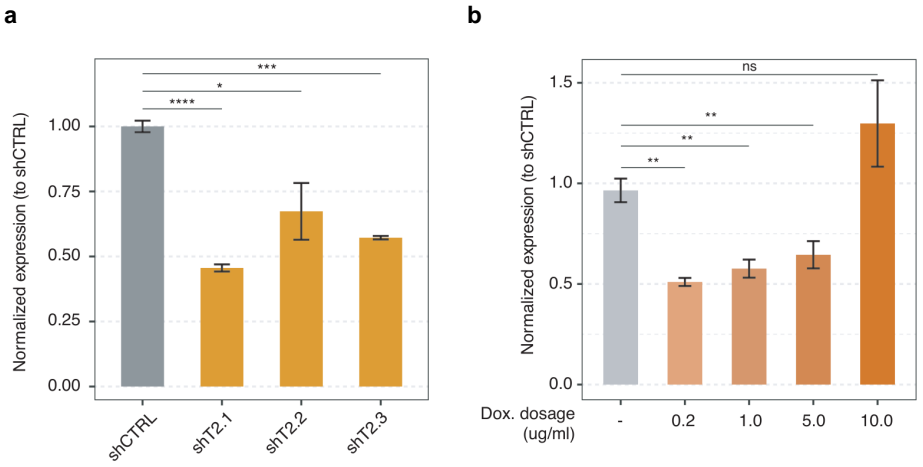

Figure R1. Quantification of *TET2* expression levels by RT-qPCR after Dox treatment. (a) Normalized *TET2* reduction efficiencies between different shRNAs. Normalized *TET2* reduction levels at different Dox dosages. Unpaired two-tailed Student's t-test, n=3, mean s.e.m, (*p<0.05; **p<0.01, ***p<0.001, ****p<0.0001).

Regardless, a consistent 50% mRNA *TET2* reduction in shT2.1 showed promising results for further evaluation, particularly since moderate *TET2* depletion might better phenocopy *in vivo* events in myeloid malignancies

patients and individuals with CHIP, which mostly present *TET2* haploinsufficiency [221].

5.1.2 TET2 depletion effect evaluation on TD efficiency

To better characterize the impact of *TET2* reduction during TD, we have focused on two lines, shT2.1 and shT2.3, which showed a similar 50% downregulation in our preliminary characterization (**Fig R4a**). To assess whether the *TET2* reduction impacts cell fate conversion through C/EBPα induced TD, we monitored the expression of surface markers by flow cytometry. Specifically, we have evaluated the CD19 B-cell and the CD11b macrophage markers. To that extent, we have treated shT2.1, shT2.3, and shCTRL BlaER cells with doxycycline for 48 hours prior to TD and analyzed the cells at initial B-cell state, 72 and 168 hours post induction (hpi), when most of the population have acquired the myeloid identity. Notably, Pre-treatments ensure a steady *TET2* depletion prior to cell induction, which should contribute to maintaining reduced *TET2* levels throughout the process.

Interestingly, both control and TET2 knock-down cells converted into induced macrophages in an expected manner. First, they lost the CD19 marker and then gained the CD11b marker. However, some slightly accelerated kinetics were observed for the shT2.3 at the 72-hour timepoint, while the control and the shT2.1 cells were nearly identical (**Fig. R2**). These findings suggest that a roughly 50% *TET2* depletion neither blocks nor delays TD to a great extent, as suggested by the analyzed cell surface markers. However, based on the very limited information provided by these analyses, we decided to extensively characterize the methylome and transcriptome landscape of the *TET2*-depleted cells. To address this issue, we have focused on the shT2.1 cell line, which has shown up to 50% *TET2* reduction by RT-qPCR and presented no adverse effects to doxycycline treatment. The follow-up experiments, described in the next sections (see: 6.3), provide a broad overview regarding the presented questions.

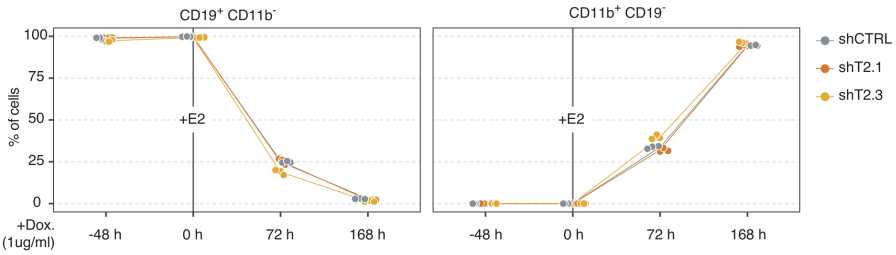

Figure R2. Flow cytometry analysis during TD and Dox treatment. Percentages of CD19+/CD11b- cells (left) or CD11b+/CD19- (right) are shown. n = 3 biologically replicates.

5.2 Identification of TET2 chromatin targets and partner proteins

Our attempts to capture the endogenous TET2 at +168 hpi had limited success, even with previously reported antibodies [299]. Which is most likely attributed to relatively low TET2 levels and/or lack of antibody affinity.

So, to circumvent this issue, it was necessary to provide a more robust system to thoroughly interrogate the TET2 interactome and chromatin distribution during myeloid cell fate commitment

5.2.1 Evaluation of TET2 immunoprecipitation strategies

Previous studies had great success capturing TET2-tagged protein in mouse ESCs. Hence, we decided to look into potential epitope knock-ins to generate a fused Tet2 protein for efficient immunoprecipitation in our cells.

We first determined which epitope-capturing strategy would more closely resemble the endogenous Tet2 distribution, as non-specific bindings and increased background noise have been well documented in tagged protein-capturing protocols.

To address this, we took advantage of three studies. Firstly, endogenous capturing in WT and KO conditions was done with a non-commercially available antibody generated in Pocono Rabbit Farm and Laboratory and was used as an

endogenous control [300]. Secondly, two independent studies have engineered 2xFLAG or FLAG-biotin tags into the endogenous C-terminus of Tet2, and have successfully mapped the chromatin localization of the protein [181,295].

Having analyzed the data, we have detected confident Tet2 peaks at the previously described *Oct4* locus [301], which was completely absent in Tet2 knockout conditions (**Fig. R3a**). Notably, we detected similar binding profiles between endogenous Tet2 and 2xFLAG immunoprecipitation, although presenting the expected increase in the background signal. FLAG-biotin dataset, presented enrichment at the expected regions, but showed an overall inferior signal. Further evaluation of total detected peaks showed that 2xFLAG shared almost 80% of regions with the endogenous conditions and although similar to the FLAG-biotin (83%), had a much higher number of total peaks (**Fig. R3b**).

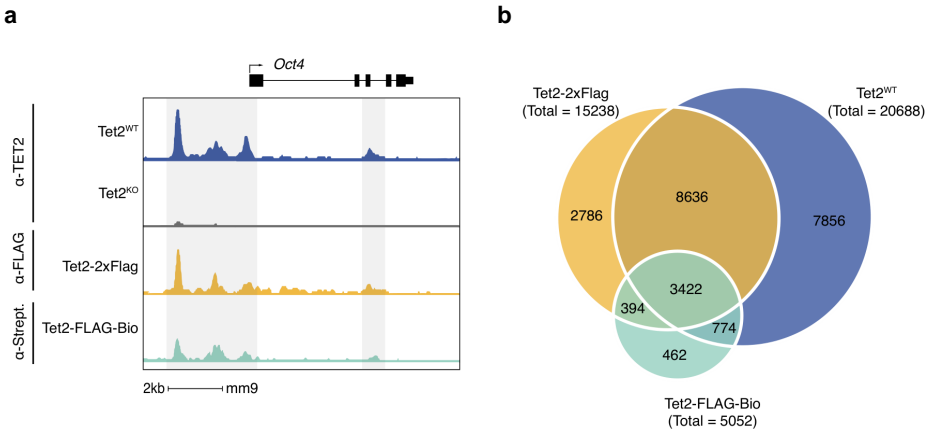

Figure R3. Comparison of Tet2 immunoprecipitation strategies in previous ESCs studies. (a) Snapshot of the *Oct4* locus showing the efficiencies of Tet2 capturing with different strategies. (b) Venn's diagram showing the total number of Tet2 peaks and their overlap between different capturing strategies. Strept: M-280 Streptavidin beads

In summary, 2xFLAG epitope capturing highly assimilated to endogenous Tet2 immunoprecipitation. Accordingly, this protein tagging strategy was used for cell line generation and downstream capturing experiments.

5.2.2 TET2-tag cell line generation and validations

To further improve the immunoprecipitation chances, we decided to use a ChIP-verified 3xFlag epitope. The triple tag should increase antibody affinity, allowing us to efficiently perform both ChIP-seq and IP-MS experiments. Despite its relatively elevated molecular weight of 2.7 kDa (22 aa) and potential interference of negatively and positively charged amino acids, we have not detected any significant changes affecting protein stability. This is most likely due to the sheer size of TET2, which stands at 224 kDa (2002 aa).

Briefly, to generate the fused protein, we introduced the 3xFlag tag at the last coding TET2 exon through CRISPR homology-directed repair mechanisms. The knock-in was designed to eliminate the stop codon and introduce a Hygromycin or Neomycin antibiotic resistance separated from the FLAG tag by a self-cleaving P2A sequence. After antibiotic selection, double-resistant cells were single-cell sorted, allowed to grow as clones, and consequentially genotyped by PCR screening. Clones with biallelic insertions were selected (**Fig. R4**)

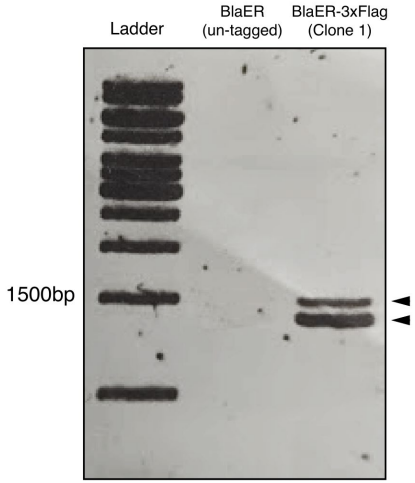

Figure R4. PCR amplification of the TET2 tagged locus. Arrow highlighted amplicons correspond to 3xFlag-HygR (~1500bp) and 3xFlag-NeoR (~1300bp) tagged alleles for one of the bi-allelically positive clones.

Once stable cell lines we established, we have assessed the capturing efficiency by IP-WB. After 72 hpi, to allow for TET2 accumulation, two positive clones (C1

and C2), un-tagged BlaER cells (negative control) and TET2-Flag ESCs (positive control) were analyzed.

To better understand the changes in immunoprecipitation efficiency, we have directly compared anti-TET2 vs anti-3xFlag capturing from the same protein extract. Accordingly, we have detected TET2 in all conditions and the Flag epitope in the tagged ESCs, as well as in both candidate clones, which showed similar signal strength. Nonetheless, we have observed a remarkably stronger signal when immunoprecipitating the 3xFlag, while endogenous TET2 capturing barely showed any band (**Fig. R5**). These results strongly suggested a competent TET2-3xFlag capturing, which could be coupled with high-throughput analysis. Moving forward, Clone 1 (C1) was carefully expanded, re-evaluated and used for future experiments.

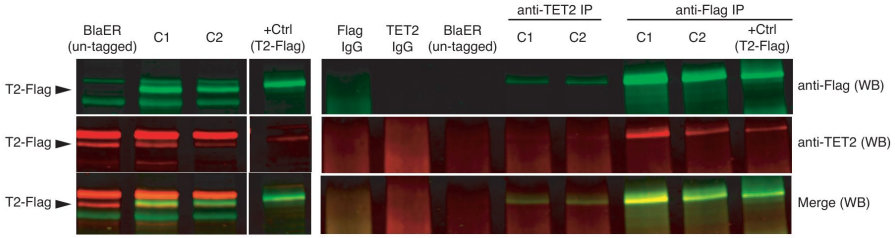

Figure R5. Western blot showing the immunoprecipitation efficiencies comparison between endogenous and 3xFlag tagged TET2.

§

Of note, the generation of stable cell models to fulfill the proposed objectives has encompassed an extensive part of this thesis work. Aside from the presented results, many alternative approaches have been attempted to faithfully modulate TET2 levels. For instance, very extensive work has been done regarding the efforts to establish an auxin-inducible degron system, which would allow for a rapid, precise and direct TET2 depletion. That could be used to directly assess the protein influence at different stages of TD based on quick inducible degradation. We have actually managed to establish several degron cell lines, circumventing complex plasmid assembly and very low knock-in efficiencies. However, validation experiments yielded no results, leading to the abandoning of this line of research. Similarly, chromatin immunoprecipitation experiments took an extensive amount of time to optimize crosslinking and protein extraction efficiencies, leading to some failed sequencing attempts even with efficient tag-capturing strategies.

Regardless, the presented models provided an invaluable resource and have been taken full advantage to confidently interrogate TET2 activity and its role in epi/genome dynamics during major cell fate rewiring.

§

(VI) TET2 chromatin and interactors profiling

To establish a confident TET2 activity map, we first pursued the evaluation of *TET2* binding sites and interacting proteins, which should provide us with the protein's distribution within the gene regulatory regions. Together with the methylation evaluation at the target sites (through the WGBS dataset), this would allow for a better understanding of TET2's direct influence on the transcriptome of the associated genes.

As previously established, we used the bi-allelically tagged TET2-3xFlag BlaER cell line for the immunoprecipitation protocols, which were used for all experiments in the following sections.

6.1 TET2 interactome characterization during TD

To identify TET2's interacting partners during TD, a TET2-FLAG immunoprecipitation coupled with a high-throughput mass spectrometry (IP-MS) strategy was used. Cells were induced to allow for TET2 accumulation and analyzed at 72 and 168 hpi. The IP-MS experiments were performed in biological triplicates. An additional replicate was used to validate the capturing efficiency by WB before proceeding with the mass spectrometer pipeline. Un-tagged BlaER cells were induced in parallel and used as controls for unspecific binding.

6.1.1 Identification of TET2 partners during cell conversion

Firstly, MS analysis at 72 hpi returned almost no significant results. TET2 was strongly enriched compared to control samples, but very few protein partners were detected (data not shown). The only interactors were Sin3A Associated Protein 18 (SAP18) and Tyrosine 3-Monooxygenase/Tryptophan 5-Monooxygenase Activation (YWHAQ). These were involved in relevant biological processes such as the previously described TET2-HDACs interaction in the case of SAP18, which is a component of the Sin3 histone deacetylase complexes. Suggesting potential TET2 implications in transcriptional repressing activities [302].

Overall, the lack of detected peptides suggested either an insufficient *TET2* accumulation after 72 hpi to pass the interactome detection threshold or simply a lack of biological interaction at that stage of TD. Regardless, the main focus was shifted towards later time points.

Interactome analysis in iMac (168 hpi) showed a better output, leading to the identification of 23 TET2 confident interactors by applying a stringent Bayesian False Discovery Rate correction (BFDR) (**Fig. R6a**). Notably, the interactors SAP18 and the YWHAQ protein, which were detected at 72 hpi, were also detected at 168 hpi. This highlights the reproducibility of our IP-MS experiments and emphasizes the potential significance of the TET2-SAP18 interaction in myeloid differentiation, especially since no other SAP members or Sin3a were found.

To further dissect the obtained results, we performed Gene Ontology analysis on the significant interactors which showed a very strong enrichment for mRNA binding and processing proteins. These included several members of the SRSF Splicing Factors family, THOC6, SON and EFTUD2, and several members of the hnRNP RNA-binding family of proteins such as hnRNPL. These findings strongly suggest a novel potential involvement of TET2 in mRNA post-transcriptional modulation, such as splicing events or RNA regulating mechanisms like nonsense-mediated mRNA decay (**Fig. R6b**).

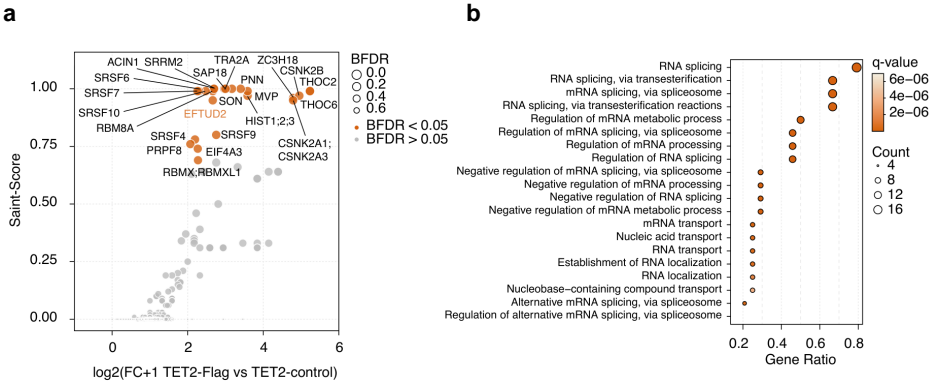

Figure R6. Significant TET2 interactors detected by IP-MS. (a) Scatter plot of detected TET2 interactors at iMac stage. Significantly enriched candidates are highlighted in

orange. (b) Gene Ontology analysis of Biological Processes (BP) significantly enriched for the TET2 interactors at the iMac stage

6.1.2 TET2 RNA splicing machinery partners validation

To confirm our findings, we performed validations by co-immunoprecipitation. Taking into account the availability of commercial antibodies, we have focused on EFTUD2, which is a GTPase spliceosome that participates in catalytic splicing and post-splicing complex disassembly. Functionally, it has been reported to maintain the survival of tumor cells and play a role in innate immune response through regulation of interferon-stimulated genes, particularly during the hepatitis C virus infection. Which might be of relevance for macrophage mediated immune response [303].

Following the same IP-MS experimental conditions, we have successfully validated TET2 interaction with EFTUD2 in two independent Flag-tag clones by co-immunoprecipitation and western blotting after 168 hours of induction (**Fig. R7**).

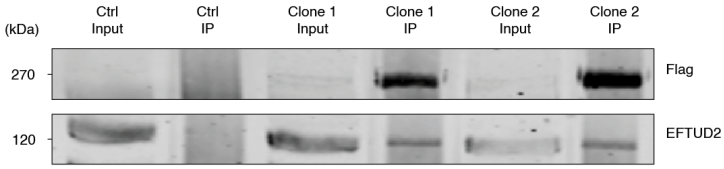

Figure R7. Western blot showing coimmunoprecipitation of TET2-Flag with ETFUD2 at iMac stage.

6.1.3 Evaluation of TET2 role in alternative splicing

To determine TET2's influence on mRNA processing mechanisms, we used our highly-covered RNA-seq dataset (50M reads/sample) collected in TET2-depleted TD conditions (see: 6.3.2). We used VASTtools [304,305] to perform a splicing analysis and observed 619 significantly different percent-splice-in (PSI) events between TET2 normal and TET2 depleted cells after 168 hours of induction (**Fig. R8a**). These included oncogenes such as VAV2, regulators of cytokine production such as PD-L2, or crucial components of the PRC2 complex, including EZH2.

The most predominant PSI type of events observed was related to intron retention (IR: 70.4%) followed by cassette exon skipping (EX: 19.8%) and alternative splice donors/acceptors (ALT: 9.6%) events (**Fig. R8b top**). Notably, quantification of PSIs indicated an increased IR in TET2 depleted condition (**Fig. R8b bottom**) suggesting a novel TET2 implication in mRNA maturation by facilitating intron removal.

Gene Ontology analysis of genes with aberrant IR revealed mildly significant enrichment for terms related to specific immune processes, such as Fc-gamma receptor function in immune processes and phagocytosis, as well as others, such as cell cycle regulation (**Fig. R8c**).

Collectively, our data indicate a potentially novel function for TET2 in modulating mRNA processing through its interaction with the splicing machinery. Thus, highlighting the role of this protein not solely as an epigenetic factor regulating transcriptional activity but also as a regulator of post-transcriptional processes

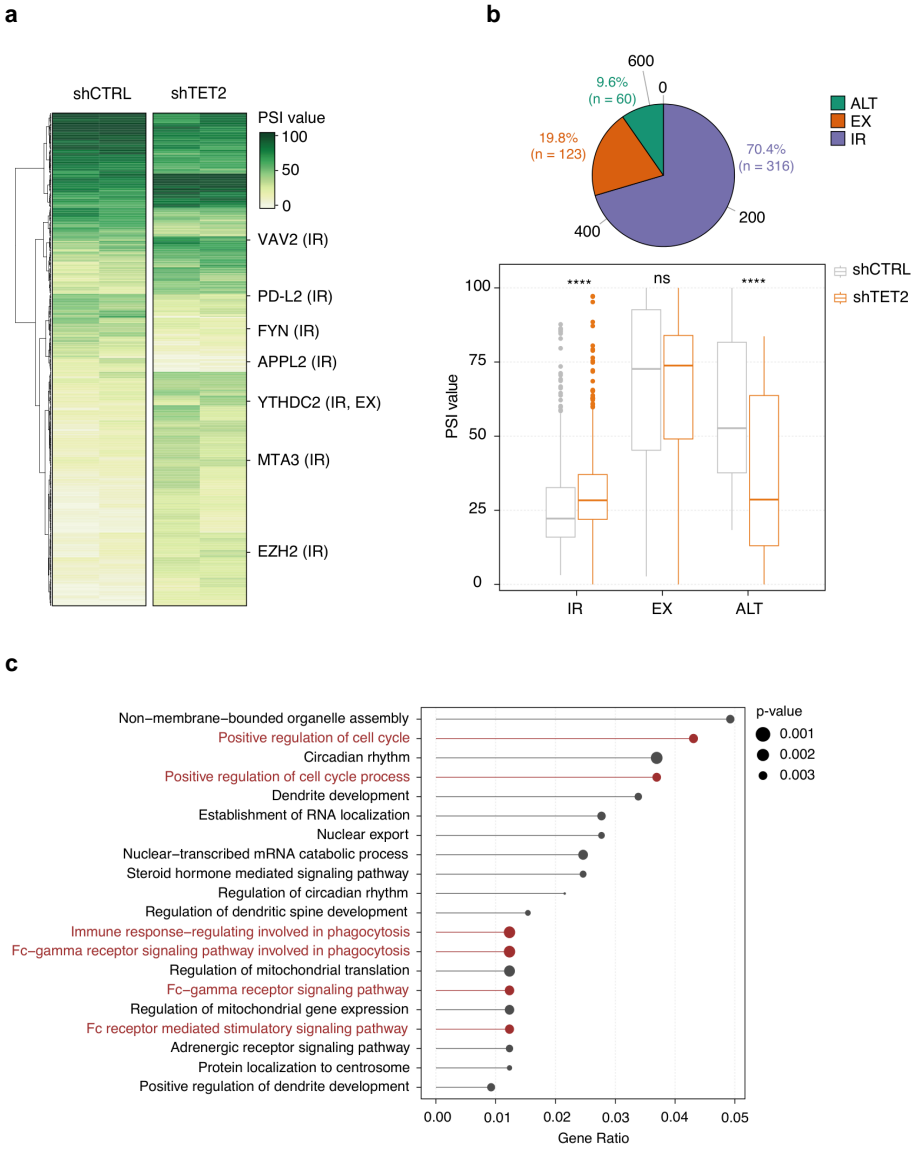

Figure R8. Evaluation of PSI alternative splicing (AS) events in iMac with reduced TET2 levels. (a) Heatmap of differentially spliced transcripts in TET2 depleted conditions at the iMac stage. Two biological replicates for each condition are shown. Biologically relevant genes are depicted on the right. IR: Intron Retention, EX: exon skipping, ALT: alternative splice donor and acceptor. (b) Top: pie chart depicting the number and percentage of

differentially spliced events of each type at iMac stage. Bottom: average PSI levels for each type of splicing event (shCTRL vs shTET2 two-sided Wilcoxon rank-sum test; ns>0.05, **** p-value<0.0001). (c) BP gene ontology analysis of differential IR genes at iMac stage. Highlighted in red are the most biologically relevant terms.

6.2 Identification of TET2 chromatin targets during TD

To fully uncover the TET2 chromatin binding landscape, we used a previously described dual crosslinking protocol [306]. It mainly consists of a combination of a protein-protein crosslinker, disuccinimidyl glutarate (DSG), and formaldehyde (FA), leading to increased immunoprecipitation of proteins that cannot directly bind to DNA, such as TET2. However, the double crosslinking also increases overall background noise compared to traditional FA-only crosslinking ChIP protocols.

As described before, BlaER-3xFlag cells were induced, to allow for TET2 accumulation, and analyzed in biological duplicates at 72 and 168 hp. Un-tagged BlaER cells induced in parallel were used as a control for unspecific binding.

6.2.1 TET2 ChIP validation by qPCR

To ensure the quality of the immunoprecipitation, a fraction of the IP'ed DNA was used to verify the capture efficiency by ChIP-qPCR. Although we lacked confident positive controls, our methylation data (WGBS-seq) strongly suggested *IL1RN* promoter region as a good candidate. The region follows a strong demethylation at 96 hpi, which is even more prominent in iMac (data not shown). Hence suggesting TET2 binding and activity. A small upstream intergenic region devoid of histone marks, transcription factor binding and not displaying DNAm changes was used as a negative control for the ChIP-qPCR.

Amplification results showed the expected enrichment at the candidate region at both time points (**Fig. R9**). Notably, ChIP at 72 hpi showed the best enrichment, suggesting a more efficient TET2-chromatin capture. The TET2 chromatin enrichment in the remaining samples was assessed by high-throughput sequencing.

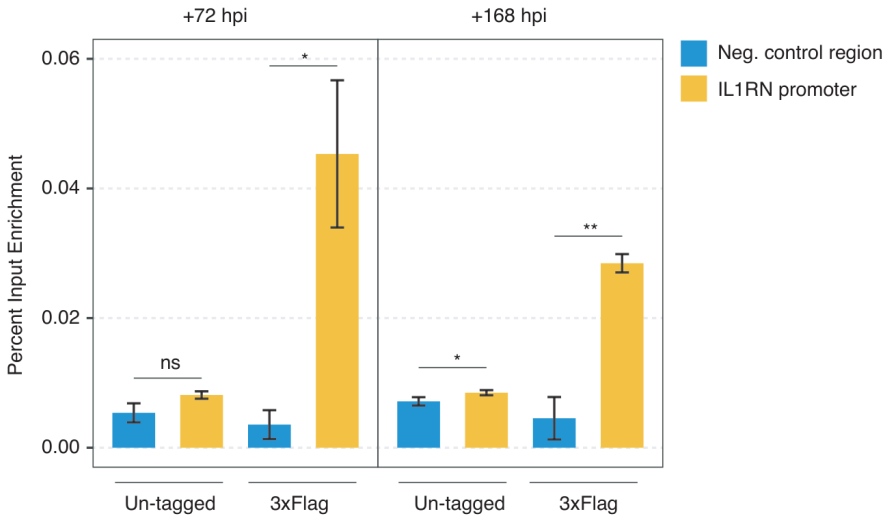

Figure R9. ChIP-qPCR showing the enrichment of TET2 at the *IL1RN* promoter region at 72 and 168 hpi. A region 5kb downstream of the *IL1RN* is shown as a negative control region. Unpaired two-tailed Student's t-test, n=3, mean s.e.m., (*p<0.05; **p<0.01)

6.2.2 Uncovering TET2 genome-wide chromatin binding sites

ChIP-seq analysis showed solid results, with more than 10,000 significant peaks at the 72 hpi time point in both biological replicates. In contrast, the 168 hpi samples showed a lower overall signal and presented less than 1,000 confident peaks. Importantly, one of the 72 hpi duplicates was lost during library preparation, which completely hindered any future statistical analysis.

Therefore, we have focused our attention on the high-quality 72-hour timepoint, which should give an exhaustive map of early binding events that may cause cell fate switching and potential loss of tumorigenicity. In contrast to later time points, which might bias TET2 recruitment to relevant genes for the physiologic macrophage functions. Of note, the preliminary analysis indicated that almost all the peaks detected at 168 hpi were included in the 72 hours' time point.

To identify high-confidence TET2 peaks within the 72 hpi replicates, we performed Irreproducibility Discovery Rate (IDR) statistical correction. We identified a total of 4,353 confident TET2 binding sites in our model (IDR < 0.05),

of which 40.56% were annotated by proximity to promoters, 54.76% to potential enhancers, and 4.68% to other genome regions (**Fig. R10a**).

GO analysis of the associated genes to the TET2 peaks revealed significant enrichment in myeloid-related terms (**Fig. R10b**). These included 'Phagocytosis', 'Leukocyte migration', or 'Mononuclear cell differentiation', among others. They encompassed crucial myeloid TFs such as *RUNX1*, *MAFB*, *IL1B,* or *JAK2.*

Notably, these results align with the previously described importance of TET2 in myeloid commitment programs recapitulated in our model [100,194,200].

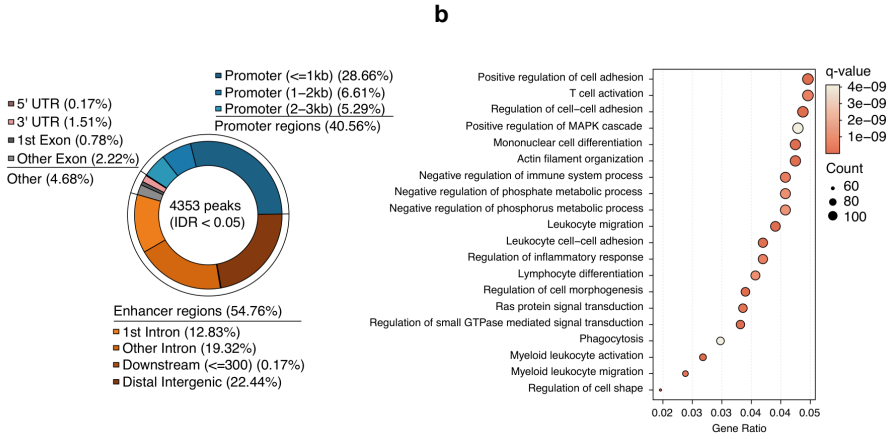

Figure R10. Overview of the confident TET2 binding sites during TD. (a) Distribution of high confident TET2 binding sites (IDR<0.05) during TD along different genomic features. (b) BP gene ontology analysis of genes associated with TET2 binding sites.

6.2.3 TET2 chromatin target profiling and classification

To gain a better understanding of the regulatory significance within TET2 chromatin binding, we further characterized the chromatin status of the TET2 target regions.

To do so, we used publicly available datasets that have extensively profiled the genome and chromatin architecture of the BlaER cells during the TD process [289,307]. As well as an unpublished in-lab generated WGBS-seq dataset for genome-wide methylation profiling [290].

6.2.3.1 TET2 binding site's role in 3D chromatin long-range interactions during TD

Preliminary analysis suggested a strong overlap of the TET2 chromatin targets with GREs, suggesting potential TET2 involvement in genome topology (**Fig. R10a**).

Analysis of 3D chromatin configuration by Hi-C [289] revealed that TET2 binding sites were associated with a significant rewiring of the chromatin organization (**Fig. R11**). TET2 sites were involved in inter-TAD, long-range interacting regions (5-10Mb). Notably, these interactions have been described as crucial for TD-associated compartmentalization changes [289]. TET2 binding regions formed interacting clusters specifically at the iMac stage, suggesting the protein's implication in far-reaching chromatin rewiring during cell fate commitment and not just as a classical intra-TAD player.

Mechanistically, it is still unclear how TET2 participates in the process since we have not found any significant association between the methylation levels and rewiring strength, suggesting methylation-independent implication.

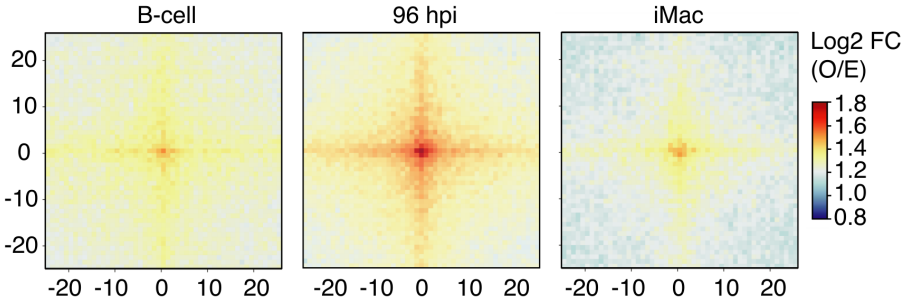

Figure R11. Aggregate metaplots (10-kb resolution) showing long-range interactions (5-10Mb) between TET2 bound regions during TD.

6.2.3.2 TET2 binding sites classification and identification of dynamic regulatory regions

Next, we assessed the association of TET2 binding sites with GREs during TD. To that end, we overlapped all chromatin-accessible regions (ATAC-seq) detected during TD with either transcription starting sites (TSS) or H3K4me1 peaks, a typical mark of poised enhancers. This filtered the TET2 peaks in promoters (P) or enhancers (E), which were then interrogated for their activity based on the H3K27ac mark (**Fig. R12a**).

Accordingly, we categorized the TET2-bound GREs into inactivated (loss of H3K27ac), maintained stable (no changes in H3K27ac) or activated (stable or transient gain of H3K27ac) (adapted from Stik Nature Genet 2018). By doing so, we successfully classified the majority of TET2 binding sites (n=3,997, 91.7%), which were further considered for the downstream analysis. The unclassified TET2 peaks (n=356) were filtered out of the analyses because they did not pass at least one of the setup thresholds. However, they were equally located at regulatory regions of important TFs such as *HES1*, *KLF2* or *EIF5*.

Globally, a large portion of TET2 binding sites were associated with enhancers (n=2,534), while a minor portion was bound to promoters (n=1,463). The latter highlights the biological robustness of the classification compared to the initial annotation (**Fig. R10a**), mainly by biasing TET2 sites towards the known preferential enhancer-binding. Interestingly, while nearly half (46.9%) of the TET2-bound GREs overlapping promoters did not exhibit any significant H3K27ac changes during TD, 60.5% of the TET2-bound GREs overlapping enhancers showed a consistent gain or transient gain in H3K27ac during the conversion (**Fig. R12b**). This strongly suggests the dynamic nature of these regions, which are often associated with DNA demethylation (Rasmussen Genome Research 2019).

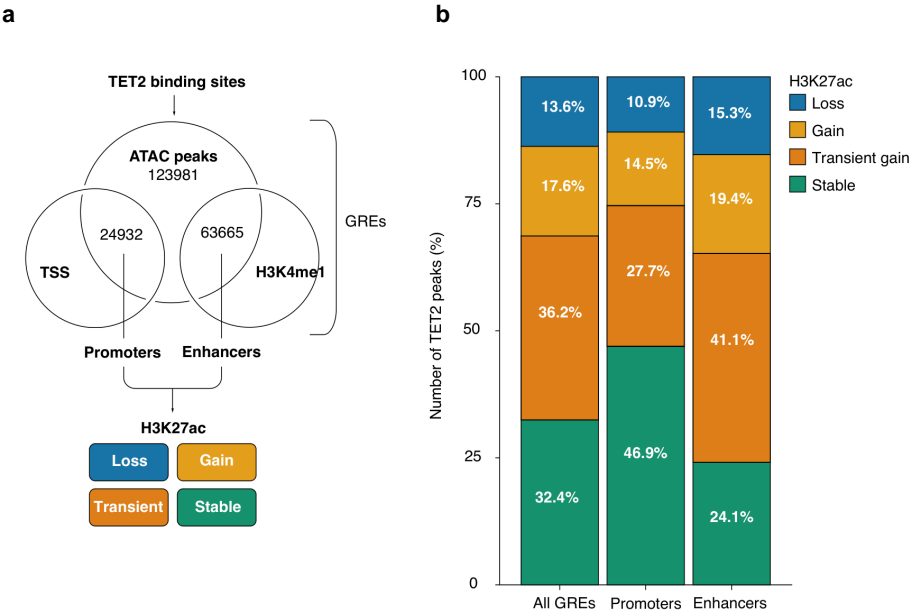

Figure R12. Chromatin based GREs classification of TET2 binding sites. (a) Schematic overview of the strategy to classify TET2 binding sites into different subsets of gene regulatory elements (GRE) based on chromatin features. (b) Distribution of the TET2 binding sites associated with different chromatin activity states (based on H3K27ac dynamics) along promoter or enhancer regions. All GREs are depicted as a distribution control.

6.2.3.3 TET2 binding at GREs associates with gene expression

To examine the transcriptional impact of TET2 binding at the different chromatin subsets, we analyzed publicly available expression data along TD [289]. We observed that TET2-bound genes associated with gain, and to a lesser degree, transient H3K27ac gain, were strongly upregulated during TD. While the subset with stable or losing H3K27ac did not show great changes (**Fig. R13**). Overall, this supports the idea of TET2 binding at active GREs, which usually leads to gene upregulation.

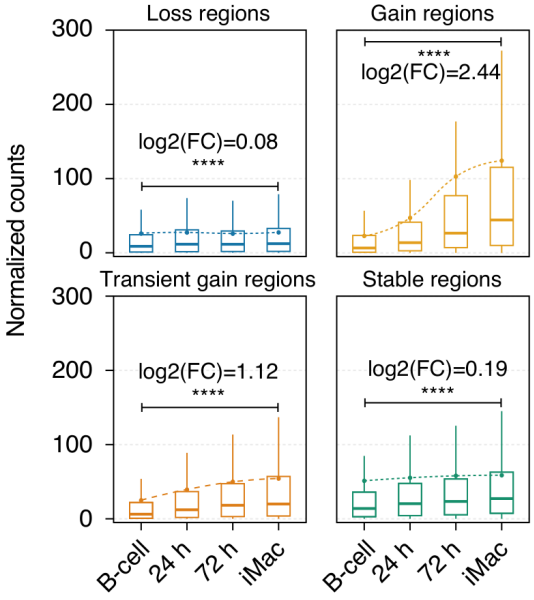

Figure R13. Gene expression dynamics, during TD, at the TET2 binding sites associated with different chromatin activity states (based on H3K27ac dynamics) (B-cell vs iMac two-sided Wilcoxon rank- sum test; **** p-value<0.0001). Colored dots show mean expression values.

6.2.3.4 DNA methylation and activity at TET2 binding sites

We further assessed DNAm levels at all TET2 binding sites by integrating them with the WGBS-seq data. We analyzed our TET2-bound regions of interest and categorized them into Methylation Steady Regions (MSRs) and Methylation Dynamic Regions (MDRs) based on their methylation changes between the B-cell (0 hpi) and the iMac (168 hpi) stages. MDRs display at least a 20% methylation change between the two-time points, while MSRs were maintained below that threshold.

Remarkably, the majority of TET2 binding sites were associated with MSRs (n=3,214) that were devoid of methylation in B cells and maintained as such during TD. Meanwhile, MDRs, which were the remaining 18.8%, underwent changes during TD (n=604) (**Fig. 14a**).

These findings suggest that DNAm at TET2 binding sites is mainly maintained at low and stable levels, typical of CGIs. Meanwhile, only around 20% of regions are the catalytic TET2 targets, and loose DNAm along TD. The latter highlights the selectivity of TET2 activity despite its distribution along all types of GREs.

Interestingly, both genes associated with MSRs or MDRs were upregulated during TD, highlighting the known TET2 function as a positive transcriptional regulator expression (**Fig. 14b**). However, the genes associated with MDRs displayed more than 2-fold larger differences between B-cell and iMac stages, even when starting at lower levels. Overall, these findings confirm the importance of methylation-dependent gene activation in cell fate commitment and the role of TET2 in promoting gene upregulation.

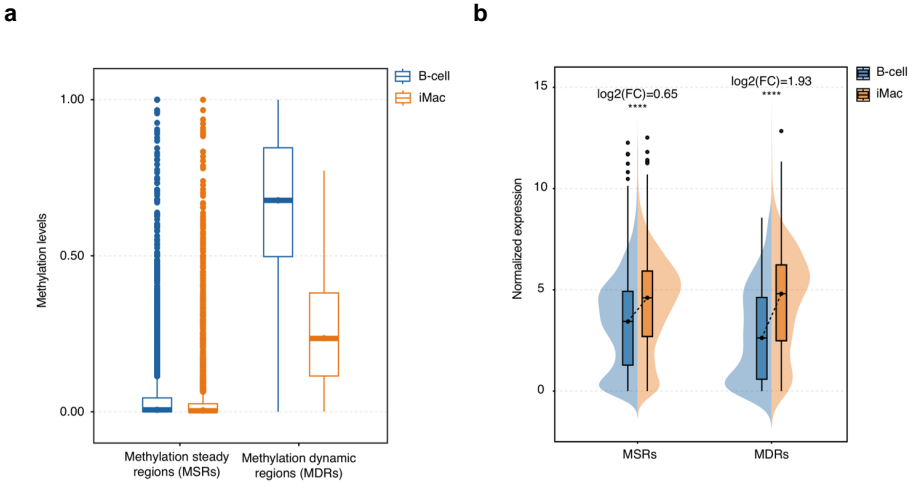

Figure R14. Classification of TET2 binding sites according to DNAm dynamics during TD. (a) DNAm levels in B cells and iMacs at TET2-bound Methylation Steady Regions (MSR) or TET2-bound Methylation Dynamic Regions (MDRs). (b) Average expression levels for genes associated with TET2-bound MSR or TET2-bound MDRs (B-cell vs iMac two-sided Wilcoxon rank-sum test, **** p-value<0.0001).

Since an important proportion of the TET2 targets are located distally from their potentially regulated genes (Enhancers: 59.2%), measuring the transcriptional impact of TET2 binding at these genes might be underrated. The previous transcriptional quantification provides a great overview, but it is limited by the idea of a single E-P regulation by proximity. In reality, enhancer elements most

likely regulate more than one gene. On the other hand, multiple enhancers can often regulate a single gene.

Regardless of the scenario, eRNA provides a direct proof of enhancer activation. Therefore, we used transient transcriptome sequencing (TT-seq) data to measure enhancer RNA (eRNA) synthesis [308].

Specifically, we took the publicly available TT-seq data during TD [307] and evaluated the eRNA synthesis activity at the available TD timepoints in DNAm classified TET2 sites.

The interrogation showed that MDRs showed upward tendencies with higher eRNA synthesis in both strands compared to MSRs (**Fig. R15**). Some transient activation was observed at early time points (12-24 hp) in all cases, most likely as a consequence of CEBPA pulse. Nonetheless, MDRs appear to significantly increase their eRNA at +72 hpi, once again coinciding with the previously mentioned strong TET2 accumulation. eRNA evaluation at later TD timepoints (120-192 hpi), might show even clearer enhancer activation, if such, is DNAm dependent. Since our WGBS data indicate the strongest demethylation at 168 hpi [290]

Regardless, this provides additional information regarding how specifically methylation dynamic regions get activated during TD transcriptional remodeling.

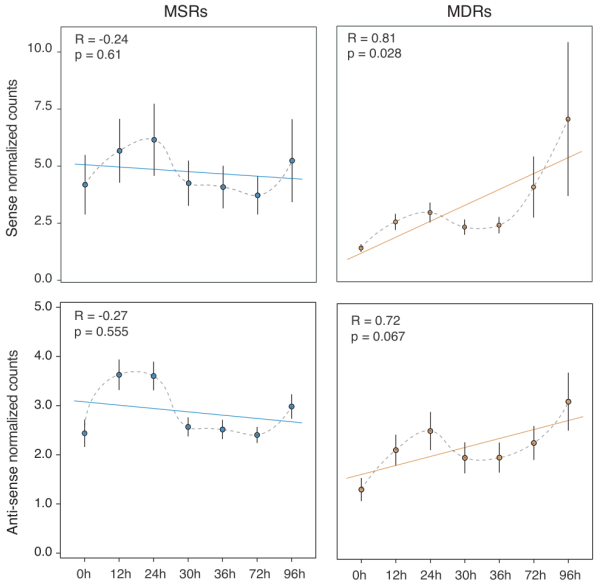

Figure R15. Activation of DNAm classified TET2 binding sites, according to eRNA synthesis (TT-seq). Both sense and antisense strand eRNA synthesis is shown. Linear regression (R) and its p-value is shown to illustrate the tendencies along TD timepoints.

6.2.3.5 Motif analysis of TET2 binding sites and characterization of transcription factor co-occupancy

To better understand additional transcription factors that cooperate with TET2 for methylation-mediated gene modulation, we performed a differential motif analysis between the MSRs and MDRs subsets. Our results showed that MDRs are differentially enriched in the motifs of the POU Class Homeobox family, SPI1, and, importantly, also in the binding motifs of the CEBP-family (**Fig. R16a**). As previously mentioned, CEBPA is a crucial transcriptional factor for myeloid cell fate commitment and a known interactor of TET2, and, therefore, likely mediates TET2 recruitment to specific GREs during TD. Importantly, CEBPA binding has also been associated with extensive chromatin rewiring, particularly at enhancer regions, which are the preferential binding sites of both proteins [100,230,295].

Therefore, we interrogated the DNAm classified TET2 sites with CEBPA ChIP-seq, ATAC-seq and a bona fide enhancer mark H3K4me1 data. As previously

suggested by motif enrichment analysis, we observed that in iMac, MDRs are strongly co-occupied by TET2-CEBPA, which was not the case for MSRs. Further comparison between B-cells and iMac, also indicated strong chromatin aperture and decoration with the H3K4me1 histone, specifically at MDRs in iMac (**Fig. R16b**). Unsurprisingly, the lowly methylated status observed in the MSRs correlated with stably open chromatin and high H3K4me1 levels, suggesting lack of GRE activation during TD.

To sum up, although TET2 binding was mostly detected at regions with low and stable methylation, the dynamic subset was the one associated with a strong upregulation during TD. Extensive dataset integrations and motif enrichment analysis, suggest that this is likely mediated by the interplay between TET2-CEBPA. Both factors preferentially bind to methylation dynamic enhancers that get chromatin open and activated.

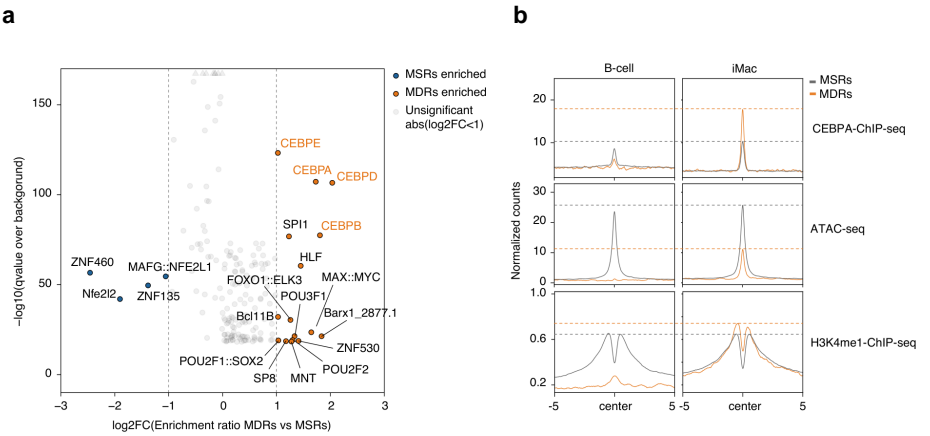

Figure R16. TF co-occupancy and chromatin state evaluation at DNAm classified TET2 binding sites. (a) Motif enrichment analysis (MEME suit) of MSR/MDRs TET2 binding sites. (b) Profile plots showing normalized signal for CEBPA ChIP-seq, chromatin accessibility (by ATAC-seq) and H3K4me1 ChIP-seq data at TET2- bound MSR or MDRs in B-cells and iMacs. The dashed lines represent values in iMac for each dataset.

6.2.3.5 TET2 binding sites functional classification overview

Finally, to give a full chromatin picture at TET2 binding sites, we integrated the methylation-based classification (MSRs/MDRs) with GREs H3K27ac-based dynamical classification (Loss/Gain/Transient gain/Stable).

Remarkably, the overlap of the two classifications showed strong enrichment for H3K27ac gain regions at enhancer and, to a lesser degree, promoter MDRs (**Fig. R17**). MSRs, on the other hand, presented all types of H3K27ac regions and were not particularly enriched in any of them.

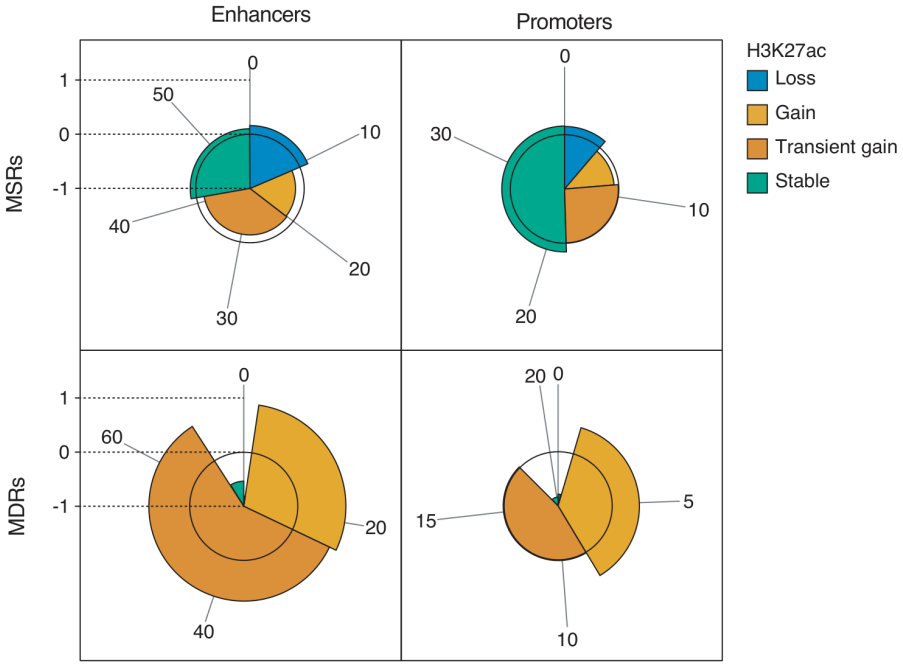

Figure R17. TF co-occupancy and chromatin state evaluation at DNAm classified TET2 binding sites. (a) Motif enrichment analysis (MEME suit) of MSR/MDRs TET2 binding sites. (b) Profile plots showing normalized signal for CEBPA ChIP-seq, chromatin accessibility (by ATAC-seq) and H3K4me1 ChIP-seq data at TET2- bound MSR or MDRs in B-cells and iMacs. The dashed lines represent values in iMac for each dataset.

Ultimately our findings strongly suggest the role of TET2-CEBPA mediated enhancer activation during TD. TET2 demethylation at specific myeloid GREs subset, coincides with strong chromatin aperture and active enhancers H3K27ac/H3K4me1 marks decoration. Which happens together with transcriptional mRNA and eRNA upregulation.

(VII) TD methyl-transcriptome profiling in TET2 reduced conditions

The main idea regarding TET2-mediated demethylation is that it ultimately results in a transcriptional shift necessary for development and cell fate commitment. In the previous sections, we have shown how accumulation of TET2 during TD in our model results in site-specific DNA demethylation at +72 hours post-induction, which is mainly associated with myeloid-related program activation, such as phagocytosis, leukocyte migration, and adhesion (see 6.2.2). On top of that, these DNAm differences have been well correlated with promoter regions, chromatin aperture, and gene upregulation (Valcarcel BioRxiv 2024).

Conversely, we wanted to interrogate how depletion of TET2 during TD could lead to aberrant DNAm phenotypes and sub-optimal gene expression. To wholly interrogate this axis, we evaluated and integrated both layers in a high-throughput approach.

7.1 Methylome evaluation in TET2 depletion during cell fate conversion

We used the doxycycline-inducible shCTRL and shT2.1 cell lines (see 5.3) to assess the DNAm impact of TET2 depletion during TD. The Dox treatment was continued throughout the process, including a 48-hour pre-treatment stage, to ensure TET2 downregulation at the beginning of the process. Cells were evaluated in biological triplicates with the Infinium MethylationEPIC arrays at 3 different time points during cell fate conversion (0, 72 and 168 hpi). Reduced *TET2* levels were confirmed by RT-qPCR at each timepoint to ensure the validity of the downstream analysis (**Fig. R18**). Of note, during our validations, we noticed that the efficiency of the reduction was slightly affected at 72 hpi. This

was due to the kinetics of the system. *TET2* shows a sharp upregulation at this time point, so even small differences in the sample's kinetics could significantly affect the normalization of the expression levels.

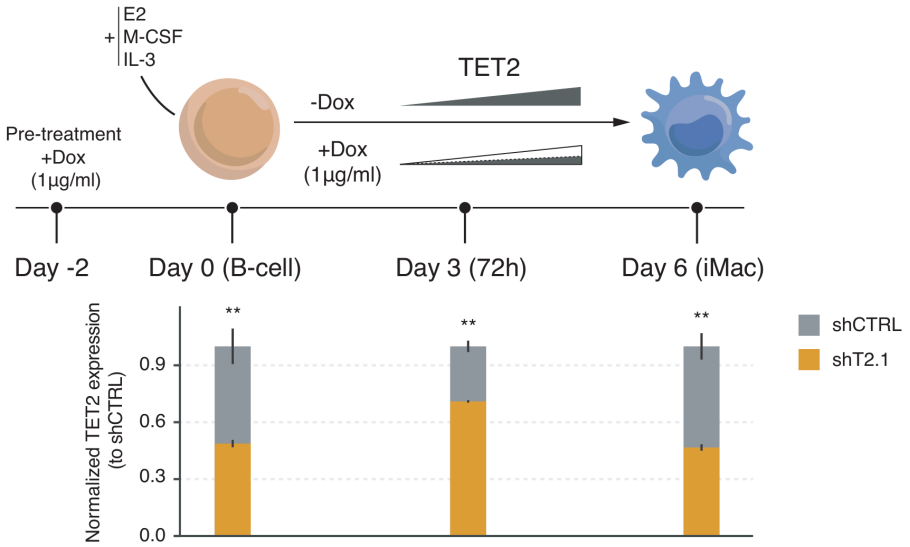

Figure R18. Schematic overview of the experimental design to evaluate consequences of TET2 depletion during TD. Validation of *TET2* reduction at each timepoint is shown. Unpaired two-tailed Student's t-test, n=3, mean s.e.m, (**p<0.01).

7.1.1 Hypermethylation kinetics during TET2 depleted TD

We detected strong hypermethylation in *TET2*-depleted cells at 72 hpi and particularly at iMac stage, where we detected up to 1794 differentially hypermethylated positions (DMPs) (**Fig. R19**). Some minor hypermethylation was found in B-cell, which was attributed to biological variances as almost none of the detected regions correspond with TET2 binding sites or GREs (data not shown). Similarly, hypomethylated positions, although significant, have not shown any particular genomic association. Nevertheless, TET2 depletion led to clear hypermethylation profiles, which strength correlated with protein accumulation during TD.

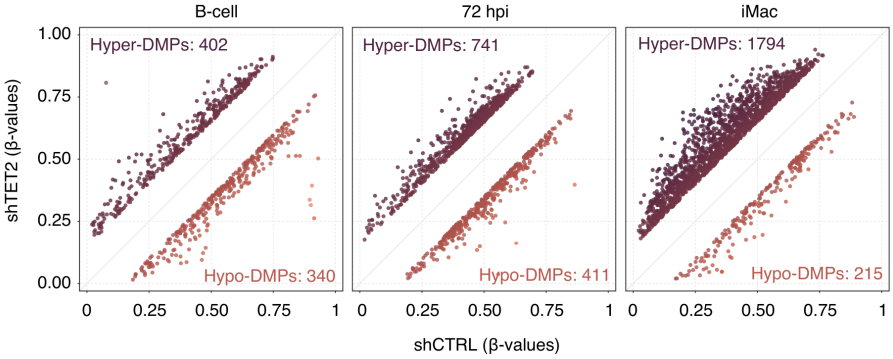

Figure R19. Scatterplot depicting significant Differentially Methylated Positions (DMPs) during TD. abs(shTET2 vs shCTRL diff>0.15); p-value<0.05.

7.1.2 Dynamic evaluation of the TET2 depleted differentially methylated landscape

To better understand the hypermethylated landscape and dynamically evaluate methylation kinetics at the affected region, we clustered all of the detected DMPs along TD, revealing 4 methylation clusters (M1-M4) (**Fig. R20a**).

Notably, the M2 and M3 encompassed most of the detected hypermethylation changes. Moreover, quantification of DNAm levels unveiled that only clusters M2 and M3 followed demethylation tendencies during TD and, more importantly, were significantly hypermethylated in *TET2*-depleted cells during the process (**Fig. R20b**). Clusters M1 and M4, although slightly hypermethylated after *TET2* reduction, were maintained at low and high levels, respectively, and were not significantly enriched for any particular biological processes.

On the contrary, M2 and M3 showed enrichment for genes associated with myeloid and immune-related programs such as *LPP*, *MAEA*, *IFNGR2* and *IL1RN*, among others. Notably, CpGs from the M2 cluster were mildly demethylated during TD and associated with genes related to myeloid leukocyte activation, phagocytosis, and macrophage differentiation. Meanwhile, CpGs from the M3 cluster followed strong demethylation tendencies and were associated with terms such as response to chemokines, myeloid migration, and calcium-mediated signaling (**Fig. R20c**).

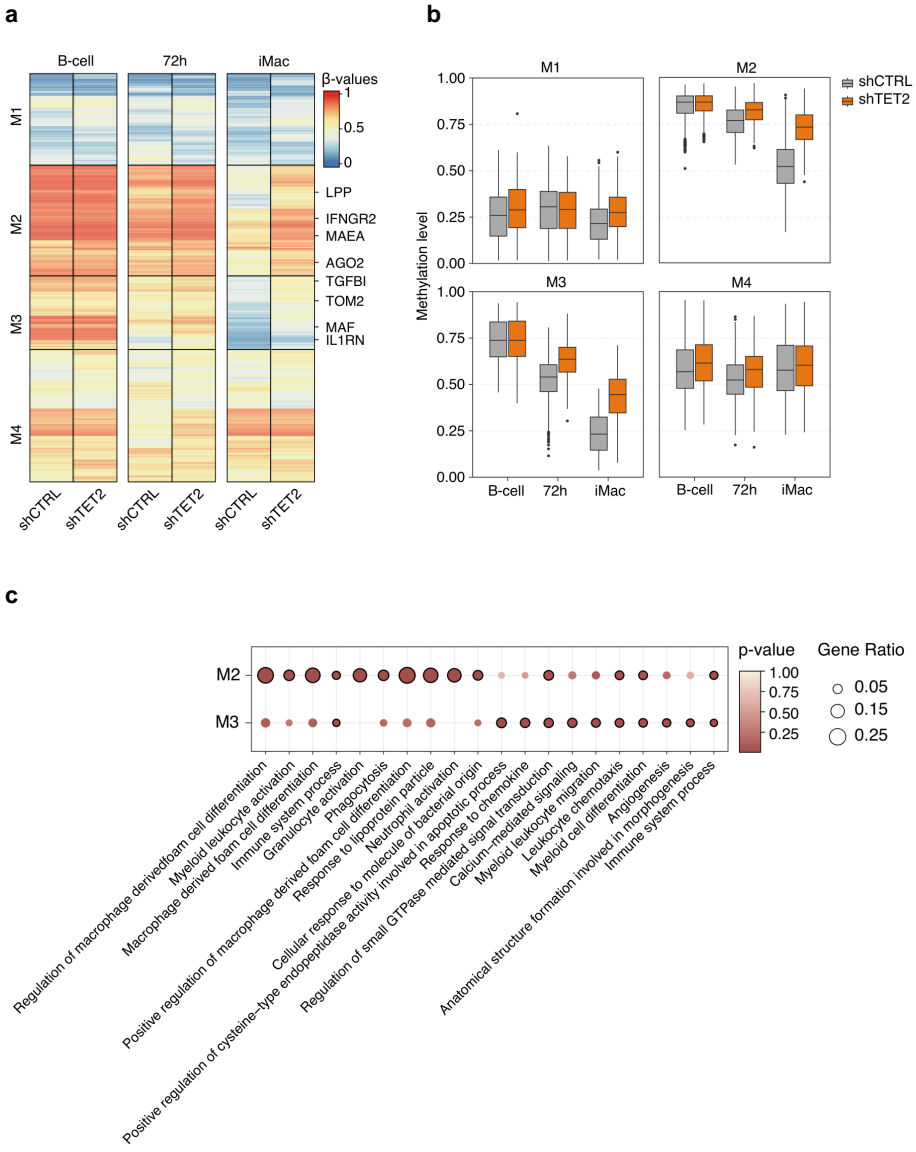

Figure R20. Evaluation of altered methylome landscape during TET2 depleted TD. (a) Heatmap of DNAm levels of significant DMPs detected in TET2 depleted conditions during TD. M1-M4 were defined by k-means clustering. DMPs associated with relevant genes are shown on the right. (b) Average methylation levels during TD within each

methylation cluster in control and TET2 depleted conditions. (c) BP gene ontology analysis of genes associated with M2 and M3 clusters.

7.1.3 WGBS corroboration of TET2 impaired activity during TD

We used our WGBS-seq dataset to conduct a more comprehensive assessment of the genomic regions surrounding the DMPs. This is because evaluating methylation using single CpG-centric methylation arrays has limitations. For instance, some distal gene regulatory regions are uncovered or covered by only one probe. Having this information is biologically significant as methylation changes rarely affect only one isolated CpG. WGBS-seq analysis unveiled that M2- and M3- DMPs are included in, on average, 200bp demethylated regions, contrasting with stable M1- and M4- DMPs (**Fig. R21**).

Therefore, the TET2 target candidates' evaluation was carefully based, among other parameters, on the methylation levels within the proximity of the detected DMPs.

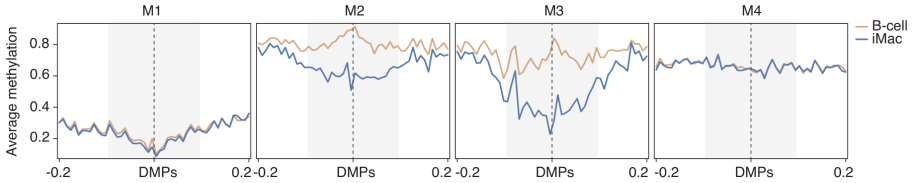

Figure R21. Evaluation of altered methylome landscape during TET2 depleted TD. (a) Heatmap of DNAm levels of significant DMPs detected in TET2 depleted conditions during TD. M1-M4 were defined by k-means clustering. DMPs associated with relevant genes are shown on the right. (b) Average methylation levels during TD within each methylation cluster in control and TET2 depleted conditions. (c) BP gene ontology analysis of genes associated with M2 and M3 clusters.

7.2 Transcriptome evaluation in TET2 depletion during cell fate conversion

Having pinpointed the clear influence of *TET2* depletion on the DNAm landscape, we next investigated its potential association with transcriptional changes. To do that, we have performed total RNA-seq in biological replicates

in the same conditions employed for the methylation arrays (**Fig. R18**), enabling a more direct comparison.

7.2.1 TET2 depletion effect on gene expression during TD

By performing differential expression analyses between shTET2 and shCTRL cells, we detected 2614 differentially expressed genes (DEGs) throughout the three TD time points. Notably, most of the DEGs were found at the iMac stage, showing 1222 significantly down and 474 up-regulated genes (**Fig. R22a**). The different transcriptomic profiles were reflected in a principal component analysis (PCA) analysis. This illustrated that *TET2*-depleted cells displayed a delayed TD transcriptional profile (**Fig. R22b**).

GO analysis at iMac stage showed decreased enrichment in expected processes related to myeloid differentiation, cell migration and response to stimuli, as well as others such as metabolic processes and amino acid transport (**Fig. R22c**). The most interesting DEGs included crucial TFs and myeloid genes, such as *IRF8*, *IL1B,* or *TP63*. No clear association with any significant biological process was observed for upregulated DEGs. Nonetheless, this subset includes genes associated with immune cell adhesion (*MCAM*, *ITGA8*, *CD2, ID1)* and other immune-related genes such as *SLAMF6* or *NOD2*.

These findings confirm TET2*'s* role in regulating myeloid pathways and suggest novel implications in processes such as metabolic rewiring, which has been described in reprogramming (Cheng Cell Rep 2020).

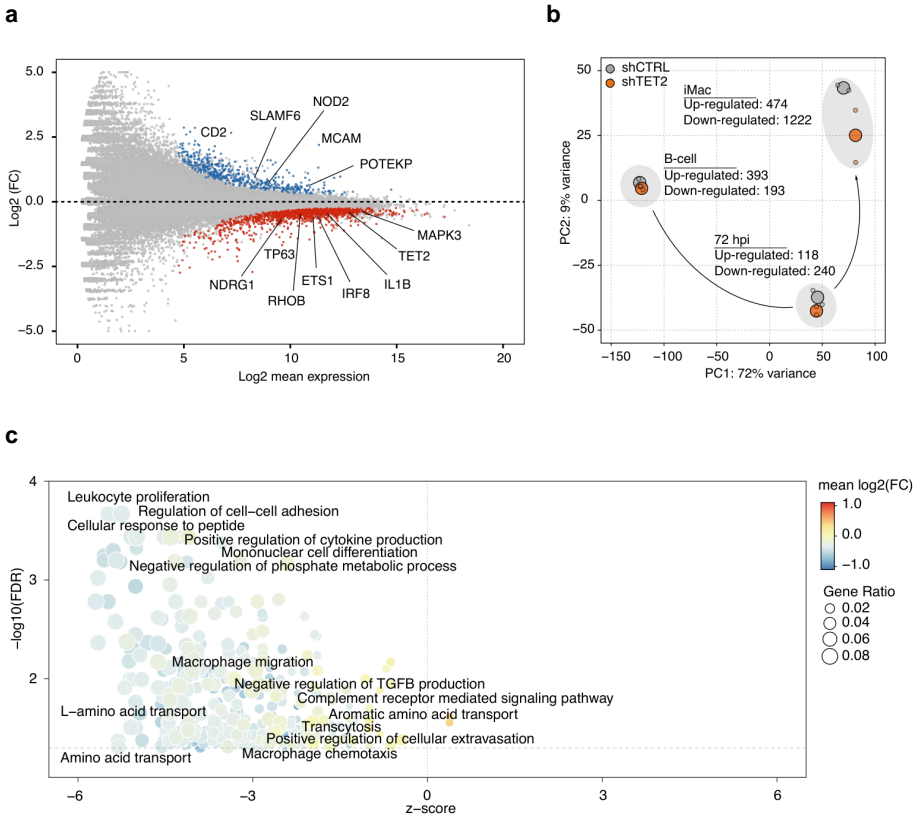

Figure R22. Evaluation of altered transcriptional landscape during TET2 depleted TD. (a) MA plot showing differentially 1,696 expressed genes (DEGs) in TET2 reduced iMac. FDR<0.05. (b) Principal component analysis showing the transcriptional kinetics in control and TET2 depleted conditions during TD. The Number of differentially expressed genes (DEGs) (p-adjusted<0.05) at each time-point is indicated. Big dots represent average values, and small dots represent individual replicates. (c) BP gene ontology analysis of DEGs at iMac stage in TET2 depleted conditions. z-score indicates the probability of decreased (negative value) or increased (positive value) term function.

7.2.2 Analysis of transcriptional dynamics during TD

After performing pairwise comparisons between the transcriptomic profiles of CTRL and TET2-depleted cells at different time points of the TD, we decided to analyze the normal transcriptional dynamics of the process. For that, we clustered all TD-associated DEGs in control samples, revealing 5 expression

clusters (E1-E5) (**Fig. R23a**). These included silencing clusters, E1-E2, that were enriched in genes related to B-cell activation and signaling programs (including genes such as *MZB1* or *IGHD*), ribosome biogenesis genes (including *NUDT21*), and genes related to DNA conformation changes (such as *HDAC5*, *CHD7*) (**Fig. R23b**).

Contrarily, activating E4-E5 clusters were related to myeloid and immune functions and included genes such as *TET2*, *IL1B*, *CCL3* or *DOK2*. Genes included in the E3 cluster followed a transient activation and were enriched in calcium-related pathways (such as *HTR2B*), which have been described to be crucial for macrophage activity, particularly in age-associated inflammation [309]. (**Fig. R23b**)

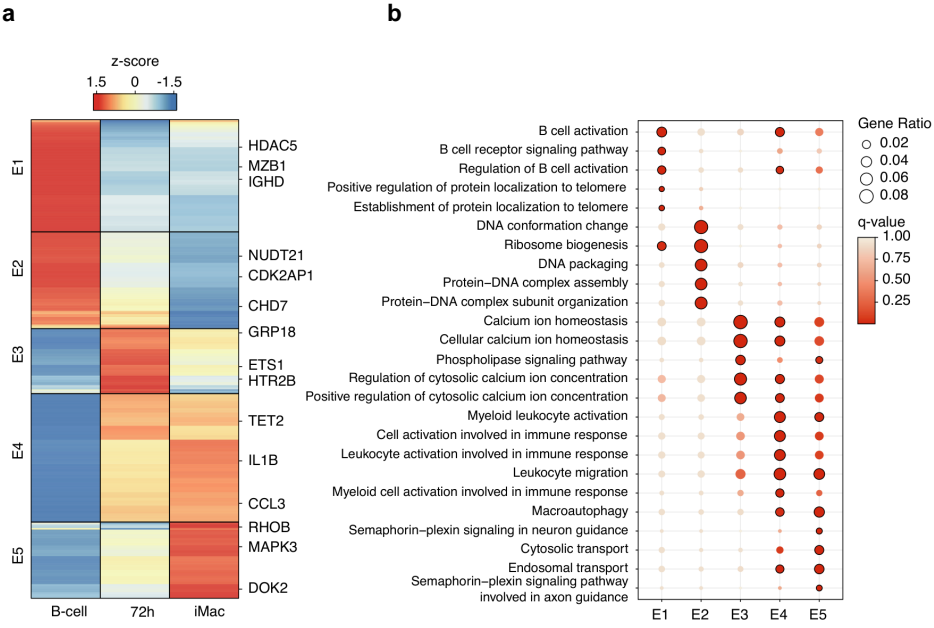

Figure R23. Evaluation of normal transcriptional landscape TD kinetics. (a) Heatmap of DEGs (absolute(log2FC)>1; p-adjusted<0.05) during TD in control conditions. E1-E5 were defined by k-means clustering. Biologically relevant genes are shown on the right. (b) BP gene ontology (BP) analysis of genes in E1-E5 expression clusters. Top 5 most significant terms of each cluster are shown. Significant terms (qvalue < 0.05) are highlighted with a black stroke

7.2.3 Analysis of transcriptional dynamics during TET2-depleted TD and its link with DNA methylation

Next, to determine the transcriptional impact that TET2 plays during the TD process, we quantified the average expression of the defined TD clusters (**Fig. R23a**) in control and *TET2*-depleted conditions.

Among all clusters, we observed abnormal transcriptional tendencies in *TET2*-depleted conditions only for the activating ones (E3, E4, and E5) (**Fig. R24a left**). Precisely, E3 showed a deficient final downregulation after transient activation at 72 hpi, and genes from E4-E5 failed to accumulate to control levels in TET2-depleted iMacs. These abnormal transcriptional dynamics might be related to hypermethylation events.

To test this hypothesis, we quantified the methylation levels of dynamic CpGs related to each expression cluster. The dynamic subset was defined based on regions of at least 20% DNAm change between B-cell and iMac (WGBS), which was used to subset EPIC array probes within the region of interest.

The analysis revealed the expected inverse relationship between expression and methylation (E3-E5), but more importantly, we also observed that TET2 depletion led to significant hypermethylation in clusters E4 and E5 (**Fig. R24a right**). This is in line with the detected impairment in gene accumulation as demonstrated by the expression dynamics of important myeloid and immune genes such as *IFNGR2*, *TLR1*, *MAF*, *ETS1,* and *NDRG1* belonging to E4-E5, (**Fig. R24b**).

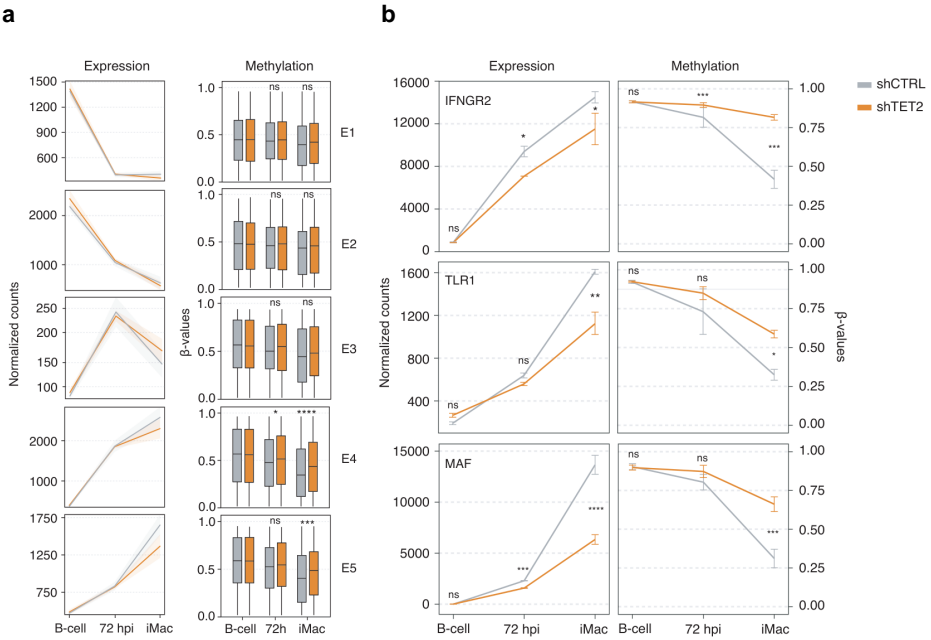

Figure R24. Evaluation of TD clusters expression and methylation kinetics. (a) Left: comparison of expression levels in each cluster during TD in control and TET2 depleted conditions. The line indicates the average value, and the shade region is the average standard deviation between the biological replicates (n=2). Right: DNAm levels at each expression cluster during TD in control and TET2 depleted conditions. Quantification was done in CpGs associated with genes that presented at least 15% methylation change during TD (B-cell vs iMac two-sided Wilcoxon rank-sum test; ns>0.05, * p-value<0.05, *** p-value<0.001, **** p-value<0.0001). (b) Altered expression-methylation profiles of selected gene candidates. Unpaired two-tailed Student's t-test, mean s.e.m, (*p<0.05, ***p<0.001, ****p<0.0001).

To sum up, our epigenome profiling allowed us to establish a confident methylation-expression map during TD and identify TET2 target genes in an unbiased manner. Through an integration of RNA-seq, methylation arrays, and supporting WGBS dataset, we determined consistent TET2 implications in myeloid and immune-related programs, including crucial macrophage physiological functions such as phagocytosis, cell activation, adhesion, and migration (**Fig. R23b**). Not only have we identified genes that are methylation-

dependent for their proper accumulation, but also whole subsets with abnormal transcription-methylation tendencies (**Fig. R24a**). Notably, *TET2* impairment during TD led to hypermethylation and an impaired gene accumulation of crucial immune-related gene subsets. The latter included important inflammation response players such as *IFNGR2*, *TLR1,* or *MAF* transcription factor (**Fig. R24b**). These observations confirm the previously described TET2's role as a positive transcriptional regulator in myeloid lineage differentiation and identify specific candidates for evaluation in normal and disease conditions.

(VIII) Evaluation of confident TET2 targets and integration of acute myeloid leukemia patient data

Once we uncovered the TET2 chromatin distribution and its depletion effect in our cellular model, we fully integrated the findings to identify bona fide targets at chromatin, DNA methylation, and transcriptional levels. To reinforce the analysis, we also incorporated publicly available data from acute myeloid leukemia (AML) patients, which should further pinpoint the overlapping events between our TET2-reduced cellular model and leukemia patients with *TET2* alterations.

8.1. Identification of DNA hypermethylation events in TET2 mutated AML patients

To evaluate the effect of TET2 alterations in a clinical setting, we have processed Infinium MethylationEPIC array data (TCGA-LAML database) (Cancer Genome Atlas Research Network Nat Genet 2013) from 15 $TET2^{MUT}$ and $TET2^{WT}$ AML patients. Importantly, the patient samples were selected excluding carriers of other genes that might impact DNAm, such as *DNMT3A*, *IDH1/2*, *WT1,* and *CEBPA* [233,310,311].

The analysis showed that $TET2^{MUT}$ patients presented a strong hypermethylated phenotype, displaying 3592 significantly hypermethylated CpGs and only 304 hypomethylated ones (**Fig. R25a**). Gene ontology of the hypermethylation

association genes showed enrichment in immune and myeloid cells processes, including response to stress, leukocyte migration and differentiation (**Fig. R25b**). Encompassing genes such as *TLR9*, *ACIN1*, *TGFB1* or *RUNX2*, among others. No significant association with any biological process was observed for the genes associated with hypomethylated CpGs.

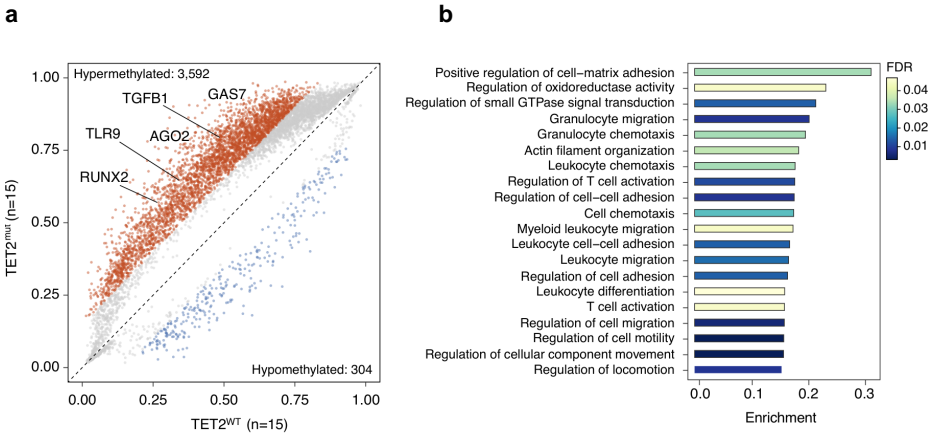

Figure R25. Evaluation of differential methylation in TET2^{MUT} AML patients. (a) Scatterplot depicting significant DMPs in TET2^{MUT} vs TET2^{WT} AML patients. abs(diff>0.15); p-value<0.05. (b) BP gene ontology analysis of genes associated with the detected hypermethylated DMPs in TET2^{MUT} AML.

8.1.1 Data integration and confident TET2 targets identification

To better characterize the hypermethylated events found in both TET2-mutated AML patients and the TET2-reduced cellular model, we investigated their connection to DNAm dynamic TET2 binding sites (MDRs) and to the specific set of genes that are differentially expressed during TD. This approach allows detecting *TET2* targets starting from a single CpG and the evaluation of the consequential chromatin levels and gene expression.

Precisely, we intersected TET2^{MUT} AML patient's hypermethylation events (AML-EPIC arrays) with DEGs (TET2-DEGs) and DMPs (TET2-EPIC arrays) found in *TET2*-reduced iMacs. For the intersection, we also considered TET2 bound regions (from the TET2-ChIP-seq) that presented at least 15% methylation reduction during TD (WGBS-seq). The output of the intersection of

the 5 datasets led to the identification of 13 bona fide TET2 target genes potentially involved in the pathology of TET2 mutated AML patients (**Fig. R26**). These include genes such as N-myc down-regulated gene 1 (NDRG1), which has been linked to proper neutrophil maturation [312], or LncRNA PVT1 oncogene which has been described to promote the malignant progression of AML [313]. However, the most interesting candidate was AGO2, whose DNAm-mediated regulation has never been reported and could provide novel insights into TET2^{MUT} leukemogenesis.

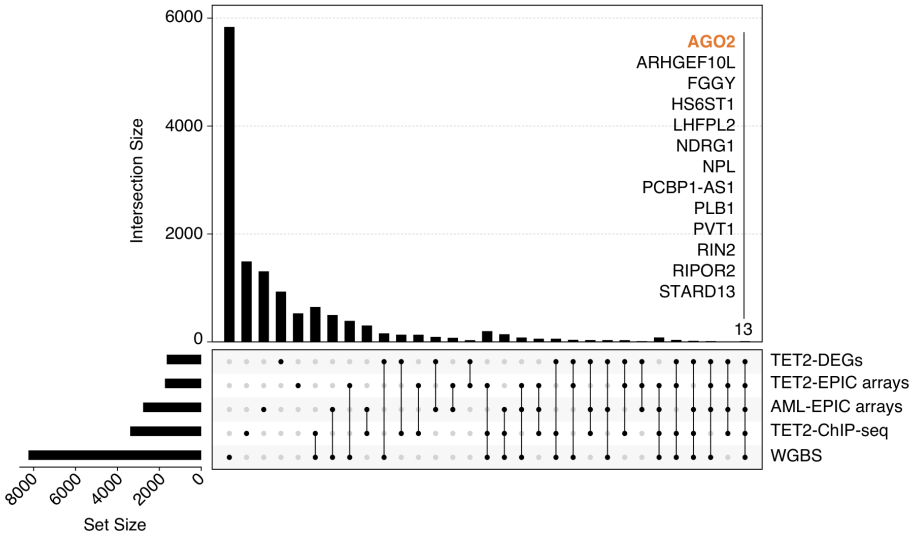

Figure R26. UpSet plot depicting the integration of in vitro BlaER generated datasets and TCGA LAML Infinium MethylationEPIC array data. 13 candidate genes present in all datasets are highlighted.

8.1.2 AGO2 evaluation as a relevant TET2 target in disease

Argonaute 2 (*AGO2*) is a core component of the miRNA-induced silencing complex (miRISC). Briefly, AGO2 binds to microRNA (miRNA) or short interfering RNA (siRNA) and guides them to homologous mRNAs for translational suppression or degradation. Importantly, AGO2 is the only member of the AGO family with endonuclease activity. It has been extensively linked to oncogenesis, immune response, and even prognostic value, although with contrasting findings [314].

To unveil the potential role of AGO2 in AML, we explored publicly available datasets for Acute Myeloid Leukemia (TCGA-LAML) to preliminary evaluate its gene expression and possible influence on survival. Analysis with the GEPIA2 tool (Tang Nucleic Acids Res. 2019) showed that *AGO2* was overexpressed in several tumor datasets such as Lymphoid Neoplasm Diffuse Large B-cell Lymphoma (DLBC) and Thymoma (THYM), Head and Neck squamous cell carcinoma (HNSC), Pancreactic adenocarcinoma (PAAD) and Stomach adenocarcinoma (STAD) (data not shown). Moreover, a subset of AML patients also presented elevated *AGO2* levels (**Fig. R27a**), which is in line with previous findings that pointed to *AGO2* as a tumor maintenance oncogene (REF). Additionally, the analysis of the overall survival suggested that high *AGO2* expression patients present a lower survival than *AGO2*-low patients. Therefore, being classified as a high-hazard ratio event (HR=2) (**Fig. R27a**). These data strongly suggest that AGO2 might participate in leukemic initiation and/or progression. This has led to its further evaluation through our cellular model and other available datasets.

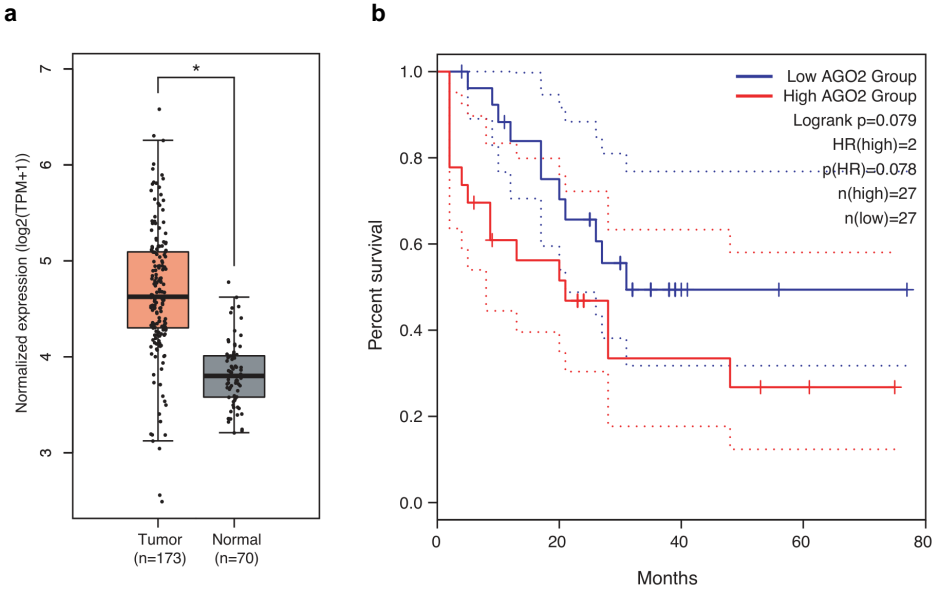

Figure R27. Evaluation of AGO2 in Acute Myeloid Leukemia (TCGA-LAML) dataset. (a) AGO2 expression levels in LAML tumor vs paired healthy samples. One-way ANOVA,

mean s.e.m., (*p<0.05). (b) Overall survival of AGO2 high vs AGO2 low expression patients. The data was analyzed by Mantel-Cox Log-rank test.

8.2. AGO2 evaluation in normal and malignant hematopoiesis

As the first approach to integrate our findings, we took advantage of the hematopoietic epigenome data from AML patients and healthy individuals generated as part of the Blueprint project (http://blueprint-data.bsc.es/#!/about).

8.2.1 AML patient's genome-wide DNA methylation exploration

We have selected 21 AML patients from the Blueprint project, who were extensively characterized (including WGBS-seq, RNA-seq, DNAse-seq and histone marks by ChIP-seq) to interrogate the status of *AGO2* in leukemic cells.

We classified the selected cohort, based on average methylation levels at the *TET2* bound regions detected in our cell model, into Hypo-methylated-AML (Hypo-AML, n=12) and Hyper-methylated-AML (Hyper-AML, n=9) patients (**Fig. R28a**). As expected, the comparison of the DNAm levels between Hypo-AML and Hyper-AML patients showed significant differences (**Fig. R28b**).

Site-specific comparisons between the two groups showed clear hypermethylation at *TET2* chromatin targets in Hyper-AML patients, with 340 regions having at least 20% change between the two groups (**Fig. R28c**). Among the top candidates, we found inflammation (*IL2RA*, *IL1B*), retinoic acid receptor (*RXRA, RARA*) and cell cycle (*GAS7, CDC7*)-related genes, in addition to *AGO2*.

Remarkably the evaluation of all TET2 binding sites (BlaER ChIP-seq) at AGO2 locus showed that, four out of five regions were hypermethylated (**Fig. R28c**) with an average of 30% increase (data not shown). Which strongly suggest that not only the previously characterized candidate enhancer goes through methylation changes, but also other TET2 bound regions within and in the vicinity of the *AGO2* locus. This finding suggests the presence of a tight methylation-mediated regulatory region, which most likely affects *AGO2* expression.

To further support this hypothesis, we analyzed gene expression profiles of the same AML patients and found well-correlated *TET2* levels based on our classification, which, although not significant, were clearly lower in the hypermethylated group. *AGO2* also showed lower expression in Hyper-AML cohort, supporting the finding that gene locus hypermethylation leads to reduced expression (**Fig. R28d**). Integration of more WGBS profiled AML patients and evaluation of TET2-AGO2 expression should give more statistically significant results. Nonetheless, these findings show how TET2-mediated *AGO2* demethylation constitutes a biologically relevant event during myeloid cell fate commitment, which might directly or indirectly modulate *AGO2* expression.

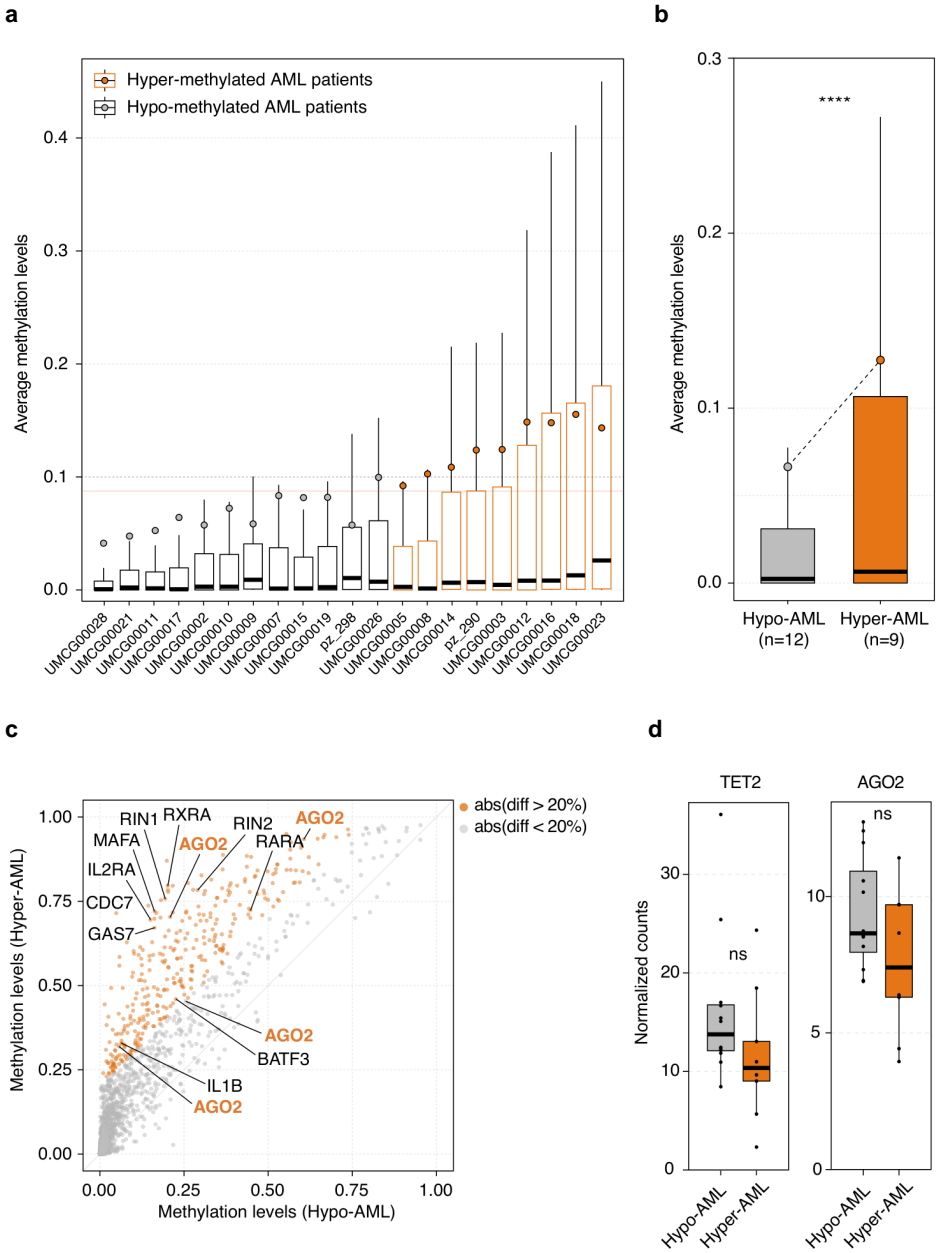
a
Hyper-methylated AML patients
Hypo-methylated AML patients
Average methylation levels
0.0
0.1
0.2
0.3
0.4
UMCG00028
UMCG00021
UMCG00011
UMCG00017
UMCG00002
UMCG00010
UMCG00009
UMCG00007
UMCG00015
UMCG00019
pz_298
UMCG00026
UMCG00005
UMCG00008
UMCG00014
pz_290
UMCG00003
UMCG00012
UMCG00016
UMCG00018
UMCG00023
b
Average methylation levels
0.0
0.1
0.2
0.3
****
Hypo-AML (n=12)
Hyper-AML (n=9)
c
abs(diff > 20%)
abs(diff < 20%)
Methylation levels (Hyper-AML)
Methylation levels (Hypo-AML)
0.00
0.25
0.50
0.75
1.00
RIN1
RXRA
RIN2
AGO2
MAFA
AGO2
RARA
IL2RA
CDC7
GAS7
AGO2
BATF3
IL1B
AGO2
d
TET2
AGO2
Normalized counts
ns
ns
0
10
20
30
0
5
10
Hypo-AML
Hyper-AML
Hypo-AML
Hyper-AML

Figure R28. AML patients' classification base on DNAm levels. (a) Methylation levels at TET2 binding sites (BlaER TET2-ChIP-seq dataset) in 21 selected AML patients (Blueprint data). Solid orange line (0.9) shows the methylation threshold used to define hyper/hypomethylated AML patients' cohorts. Colored dots show mean methylation values. (b) Average methylation levels at TET2 binding sites in hyper/hypomethylated AML patients' cohorts. Colored dots show mean methylation values. Two-sided Wilcoxon rank-sum test; mean s.e.m, (**** p-value<0.0001). (c) Scatterplot depicting average TET2 binding sites methylation differences between hyper and hypomethylated cohorts. Regions with at least 20% change between the groups are highlighted in orange. (d) Average expression levels (TPMs) of genes of interest in hyper and hypomethylated cohorts.

8.2.2 In vivo hematopoiesis overlaps with TET2-focused BlaER characterization

In light of our findings, we wanted to give a global overview on how our cellular model and the analyzed AML patients compared to healthy individuals. Once again, we performed an extensive methylation profiling from the Blueprint database to extract methylation levels in human hematopoietic stem cell differentiation [196]. To provide the most accurate map and integrate methylation information of TET2-reduced cells, we only focused on CpGs covered by Infinium MethylationEPIC arrays, that were bound by *TET2* and presented at least 20% change during TD in our cellular model (n=224). This subset was applied to an extensive list of healthy hematopoietic cells at all stages of differentiation (progenitors: HSCs, MPPs, CMPs, MEPs, GMPs, CLPs, MLPs | mature cells: CD4+ T-cells, T-CD8+ T-cells, NK-cells, B-cells, MKs, Neutrophils, Monocytes), previously classified Hyper-Hypo AML patients and BlaER cells during TD in normal and *TET2* reduced condition. PCA analysis showed clear differences between cell populations, with defined theoretical trajectories for both myeloid and lymphoid lineages (**Fig. R29**).

Focusing on the myeloid lineage, we observed that within the AML cohort, Hypo-AML patients presented blasts that were substantially more differentiated compared to the Hyper-AML group that slightly skewed towards the lymphoid maturation branch and overall showed a more immature state. Furthermore, control BlaER cells, although initially different, progressively transdifferentiated towards healthy monocytes and neutrophils.

Notably, *TET2* reduction presented an abnormal TD trajectory, resulting in a delayed myeloid acquisition similar to the one found in AML patients (**Fig. R29**). These findings support the idea that the data obtained from our *TET2* knockdown cell model resemble the delayed maturation found during the leukemic cell arrest. Despite overall expected differences between *in vitro* and *in vivo* cells, we observe that demethylation events in both cases are similar, as clearly illustrated by the PCA trajectories.

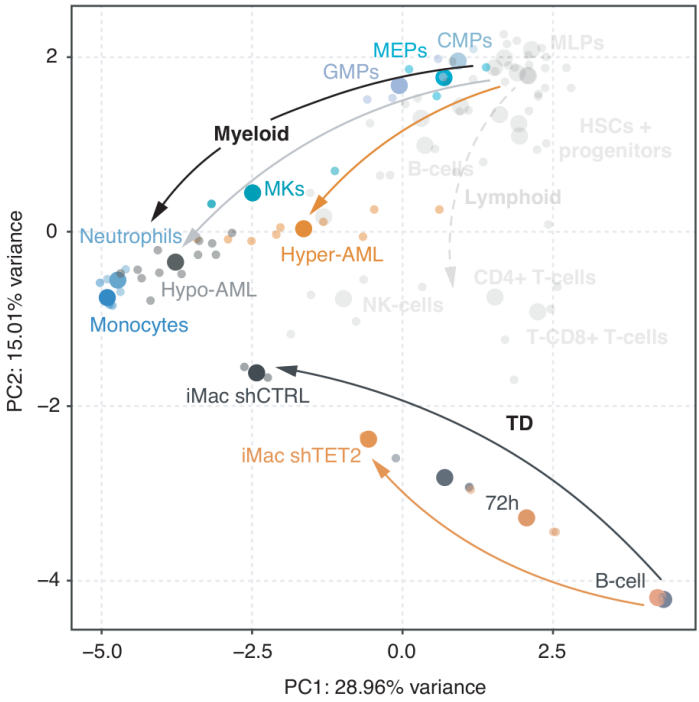

Figure R29. Principal component analysis of methylation kinetics in healthy hematopoiesis, hyper/hypomethylated AML cohorts and BlaER TD in control vs TET2 reduced conditions. Myeloid lineage cells are highlighted for better visualization. Big dots represent average values and small dots individual samples. Arrows represent trajectories for each dataset.

To sum up, we have faithfully integrated our TET2 chromatin/methylation and expression data with publicly available normal and AML datasets. We have identified *AGO2* as a confident target for methylation-mediated modulation

during myeloid cell fate commitment. Expression and methylation profiles in AML patients, along with our data, suggest its implication in disease maintenance, although the exact mechanisms are yet to be elucidated. Finally, we have shown how a specific subset of dynamically methylated loci is shared among *in vitro* and *in vivo* datasets, which provides a highly confident and previously unknown map of *TET2*-regulated regions during myeloid terminal differentiation.

8.3. AGO2 evaluation during TD-driven loss of tumorigenicity

Having defined the potential of *AGO2* as an oncogene (**Fig. R27**), we aimed to functionally dissect it in our cellular model, which is associated with the loss of tumorigenicity as cells transdifferentiate from leukemic B cells to macrophages.

We first checked AGO family transcriptional dynamics during the conversion. Although all members showed some level of gene upregulation, *AGO2* exhibited the highest accumulation. Interestingly, we have detected relatively transient transcriptional kinetics in *AGO2* and *AGO4*, that accumulated to the maximum levels at 72-96 hpi and then decreased at the iMac stage (**Fig. R30a**). This temporary accumulation could indicate potential *AGO2* involvement during intermediate myeloid commitment phases and is not necessary for terminally differentiated cell functions. Which could be highly relevant during leukemic blast arrest.

Regardless, our analysis shows that the *AGO2*-related DMP (cg00288598) was demethylated along normal TD and presented strongly hypermethylated in *TET2*-reduced conditions at 72hpi and very significantly in the iMacs stage (**Fig. R30b top**). This hypermethylation correlates with abnormally low *AGO2* levels, which were maintained at uninduced cell levels, with only slight accumulation at 72 hpi (**Fig. R30b bottom**).

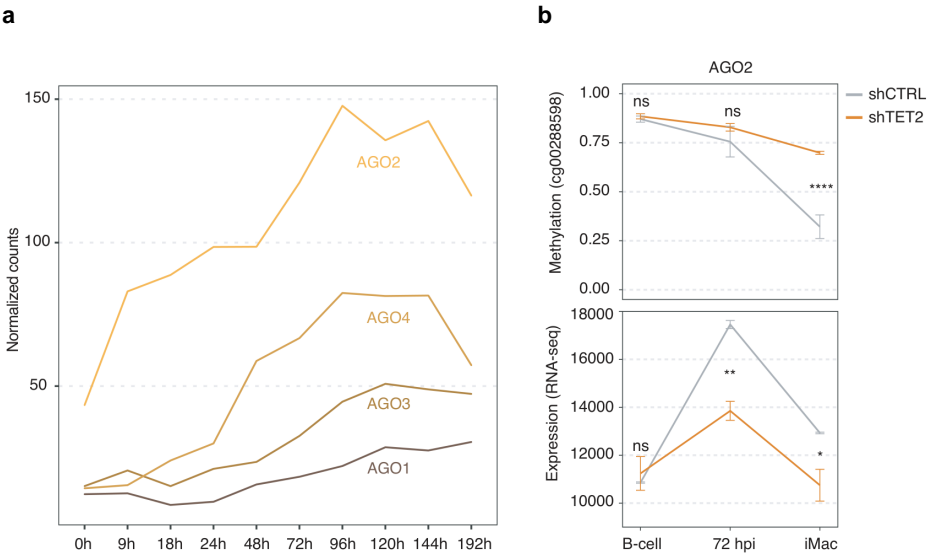

Figure R30. Evaluation of AGOs during normal and TET2 reduced levels. (a) Transcriptional kinetic of the Argonaute family members during TD. (b) Top: methylation levels at the detected AGO2 DMP during TET2 reduced TD. Bottom: abnormal transcriptional AGO2 kinetics during TET2 reduced TD. Unpaired two-tailed Student's t-test, mean s.e.m, (*p<0.05; **p<0.01, ****p<0.0001).

8.3.1 AGO2 epigenetic characterization during TD

A closer evaluation of the epigenomic status of the region surrounding the *AGO2* DMP suggested it might function as an intragenic enhancer (**Fig. R31, shaded region**). Notably, the region is extensively bound by *TET2*, but not exclusively (**Fig. R31 TET2 ChIP-seq**). Actually, we detected a total of 6 confident *TET2* peaks distributed along promoter and intronic regions of the gene, showing evidence of a TET2-mediated *AGO2* regulation.

Further evaluation of the candidate enhancer during TD, showed its occupancy by CEBPA and its chromatin opening from 24 hpi onwards. However, the region is only transiently marked with H3K27ac, which reached its maximum levels at 72 hpi and was reduced, but not lost, in iMac (**Fig. R31 middle**). Remarkably, the H3K27ac kinetics coincide with the previously described *AGO2*

transcriptional profile, strongly suggesting the role of this enhancer in regulating *AGO2* levels.

To determine whether our intragenic enhancer has a direct 3D connection with the *AGO2* promoter (E-P contact), we analyzed publicly available Hi-C dataset collected during TD [289]. The analysis showed an overall gain in interactions in the *AGO2* locus throughout TD, including forming a specific loop between the candidate enhancer and the TSS of *AGO2* in iMac (**Fig. R31 bottom**). Interestingly, other *TET2*-bound intragenic enhancers also interacted with *AGO2* promoter and intronic regions. Therefore, the 3D chromatin organization at the AGO2 locus might be mediated by DNAm changes.

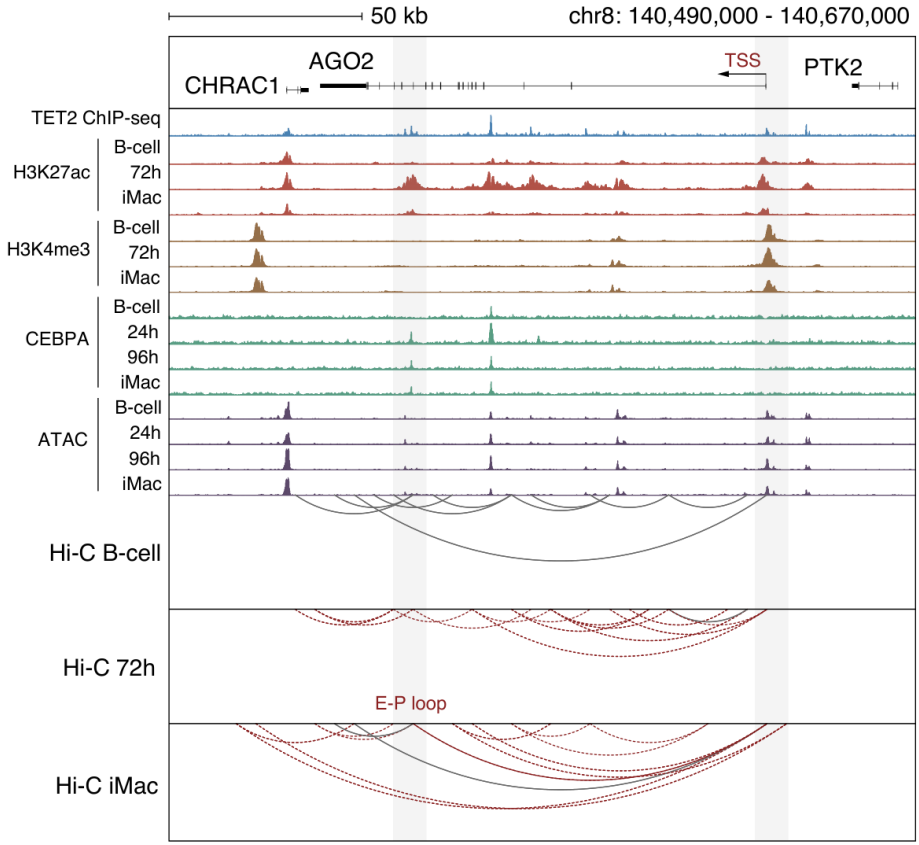

Figure R31. Snapshot of the AGO2 locus during TD. Top: histone marks, chromatin accessibility and CEBPA- TET2 binding sites. Bottom: Hi-C chromatin interactions. Only the interactions of at least 20kb and an interaction score > 1 are shown. Dashed red line indicates interactions exclusively detected at 72h and iMac. Solid red line at iMac stage shows looping between AGO2 promoter and candidate enhancer.

8.3.2 Refined DNA methylation profiling at AGO2 candidate enhancer

A thorough evaluation of the *AGO2* locus showed several small sub-enhancer regions (ER1 and ER2) of approximately 300bp within the region of interest that correlated with TET2 binding and methylation loss (**Fig. R32 top**). Importantly, we have noticed that only one methylation array probe (cg00288598) was found within the enhancer, highlighting the limitations of the arrays-based methylation evaluation.

To circumvent that and directly quantify the effect of *TET2* reduction on these important sites, we performed pyrosequencing of 3 CpGs for ER1 and 4 CpGs for ER2, at iMac stage. Of note, all the chosen CpGs are demethylated during normal TD (WGBS-seq), hence most likely affected by *TET2* alterations. Consequentially, we detected significantly higher methylation levels in *TET2*-reduced conditions in almost all of the residues (**Fig. R32 bottom**), suggesting a strong TET2 recruitment and catalytic activity at the region.

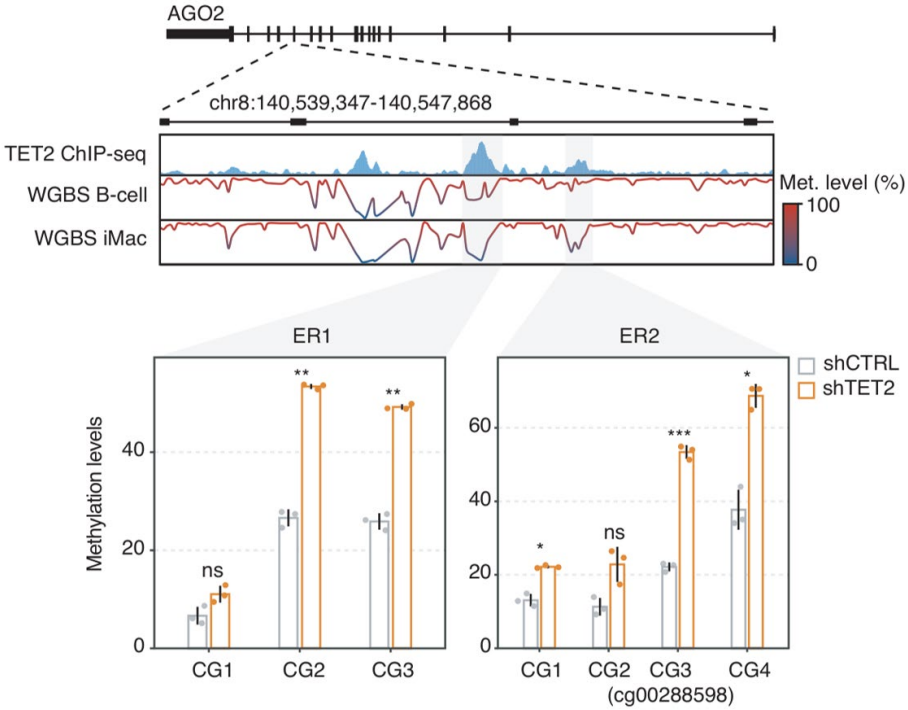

Figure R32. Top: Snapshot of AGO2 candidate enhancer. TET2 binding and DNAm profiles in B-cell and iMac are shown. Grey highlight corresponds to regions with methylation loss during TD. Bottom: pyrosequencing validation of the methylation status of selected CpGs at iMac stage in TET2 depleted conditions. Unpaired two-tailed Student's t-test, n=3, mean s.e.m, (*p<0.05; **p<0.01, ****p<0.001, ****p<0.0001).

8.3.3 Effect of AGO2 depletion during TD

Finally, to link AGO2 activity to myeloid lineage commitment, we evaluated the effect of direct *AGO2* depletion during TD. We have generated constitutive shRNA-AGO2 cell lines achieving strong depletion (up to 80%) as validated by RT-qPCR and WB at the B-cell stage (**Fig. R33a**). Although *AGO2* is already mildly expressed in B-cells, we have not detected any growth of cell death abnormalities upon AGO2 depletion at this stage (data not shown). Nonetheless, to avoid any compensatory mechanisms, *AGO2* knockdown cells were not maintained for an extensive amount of time in culture and were rapidly induced after antibiotic selection.

Conversely, we detected that reduced *AGO2* levels led to accelerated TD kinetics. Monitoring of CD19+ B-cell and CD11b+ macrophage markers showed significant differences at 72 and 168 hpi, exclusively found between the efficiently depleted shAGO2.1 and control cells (**Fig. R33b**).

These findings fall in line with the previous exploration of AML patients' data where low *AGO2* levels might be beneficial for proper terminal myeloid cell fate acquisition. However, the exact mechanisms by which TET2-AGO2 axis regulates the process are not yet well understood.

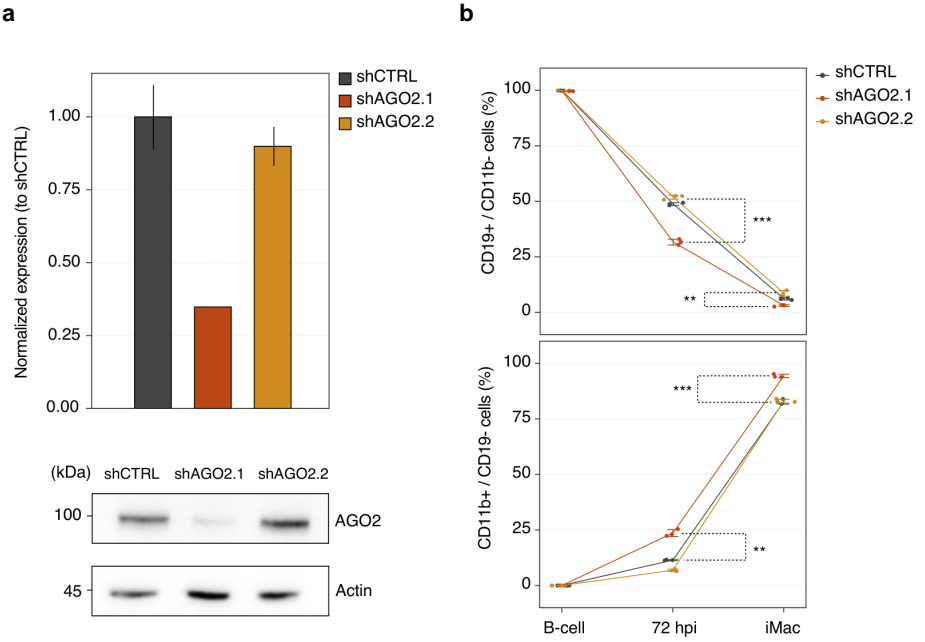

Figure R33. Evaluation of AGO2 depletion during TD. (a) Top: quantification of AGO2 relative expression in AGO2 depleted conditions. Bottom: Western blot showing protein depletion in shAGO2 conditions (BlaER-B-cells). (b) TD kinetics by flow cytometry in AGO2 depleted condition. Percentage of CD19+/CD11b− and CD11b+ /CD19− cells are shown. Unpaired two-tailed Student's t-test, n=3, mean s.e.m, ** p-value<0.01, *** p-value<0.001).

Chapter 3 |

Discussion

1. DNA methylation: its relevance and how can we study it

As already mentioned, and shown in the results of this thesis, DNA methylation correlates with gene expression and influences cell identity and tumorigenesis in mammalian cells [315]. However, it is still not clear whether the dynamic interplay between DNA methylation and chromatin-associated proteins is a driving force of cell fate decisions, such as during the emergence of pluripotent stem cells or terminal differentiation.

The elevated mutational burden of *DNMT3A* and *TET2* genes in blood alterations is not a coincidence but should be carefully evaluated. Although both genes are frequently mutated in myeloid malignancies such as myelodysplastic syndromes and acute myeloid leukemias (Woods Immunol Rev 2015), they are often found together with mutations on *NPM1* or *FLT3*. Which have been described to be the predominant leukemic drivers [316].

The presence of *DNMT3A* and *TET2* mutations in CHIP strongly suggest their role as drivers for mild clonal expansion in pre-leukemic conditions, which are exacerbated during chronic inflammation and emergency hematopoiesis [317]. Particularly, TET2-disrupted clonal hematopoiesis promotes expansion, myeloid bias, and out competition of wild-type cells [256,294,318].

Although recent findings have identified events that could be involved in TET2 recruitment to regulatory regions of genes involved in the aforementioned processes, it is highly desirable to acquire a broader knowledge of TET2 function during cell fate commitment chromatin rewiring. Which, when possible, should be tackled in a dynamic cell fate commitment setting and not merely at the final stages of leukemic blast arrest. As, most findings imply a greater DNAm influence during differentiation of hematopoietic stem and progenitor cell.

2. The BlaER system as a study model

In the present thesis, we have finely tackled TET2 cell commitment roles in a cell conversion model to address the outstanding questions related to active demethylation. Specifically, our findings provide an extensive characterization of TET2's role during the TD of leukemic B cells into non-tumorigenic macrophage-like cells [288] (**Fig. O3**). The system implies not only a highly reproducible cell rewiring but also the potential activation of TET2-mediated tumor-suppressing programs. Self-evidently, these findings could be highly prevalent in understanding leukemogenesis, always when appropriately integrated and validated in an in vivo setting.

Although the system is an excellent model for the study of DNA methylation dynamics, we are fully aware of the intrinsic limitations of the cells, such as the obvious lack of inherent complexity of *in vivo* systems. As well as the forced cell fate switch through an overwhelming translocation of abnormal and exogenous CEBPA protein levels to the nucleus, which does not mimic a naturally occurring terminal myeloid transition (HSCs-CMPs-Monocytes-Macrophages) (**Fig. I5**).

In fact, during the evaluation of early chromatin-related events (0-24 hp) in TD, we observed an abnormally high CEBPA genome occupancy, chromatin aperture, and histone mark deposition (data not shown). Suggesting how the wave of CEBPA completely shocks the B-cell's chromatin landscape. The CEBPA binding, although strongly associated with proximal and distal enhancer GREs, was mostly buffered at later stages of TD, with many sites lost at the iMac stage [319].

Importantly, CEBPA was described as recruiting TET2 to GREs for their demethylation [100]. However, those early events coincide with very low TET expression (**Fig. O3**). Hence, the possibility of protein interaction and consequential GRE demethylation during this somewhat atypical process is low. Particularly in such a short timeframe. Indeed, we barely detected any DNAm changes at 24 hpi [290]. Nonetheless, our analysis took into account all possible scenarios to evaluate the order of events that ultimately lead to transcriptional changes.

Regardless, we believe that the BlaER TD system is a unique and suitable model to investigate how TET2-induced genome DNA methylation changes relate to transcriptional control and cell fate transitions.

3. TET2 interactome evaluation during TD

TET2 is a large protein containing a conserved catalytic and an extensive disordered domain. It has the potential to interact with a wide array of proteins, which is most likely modulated by PTMs or mutations that occur along the full length of the protein or the gene, respectively (**Fig. I6**). However, the TET2 interactome has mainly been mostly studied in the context of the pluripotent state [228] and not using a high-throughput approach in myeloid cells, or at least not has been publicly released [239].

Several chromatin-related factors have been established as bona fide TET2 partners, including WT1 [233], CEBPA, KLF4 [100], HDACs [239] and OGT. The latter, for instance, boosts histone O-GlcNAcylation and gene activation through TET2 interactions independent of its catalytic activity [320]. Strongly supporting the potential non-catalytic functions of TET2 as a scaffold for protein complex formation.

3.1 Candidate partners evaluation

Through our TET2-Flag IP-MS/Spec data, we have identified a previously unrecognized potential involvement of TET2 in RNA biology. Notably, TET2-depleted cells exhibited changes in AS events during TD, suggesting a new implication of TET2 in post-transcriptional processes, which might or might not be methylation-dependent. We propose a novel role for TET2 in the biology of the spliceosome complex which might be crucial to maintaining a balanced alternative splicing landscape in development and disease [321,322].

Importantly, some previous findings have shown TET2 partnering with RNA binding proteins such as PSPC1, although mechanistically did not directly influence the regulation of mRNA splicing, but affected transcriptional destabilization through 5hmC modification [228].

Moreover, recent reports also suggest potential involvement in splicing mechanisms of other TET members, such as TET3. Which interactome in T-cells

was linked to abundant biological processes such as transcriptional regulation, RNA polymerase elongation and splicing [323]. However, our analysis has not shown an extensive overlap with the reported interactors, which were heavily enriched in SF3b complex members.

We found numerous members of the Serine/arginine-rich protein-specific kinase group (SRSF) or the elongation factor EFTUD2 (**Fig. R6**), which was validated in two different clones by Co-IP/WB (**Fig. R7**). Interestingly, we have detected that only a fraction of the total EFTUD2 protein interacted with TET2, indicating a possible supporting role of the protein within the spliceosome complex to regulate specific mRNA maturation events and not as a determining regulator of the process, which was in line with potential non-catalytic TET2 functions. Evaluation of TET2-ETFTUD2 genome (by ChIP-seq) and transcriptome (by RIP-seq) co-occupancy might give an idea about the extent of DNA methylation dependency, although the latter has not been explored yet.

Surprisingly, using the stringent cut-offs for significance recommended for analyzing IP-MS data, we have not detected any of the previously described TET2 interactors. Although some partner proteins detected, such as SAP18 (**Fig. R6**), clearly illustrate the known TET2 implication in HDACs complexes, other members were absent in the output. Among related partners, SAP30 and Sin3A but not SAP18 were detected to interact with all members of the TET family [187,324,325]. The precise role of TET2-SAP18 interaction in HDAC complexes is not fully clear. However, the most likely scenario is that affects the HDAC complex recruitment to active enhancers, such as in the case of TET2 - Sin3a interaction. The evaluation of potential methylation dependency was not further explored in our system but has been described to be methylation-independent, which highlights the importance of this analysis [325].

Additionally, the expected CEBPA-mediated TET2 interaction was also undetected, which was particularly interesting since we have observed large genomic co-occupancy of these proteins during TD (**Fig. R16**).

It is feasible that the “missing” factors do not partner with TET2 or do so at different stages of cellular conversion, although technical limitations could also be an explanation. Protocol optimization, such as different affinity-purification approaches or cross-evaluation with more sensitive/alternative techniques, such as TurboID-based proximity biotin labeling, could be advantageous to fully

uncover these discrepancies. Notably, a recent evaluation with the aforementioned technique found a mild enrichment for RNA binding and processing proteins among TET2 interactors, which was mediated by condensate formation through a specific low-complexity domain within the catalytic region of TET2 protein [326]. However, the findings were not discussed or further addressed in this manuscript.

3.2 Alternative splicing evaluation

Alternative splicing analysis in TET2 reduced TD conditions revealed the presence of significant alterations in iMac. Particularly, we have detected an increase in IR events in *TET2* knockdown cells (**Fig. R8**). The affected genes were related to cell cycle control and phagocytosis, further linking potential TET2 involvement in splicing events.

We have not found significant global changes in transcription or methylation at the PSI regions and the associated genes (data not shown), although thorough evaluation at intron and exon boundaries should be done in a gene-specific manner. For instance, we have detected up to four hypermethylated CpGs in the *EXT1* tumor suppressor large intronic region (272 kb), that presented intron retention in *TET2* knockdown cells. The gene is upregulated during TD and presents a high level of intronic reads at the detected region, suggesting the presence of alternatively spliced forms (data not shown). The exact influence of the differential DNAm on splicing is yet to be elucidated, making EXT1 an interesting candidate to evaluate for different splice variants in TET2 depleted conditions.

Importantly, AS events have been reported to be globally regulated through chromatin modulation, including nucleosome assembly, histone modifications, and CpG methylation [327,328]. Moreover IR, which appears to be mostly affected in our model, specifically has been linked to reduced CpG methylation levels [329]. For instance, increased 5hmC in exons has been found to enhance the inclusion of specific exons in mouse granulocytes and influence the expression of genes encoding regulators of macrophage transcription, phagocytosis, and inflammatory signaling in human monocytes and macrophages. This effect should be somewhat mirrored in BlaER-derived macrophages [329,330].

To confidently pinpoint the specific TET2 events during alternative splicing, a proper functional evaluation through deep coverage sequencing (ideally >150 million reads) and candidate AS events validation should be done. Additionally, it would be interesting to identify any preferential TET2 splicing facilitators (SR proteins) or repressor interaction (hnRNPs). And if this partnering affects core spliceosome activity [331]. Consequentially integration with cells methylome to interrogate atypical DNAm sites, such as intron-exon boundaries, should give a proper idea about catalytic dependency during the process.

4. Evaluation of TET2 chromatin targets during TD

The evaluation of TET2 chromatin targets is crucial to directly pinpoint its role in catalytically active regions.

Most of the reported datasets have been generated in mice where Tet2 has shown preferential enhancer binding, which facilitates transcription factor recruitment in normal and malignant hematopoiesis [295]. In human cells, *TET2* chromatin binding was also reported in monocyte-derived macrophages, done with a home-made antibody, which related TET2 to mitochondrial DNA-mediated interferon signaling [332]. Also, in a human breast cancer cell line where TET2 was described as responsible for estrogen receptor DNA methylation-mediated regulation, highlighting the role of TET2 in non-hematopoietic malignancies [299]. Regardless, the proteins genome-wide chromatin binding landscape during CEBPA mediated myeloid commitment, has never been described and/or integrated with the epigenetic profile of the cells. Hence the need to address this to better understand healthy and malignant myeloid development.

Our high-quality TET2-Flag ChIP-seq data has identified over 4000 high-confidence chromatin TET2 targets, which are strongly associated with GREs of myeloid and immune-related genes (**Fig. R10**).

4.1 TET2 binding at chromatin rewiring sites

Firstly, the presence of TET2 at gene regulatory regions suggested its possible role in modulating chromatin interactions. In fact, we detected that TET2 binding is associated with a large-scale rewiring of the chromatin organization (**Fig. R11**). However, the exact TET2 function in the process is not clear, as we have

not found any significant differences in DNAm stable or changing regions, suggesting possible non-catalytic implications (data not shown). This falls in line with a genome-wide TET2 GREs binding but with limited demethylation consequences (**Fig. R14a**).

One of the explanations could be related to chromatin remodeler PRC2, which can enhance long-range chromosome folding [333,334]. TET2-PRC2 interaction has been previously described, hence, it could potentially lead to cooperation between these factors during chromatin rewiring [335]. Other recent studies have linked CEBPA-bound regions to the formation of long-range 3D chromatin hubs in the BlaER transdifferentiation model. The process was mediated through phase separation, leading to the formation of condensates [319], where CEBPA and TET2 could have cooperative functions. The latter was further reassured by recent findings showing the TET2 low-complexity insert domain involved in biomolecular condensation [326].

Importantly, recent findings also link TET-mediated DNAm dynamics to chromosome organization. In Tet-triple knock-out ESCs, reduced compartmentalization through hypermethylation in TAD boundaries and chromatin loop anchors has been observed, which impairs CTCF binding [336]. However, the exact role of TET2 in the process has not been addressed, nor has the presence of these regulatory events in myeloid commitment.

4.2 Genomic classification of TET2 chromatin targets

Having access to an extensive chromatin characterization library in BlaER cells, we fully integrated TET2 binding sites with a collection of datasets of histone marks, chromatin accessibility, DNA methylation, and chromatin interactions. This has allowed us to establish a thorough and biologically significant TET2 chromatin binding context.

Our findings indicated strong TET2 catalytic activity at distal regulatory elements of myeloid genes. We identified methylation dynamic enhancers that, upon TD, undergo an increased chromatin aperture, deposition of active enhancer marks (H3K27ac + H3K4me1) and CEBPA binding (**Fig. R16**). These events ultimately led to full enhancer activation (detected by TT-seq) (**Fig. R15**) and associated gene upregulation (detected by RNA-seq) (**Fig. R14**).

DNAm is most likely a “late” mechanism in the process, as the protein does not greatly accumulate until 72 hpi (**Fig. O3**). Most TET2-bound active GREs present early activation (24-48 hp) (data not shown), so DNA demethylation seemingly occurs after chromatin aperture and histone deposition events. This might be necessary for proper transcription factor recruitment, such as gene activators or chromatin modulators responsible for maintaining the active state of the regions. Importantly our experimental approach cannot distinguish between 5mC and 5hmC and therefore you cannot exclude an early 5hmC peak at certain regions. Previous findings reported synchronous and preceding methylation changes to chromatin opening events during reprogramming [100]. Hence, we would need to generate high coverage bisulfite and ox-bisulfite sequencing (BS-seq/oxBS-seq) at early TD timepoints to address this possibility.

Interestingly we have observed a relatively homogeneous distribution between enhancer (59.2%) vs promoter (40.8%) regions, in contrast to the previously reported TET2 bias towards distal regulatory elements [295]. Most promoters contain lowly methylated CGIs that do not present any DNAm changes during cell commitment. Consequentially, TET2 recruitment to stable and devoid of methylation promoter regions, is not fully understood. Several studies suggest that the CXXC domain, may be involved in the recruitment of Tet1, Tet3 and IDAX to CGIs [174,337]. However, CXXC is absent in TET2*,* which suggests some other recruitment mechanism.

On one hand, the accumulation of TET2 might lead to formation of transcriptionally active complexes at promoter sites through methylation independent functions, or as a fail-safe component against abnormal methylation by DNMT3A. Which although excluded from CGI due to their chromatin structure (by H3K4me3) might still lead to CGIs hypermethylation in disease or stress conditions [338]. On the other, the detected peaks might correspond to enhancer-promoter chromatin looping, where active demethylation at distal regulatory elements might favor the interaction leading to TET2 detection at both regions. Regardless the localization of TET2 at gene regulatory elements indicates its importance in transcriptional modulation.

Collectively, our findings provide a thorough chromatin characterization of TET2 binding sites during myeloid fate commitment. We highlight the relevance of

GREs dynamics and particularly the importance of CEBPA-TET2 co-occupancy at biologically relevant enhancers that lose methylation during TD. This is crucial for robust GRE activation through chromatin opening and deposition of activating histone marks and, finally, upregulation of myeloid commitment-related genes. Although extensively explored during this thesis, this data provides a valuable resource to further interrogate TET2 methylation-dependent and independent function, such as the influence of TET2 on chromatin rewiring, RNA processing, or condensates formation. Importantly, these findings could be reliably replicated in the future thanks to the robustness of the TD model.

Of note, exploring TET2 chromatin binding along different TD timepoints could provide improved temporal characterization. Our attempts to profile TET2 binding sites at 168 hpi failed due to technical limitations. Although it is possible that TET2 chromatin distribution changes along TD, our preliminary data do not support this hypothesis, as the detected 72 TET2 binding sites were part of the 168 hpi sites (data not shown).

The latter is further supported by our WGBS-seq and EPIC array methylation data, which indicate the presence of progressive DNA demethylation events from intermediate (72 hp) to later stages of TD (168 hp). Hence, the binding rewiring seems improbable, as the maintenance of TET2 at the target regions appears to be requiered to achieve the necessary decrease in DNAm.

5. *TET2* knockdown effect on methylome and transcriptome during TD

Based on the literature, as extensively described in the introduction, DNA methylation changes should be mostly paired with transcriptional modulation. To that extent, we have also interrogated the methylation and expression landscape during TD in TET2-reduced conditions.

5.1 Evaluation of knockdown efficiency

Despite our efforts to establish a fully inducible TET2 knockdown cell model, we have mostly achieved a moderate *TET2* reduction of approximately 50% (**Fig. R1**). Altogether, the explored strategies to improve the reduction efficiency (see 5.1) targeted different system components with very limited success. The detected limitations are most likely related to the biology of the cellular model,

which may buffer the effect of the inducible mechanisms in some way. The evaluation of other inducible silencing systems such as shRNAmiR [339,340] or degron tags (SMASh, dTAG, HaloTag) might provide a feasibly implemented alternative to the current working pipeline. However, these have not been further explored in this context.

Of note, a consistent 50% mRNA *TET2* reduction could yield more biologically relevant results. Myeloid malignancies' patients and individuals with CHIP have a predominant *TET2* haploinsufficiency [221], which is sufficient to establish hypermethylation patterns [296]. Indeed, the moderate *TET2* depletion might better phenocopy in vivo leukemogenic or clonal expansion events, which provide the base for modest transcriptional changes sufficient for an improved cell fitness.

During disease development, *TET2* mutations by themselves, do not lead to leukemic transformation but are typically present in concert with other alterations such as *FLT3* [341]. Which further highlights the idea of mild DNA hypermethylation as a mean for cell expansion and acquisition of additional mutations, which are also typically associated with age [342].

Several cases of bi-allelically mutated *TET2* in AML and MDS have been reported to influence leukemogenesis and treatment resistance. This is particularly relevant in the subset of patients treated with hypomethylating chemotherapy [343,344]. Additionally, CMML patients have been reported to frequently present *TET2* biallelic alterations. This is not surprising given the extremely high *TET2* mutational rates in this neoplasia, which are not yet fully understood [345]. The evaluation of mild vs strong hypermethylation within variable *TET2* mutational states might provide useful information regarding protein distribution, activity, and transcriptional deregulation. It could also help us understand potential compensatory mechanisms by other TET family members. However, this aspect has not been explored in this work.

5.2 Evaluation of TD efficiency in TET2 reduced conditions

Our preliminary evaluation of TD efficiencies by FACS in TET2 reduced conditions showed very minimal differences. Both shCtrl and shTET2 lines transitioned to iMac at the same rate with no deviations at intermediate timepoints (**Fig. R2**).

On the one hand, the relatively mild reduction of TET2 achieved is unlikely to significantly perturb the transcription and methylation networks enough to prevent cell conversion. However, relying on only two surface markers provides a limited understanding of the molecular changes within the cells. In fact, previous reports suggest that alterations in important transcription factors, such as *CTCF*, also do not greatly affect the TD process but can lead to extensive chromatin changes that affect crucial macrophage functions, such as responding efficiently to external stimuli [289].

Accordingly, a high-throughput characterization with EPIC methylation arrays and RNA-seq was pursued.

5.3 Discrepancies in TET2 depletion efficiencies

Importantly, we have observed some discrepancies in TET2 reduction efficiencies by RNA-seq. While *TET2* followed a strong upregulation tendency and was detected as a significantly downregulated gene at all time points, we found that the extent of doxycycline treatment only led to approximately 30-40% reduction (data not shown), compared to previously detected 50% by RT-qPCR (**Fig. R18**).

This might be related to technical errors during library preparation, as the same RNA samples were used to validate TET2 reduction by RT-qPCR prior to sequencing. Nonetheless, even relatively limited *TET2* depletion appears to have a profound effect on the methylation and expression landscape. We observed extensive changes along the TET2 reduced TD, primarly affecting the iMac stage and, to a lesser extent, the intermediate 72 hpi time point.

5.4 Evaluation of DNAm landscape in TET2 reduced TD

DNAm characterization (by EPIC arrays) revealed a progressively increasing hypermethylation phenotype along the TET2-reduced TD, compared to shCTRL TD (Number of hypermethylated DMPs; B-cell: 402, 72hpi: 742, iMac: 1794) (**Fig. R19**). We clustered all DMPs to evaluate the dynamics of these changes, highlighting two differentially methylated subsets (M2 and M3) between TET2-reduced and shCTRL TD (**Fig. R20**).

Both clusters (M2 and M3) lost methylation during TD and were significantly hypermethylated in *TET2* knock-down conditions. Importantly, only these

subsets were enriched in myeloid and macrophage-related processes, compared to the M1 and M4 groups, which have shown stable low and high DNAm levels, respectively (**Fig. R20**). Although M1 and M4 were also slightly hypermethylated in *TET2* knock-down conditions, no clear association with any biological process was established, and no specific interesting gene candidates for additional validation were detected.

Interestingly, the M2 and M3 clusters presented variable degrees of demethylation during normal TD. Being mildly and strongly demethylated, respectively. These findings suggest that genes crucial to myeloid cell fate programs in M2 might only need mild demethylation for gene activation. Meanwhile, genes related to stimuli response and intrinsic macrophage functions in M3 follow almost complete demethylation later in the TD process, possibly to fine-tune the functions of fully mature macrophages and avoid abnormal immune responses from immature cells.

Unexpectedly, some changes were also found at the B-cell stage, where TET2 levels are low. Most of the changes were mild and were not associated with any particular biological process. However, it is possible that reducing *TET2* activity in B-cells, even at low levels, might lead to specific demethylation events. Further exploration might require a complete *TET2* depletion and a map of B-cell chromatin distribution to confidently evaluate the aforementioned possibility.

Finally, a small hypomethylated subset of CpGs was also found at all time points but was mostly regarded as a biologically irrelevant event. These DMPs presented generally low changes and did not correspond to DNAm dynamic regions [290] or with TET2 binding sites (data not shown). It is possible that the previously described TET2 and DNMT3A interactions [346] could somehow lead to abnormal DNMT3A targeting during TET2 reduction. Some studies suggest that TET proteins might be required, directly or indirectly, for optimal DNMTs activity. Particularly in heterochromatin, where TET deficient cells present hypomethylated profiles contrarily to the expected hypermethylation, which was preferentially found in euchromatin regions. These events were apparently stochastic but led to transposable elements re-activation which might be related to oncogenic transformation [347]. The possibility of these events could be

further investigated in our model by genome wide methylome and chromatin interactions profiling in TET2 depleted conditions.

5.5 Evaluation of the transcriptional landscape in TET2 reduced TD

Likewise, the evaluation of the transcriptome rewiring (by RNA-seq) in TET2 reduced TD presented similar patterns to DNAm changes. Cells accumulated transcriptional alterations along TD, culminating in highly affected iMacs with 1,222 downregulated genes mostly involved in myeloid and macrophage-related processes (**Fig. R22**).

Evaluation of upregulated DEGs, has not been significantly associated with any biological functions, although they included immune cell adhesion genes (such as *MCAM*, *ITGA8*, *CD2,* or *ID1)* and other immune-related genes such as *SLAMF6* or *NOD2* (**Fig. R22**). The latter is particularly relevant in NK and T-cells, whose development and function have been linked to TET activity [193]. However, this upregulation was not associated with any DNAm changes in our model (data not shown). It is possible that these transcription modulation mechanisms are methylation-independent or are simply regulated by distal elements not directly associated with the affected genes. How exactly *TET2* depletion leads to these genes' upregulation, has not been further explored in this study.

Notably, since *TET2* is barely expressed, we did not expect transcriptional changes at the B-cell stage; however, some limited differences were found. The detected DEGs at the B-cell stage appear to be related to spontaneous activation of a small number of myeloid gene, which led to some significant DEGs between cell lines. The latter most likely is a biological anomaly or technical error, and if it had any effect, it was not taken into account for downstream evaluation.

5.6 Evaluation of TET2 role in DNAm mediated transcriptional modulation

Finally, we interrogated the methylation-expression axis to pinpoint DNAm-dependent genes. To better understand how TET2 might influence the transcriptional dynamics we wanted to give a global overview of the TD process in normal and reduced TET2 conditions.

We have clustered the most dynamic genes along TD and identified the expected silencing (B-cell genes) and activating (myeloid genes) subsets (**Fig. R23**). These were interrogated for methylation and expression changes upon TET2 reduction. The findings showed that TET2 reduction mostly affected the activating myeloid clusters (E4 and E5), which presented abnormal methylation loss and impaired gene accumulation during TD. They included crucial myeloid differentiation genes such as *IFNGR2* or *MAF* (**Fig. R24**)

These results align with the previously known idea of TET2 functions during myeloid cell fate commitment, which is the demethylation of GREs for transcriptional activation. Our dynamic analyses allowed us to confidently pinpoint TET2's influence on transcriptional modulation coupled to DNA demethylation along myeloid cell fate commitment.

6. Evaluation of TET2 chromatin targets in healthy and malignant hematopoiesis

It is crucial to note that our findings regarding TET2 function during TD need to be faithfully cross-examined in a more significant biological system. The latter was addressed to some extent with the evaluation of methylome profiles from the Blueprint consortium.

Applying our findings from the BlaER model, we could evaluate the DNAm dynamics at the TET2 bound sites during hematopoiesis in detail. To that extent, we integrated healthy and AML patients' datasets. We classified the patients according to low (Hypo-AML) or high (Hyper-AML) global DNAm levels. Unfortunately, we did not have any genotype data regarding the profiled AML patients. Hence, the influence of typically mutated *DNMT3A* or *TET2* was not precisely evaluated.

Regardless of the integration of healthy samples, AML patients and BlaER cells showed remarkable results. We have observed clear DNAm trajectories along normal hematopoiesis, which were abolished in AML patients, where the blasts were more or less differentiated according to their classification into Hypo and Hyper-AML (**Fig. R29**). Notably, a similar commitment delay was found in TET2-reduced TD, showing the high applicability of our findings for AML patients' evaluation.

Importantly, we used a subset of dynamically demethylated probes from the EPIC arrays (n=224) to perform all comparisons. Although a relatively small number, it seems to provide confident sites that get demethylated during myeloid differentiation both *in vivo* and *in vitro* settings. The evaluation of genome wide DNAm profiles, which are mostly stable, resulted in decreased clustering by PCA, suggesting global differences between the BlaER and primary samples methylome (data not shown). Which although surprising could be related to, cross-examination of DNAm levels from WGBS-seq data (primary samples) and EPIC arrays (BlaER cells). Which could be homogenized through a genome-wide DNAm profiling of TET2 depleted TD (by WGBS).

7. TET2 targets identification and evaluation

Altogether, we incorporated our findings into publicly available data from AML patients and identified novel gene candidates that might influence leukemic maintenance and progression in a TET2-dependent manner.

7.1 TET2 candidates' identification

To provide a better *in vivo* assessment of our findings, we analyzed DNA methylation profiles (by EPIC arrays) from TET2 mutated AML patients. The results showed strong hypermethylation (3,592 DMPs) which was related to myeloid cells functions, particularly for migration and adhesion (**Fig. R23**).

Although biologically relevant, we have not found a clear enrichment for cell commitment or differentiation pathways as detected in our model (**Fig. R20**). Expectedly, the overlap between the datasets showed limited results. We detected only 19 hypermethylated DMPs (data not shown) shared among the 3,592 and 2,937 CpGs from AML patients and the BlaER system, respectively. Despite the low number, the obtained results were considered very confident events and were related to some interesting genes, which were thoroughly evaluated.

For instance, we overserved an overlap in a CpG associated to the proliferation-related gene GAS7. This CpG showed 25% and 20% increase in methylation in AML patients and TET2 depleted iMac, respectively. Upon closer evaluation of the hypermethylated site in the cellular model, it was discovered that it is located within an intragenic enhancer. This enhancer is bound by TET2-CEBPA and

undergoes demethylated after induction. In addition, it exhibited extensive deposition of H3K27ac and displayed chromatin accessibility typical of an active enhancer during TD.

However, we did not observe transcriptional changes in GAS7 upon TET2 reduction. The latter suggests a lack of methylation-transcription dependence at this specific GAS7 element. The most feasible explanation is that this enhancer simply does not control the expression of the associated gene in our cellular model. A closer look at the H3K27ac distribution indicates that the candidate site is located within a broad region covered by the histone mark, as well as other TET2 and CEBPA peaks. These patterns are typical of "super-enhancers" (Pott Nat Genet 2014), known for controlling multiple genes, which could be this particular case.

7.2 Evaluation of AGO2 as a direct TET2 target in BlaER system

The integration of all datasets allowed for the identification of an *AGO2* locus, which fulfilled all the requirements to be considered as a bona fide TET2 chromatin target (see 8.1.1) (**Fig. R26**).

Importantly the protein has been extensively described in disease. For instance, low *AGO2* was linked to tumor progression in breast cancer patients and childhood ALL [348,349]. Meanwhile, during acute inflammation, such as in COVID-19 patients, AGO2 endonuclease and small RNA-binding functions seem to be necessary for proper antiviral activity [350]. The mechanisms underlying all these contrasting AGO2 functions remain unknown.

Very limited reports also indicate implication of AGO2 on myeloid cancer cells, with some findings suggesting AGO2 oncogenic role for myeloid leukemia maintenance [351,352]. Therefore, the role of the protein in myeloid malignancies remains to be fully elucidated, hence it was further explored.

In BlaER cells the identified region was classified as a TET2-bound enhancer that started to lose methylation +72 hpi and presented typical activation traits such as H3K27ac decoration and slight chromatin aperture, in addition to early CEBPA binding (24 hp) (**Fig. R31**). Interestingly, within the region, we have found three TET2-bound sites in close proximity to each other (approx. 1000

bp) (**Fig. R32**). All of them lost DNA methylation during TD, although to different extents. The exact reason for this pattern is not clear.

According to the H3K27ac mark, all three sites form part of the same GRE but still display some distinct properties. One subregion coincides with accessible chromatin and CTCF binding, while the other is bound by CEBPA. Meanwhile, the most downstream one only has DNAm changes. Regardless, the region presents clear enhancer traits.

Interestingly, the described three-peak distribution contains a small exon between the two peaks. Specifically, the demethylating sites ER1 (intron 14) and ER2 (intron 13), encompass *AGO2* exon 14. Such a specific pattern hints towards strict regulation at the region and could be related to splicing mechanisms. Some recent findings have described non-catalytic functions of DNMT3A at intron-exon boundaries to coordinate splicing [353], which could also be possible for TET2. However, in this case, clear TET2 catalytic activity is detected. In fact, it was validated by pyrosequencing that showed significant hypermethylation of both ER1 and ER2 sub-regions in TET2-reduced conditions (**Fig. R32**). The evaluation of this particular TET2 binding pattern to encompass small exons might be an interesting event to evaluate in future analyses.

Overall, any differential TET2 binding along such a small region seems to be improbable but would need to be further dissected to make any claims. For instance, selected sub-region deletion with CRISPR tools might give a better idea about the importance of each site for AGO2 transcriptional modulation. Additionally, evaluation with DNAm editing tools already set up in our lab [290] could establish a direct transcription-DNA methylation axis to better understand the interplay between the two layers during TD. However, in this case, DNAm modulation could be challenging due to the close proximity of each sub-region.

Finally, we used publicly available Hi-C data during TD [289] to interrogate possible interactions in the *AGO2* locus. Indeed, we have detected direct candidate enhancer contact with the *AGO2* promoter site, specifically in iMac (**Fig. R31**). Confirming a transcriptional regulatory role for this enhancer. Additionally, looping with nearby Chromatin Accessibility Complex Subunit 1 (*CHRAC1*) gene was also detected, although that did not lead to *CHRAC1* expression changes in our system, but might still be relevant in other biological settings. No other distal contacts were found.

Of note, we have also identified four other confident TET2-bound regions along and in proximity of *AGO2*. However, none presented clear demethylation loss during TD (data not shown). However, all of them displayed similar chromatin properties to the candidate regions and looped among them and with the *AGO2* promoter (**Fig. R31**). This strongly suggests a strong chromatin rewiring at the region, which could be affected by TET2 binding. The latter, is presumably mediated by protein complexes including CEBPA and SPI1, which are enriched in MDRs (**Fig. R16a**). Interestingly evaluation of AGO2 ER1 showed enrichment for the same TFs but ER2 was strongly enriched for AP-1 dimeric transcription factor (data not shown). Nonetheless all of them are known TET2 interactors (**Fig. I7**) and important myeloid commitment regulators [354–356]. And could collaborate for *AGO2* locus chromatin looping and transcriptional activation.

7.3 Evaluation of AGO2 as a direct TET2 target in AML

Very little is known about the implication of *AGO2* in myeloid commitment and AML. To that extent, we interrogated the TCGA-LAML dataset and found significant *AGO2* overexpression and a high hazard ratio of overall survival in AML patients (**Fig. R27**). These findings suggest that AGO2 functions as an oncogene in myeloid malignancies. This is in line with a previous report that showed myeloid leukemia cells' dependence on *AGO2* expression [351].

To better understand the TET2-AGO2 axis in a physiological setting, we took advantage of the Blueprint dataset. *AGO2* evaluation in previously defined Hyper vs Hypo AML samples showed drastic DNAm differences. We have found that the *in vitro* detected enhancer, as well as other TET2 bound regions, were differentially methylated (**Fig. R27**). The latter strongly suggests the presence of a tightly DNAm-controlled locus that gets demethylated during myeloid cell fate acquisition. Unfortunately, we do not fully understand the influence of this event on *AGO2* expression levels.

Although hypermethylation seems to silence gene expression in our cell model and to a limited extent in AML samples (**Fig. R27**), this does not fit the idea of *AGO2* overexpression in AML. One of the possible scenarios seems to be a transient upregulation, similar to the one observed during transdifferentiation (**Fig. R30**). In healthy myeloid commitment conditions, *AGO2* locus demethylation might lead to transcriptional upregulation which is then

attenuated by other mechanisms. In *TET2* mutated cases, hypermethylation could impair this accumulation and ultimately lead to myeloid network deregulation, which somehow upregulates *AGO2*.

Another hypothesis could be related to unconventional TET2 functions, such as demethylation-mediated *AGO2* attenuation. Low *AGO2* levels might be necessary for terminal myeloid differentiation, where TET2 is responsible for demethylating the locus, leading to the recruitment of silencing machinery. This would be impaired in TET2-depleted conditions, where *AGO2* would be maintained at high levels, limiting the cell's capability to fully differentiate.

Importantly, this also does not address the AGO2-mediated miRNA-mediated gene silencing. Unfortunately, the evaluation of the AGO2-related miRNA network was not further pursued, being out of the scope of this thesis. However, future evaluation through high-throughput techniques such as miRNA-seq could provide a thorough network of *AGO2* activity.

7.4 Evaluation of AGO2 depletion during TD

Finally, to better understand the influence of *AGO2* in our cell model, where we detect confident methylation-expression interplay, we performed *AGO2* knockdown experiments. The findings revealed faster TD kinetics in AGO2-depleted conditions (**Fig. R33**). This falls in line with the idea of the necessity to downregulate *AGO2* for terminal myeloid differentiation and correlates with higher survival rates of patients with low AGO2 levels.

Evaluation of the transcriptional landscape (by RNA-seq) in AGO2-depleted conditions should give a better idea about its relevance in myeloid gene activation. This could also be approached by AGO2 overexpression, which, in theory, should limit myeloid commitment.

Altogether, there is also a need to address the TET2-AGO2 axis in *in vivo* experiments to confidently pinpoint the relevance of both protein alterations in the cell's leukemic capabilities. Currently, we are working on establishing the best experimental conditions to address this question in AML patient-derived xenograft mouse models, which will be subjected to separate and complementary *TET2/AGO2* depletion.

Main findings and conclusions

1. TET2 interactome during TD suggests novel protein implication in RNA biology through alternative splicing modulation

2. TET2 binding sites experience long-range chromatin remodeling during TD

3. TET2 extensively binds to myeloid cell fate programs GREs during TD and leads to active enhancers demethylation favoring transcriptional activation

4. TET2 depletion during TD leads to DNA hypermethylation and impaired myeloid genes upregulation

5. Integration of TET2 findings with an extensively characterized BlaER chromatin landscape allows to identify high-confident TET2 chromatin targets

6. Evaluation of publicly available normal and malignant hematopoiesis data shows similar DNA methylation tendencies between BlaER and *in vivo* myeloid commitment.

7. Overlap with publicly available $TET2^{MUT}$ AML patients methylome identifies an AGO2-related enhancer as a direct TET2 target.

8. *AGO2* locus gets extensively demethylated by TET2 during *in vivo* myeloid commitment

9. AGO2 depletion during TD leads to an accelerated myeloid cell fate acquisition

Closing remarks

Mammalian development is an infinitely complex process that incredibly guides the evolution of a zygote into a complex, multicellular adult organism. New cell types and tissues rapidly emerge and evolve with remarkable speed and robustness, which unsurprisingly involves a complex regulatory network in which players constantly interact, compete, and interplay in a highly coordinated fashion. The orchestrated developmental networks in each cell ensure, to the best of their ability, a balanced multi-omics landscape. Although unsurprising, this balance is often shifted to variable degrees. Intracellular and extracellular stress conditions lead to cell adaptation within its regulatory niche and in symbiosis with its surroundings. When irreparably damaged, cells lead to a surge of downstream events that in time manifest as disease.

An array of specific cellular processes affecting chromatin, transcription factors and genes themselves ultimately result in modulation of transcriptome-proteome networks. Among the regulatory mechanisms, epigenetic layers have been uncovered and thoroughly explored during the past decades. Epigenetic dynamics in development and disease, were fruitfully complemented by high-throughput molecular survey advances, that keep evolving at a remarkable pace. Molecular interrogation of histone modification, chromatin remodeling, noncoding RNA and DNA methylation led to exceptional advances in discovery of underlying molecular events in normal and altered development which ultimately led to the development of chemical compounds, such as histone deacetylase or DNA methylation inhibitors, which are widely used in different clinical settings. Undoubtedly, there is a major interest in studying the role of epigenetic mechanisms, regardless the perspective.

In fact, the main objective this thesis work was to provide new insights into the everchanging influence of DNA methylation in human biology. Despite the constantly growing findings on the process each year, it still remains unclear how precise events and distribution across the genome, influence cell's biology in a physiologically relevant manner. Discrepancies across differentiated, pluripotent and cancer cells have been under extensive evaluation in order to determine possible correlation with disease initiation and/or progression. To achieve that we need to question the ingrained concepts of promoter –

enhancer silencing as a sole consequential event in gene silencing. Particularly when evaluating tightly dynamic and adaptable systems such as adult hematopoiesis where temporal windows for global and/or specific DNA methylation events might be crucial for correct cell fate acquisition.

Altogether, our unbiased chromatin target identification offers a unique opportunity to investigate the relevance of DNA methylation events during myeloid differentiation. We have shown the importance of TET2 in targeting gene regulatory elements of crucial myeloid genes that get abnormally hypermethylated and downregulated upon TET2 loss of function, as seen in *TET2*-mutated AML patients. This underscores the need to investigate our bona fide TET2 targets, such as *AGO2*, for potential new therapeutic approaches to address myeloid malignancies in patients.

Appendix

Table 1: Vector construction oligonucleotides sequences

1.1 pLKO-Tet-On-shTET2 sequences

Name	Forward (5' to 3')	Reverse (5' to 3')
shCTRL	CCGGCACTTACGCTGAGTACTTCGACTCGAGTCGAAGTACTCAGCGTAAGTGTTTTTG	AATTCAAAAACACTTACGCTGAGTACTTCGACTCGAGTCGAAGTACTCAGCGTAAGTG
shT2.1	CCGGTTTCACGCCAAGTCGTTATTTCTCGAGAAATAACGACTTGGCGTGAAATTTTTG	AATTCAAAAATTTCACGCCAAGTCGTTATTTCTCGAGAAATAACGACTTGGCGTGCTG
shT2.2	CCGGCAGTCTAATGTACGAACTTTACTCGAGTAAAGTTCGTACATTAGACTGTTTTTG	AATTCAAAAACAGTCTAATGTACGAACTTTACTCGAGTAAAGTTCGTACATTAGACTG
shT2.3	CCGGCAGATGCACAGGCCAATTAAGCTCGAGCTTAATTGGCCTGTGCATCTGTTTTTG	AATTCAAAAACAGATGCACAGGCCAATTAAGCTCGAGCTTAATTGGCCTGTGCATCTG

1.2 pSicoR-AGO2 sequences

Name	Forward (5' to 3')	Reverse (5' to 3')
shCTRL	TCCTAAGGTTAAGTCGCCCTTTCAAGAGAAGGGCGACTTAACCTTAGGTTTTTTC	TCGAGAAAAAACCTAAGGTTAAGTCGCCCTTCTCTTGAAAGGGCGACTTAACCTTAGGA
shA2.1	TGCAGGACAAAGATGTATTATTCAAGAGATAAT	TCGAGAAAAAAGCAGGACAAAGATGTATTATCTCTTG

	ACATCTTTGTCCTGCTTTTTTC	AATAATACATCTTTGTCCTGCA
shA2.2	TGGATGTCCATTTCGAAGAATTCAAGAGATTCTTCGAAATGGACATCCTTTTTTC	TCGAGAAAAAAGGATGTCCATTTCGAAGAATCTCTTGAATTCTTCGAAATGGACATCCA

1.3 pMK-TET2-3xFlag

Name	Forward (5' to 3')	Reverse (5' to 3')
F1_backbone	TGGTAATCTAGAGGTGGGGGAGCTCCAATTCGCCCTA	AATAAACGTATATATACCCAGCTTTTGTTCCCTT
F2_Hyg_HA5	GAATTGGAGCTCCCCCACCTCTAGATTACCACCCAA	GTGAGTTCAGGCTTGCGAAACGATCCAGGTCCAGGG
F2_Neo_HA5	GAATTGGAGCTCCCCCACCTCTAGATTACCACCCAA	GGTAACTAGTGAACCTCTTAAACGACGGCCAGTGAACGACGG
F3_HA3	AGAGGTTCACTAGTTACCTCACTTGAAAAGACCA	CAAAAGCTGGGTATATATACGTTTATTGTCCTACAAGG

Table 2: pMK-TET2-3xFlag genotyping primers

Name	Forward (5' to 3')	Reverse (5' to 3')
Out_screen	CATGAAACTTCAGAGCCCACTTAC	AGATAACCTCTTTTGTTGCTGGTG
In_screen_1	AGTGGCCGTGGCTCCAACTCATG	CCATCTGTGACCACTTGGGAGCCAGTG

In_screen_2	GGTCTCTGAGCGGAG GAGGTTC	GAGGCTACAGTCAGTG GAGAGGAC

Table 3: WB and immunoprecipitation antibodies

Primary antibodies:

Protein / Epitope	**Reference**
TET2	Abcam (ab124297)
AGO2	Cell Signaling (C34C6)
EFTUD2	Bio-Rad (VMA00890)
FLAG M2	Sigma (F1804)
Tubulin	Abcam (ab6046)
Actin	Sigma (A1978)

Secondary antibodies:

Name	**Reference**
Goat anti-Mouse IgG (H+L) Cross-Adsorbed Secondary Antibody, Alexa Fluor 790	Invitrogen (A11375)
Goat anti-Rabbit IgG (H+L) Cross-Adsorbed Secondary Antibody, Alexa Fluor 680	Cell Signaling (A21076)

Table 4: qPCR primers

RT-qPCR primers sequences

Name	Forward (5' to 3')	Reverse (5' to 3')
HPRT	GACCAGTCAACAGGGGA CAT	CTGCATTGTTTTGCCAGT GT

TET2	TACCGAGACGCTGAGGAAAT	ACATGCTCCATGAACAACCA
AGO2	CAAGTCGGACAGGAGCAGAAAC	GACCTAGCAGTCGCTCTGATCA

ChIP-qPCR primers sequences

Name	Forward (5' to 3')	Reverse (5' to 3')
Neg. Ctrl	GCAGTCGGGGTTGGGGTAA	ACTCAGGCTAGCAGAAACCAA
IL1RN	GGAGGGTATTTCCGCTTCTC	GCCTCTGCAGATTTCCATTC

Table 5: Pyrosequence primers

AGO2 fragment amplification sequences

Name	Forward (5' to 3')	Reverse (5' to 3')
A2_ER1	TGTGTTAGGTATAGTGTTAGGGGTTTA	AATATAATAACCCCTATACCCCCACAACAT
A2_ER2	TGTTTGTTGGAAATTAGGTATATGAG	AACCTTAAAAAAAAAAACCCTATTACTAT

AGO2 fragment sequencing primers

Name	Sequence (5' to 3')
A2_ER1_s1	GTAAGAGAAATAAGAAAATAGAAG
A2_ER1_s2	AGTAAGGGTTTGTTTGGG
A2_ER1_s3	GGGTTAGGGGTTGTG
A2_ER1_s4	AGGTATAGTGTTAGGGGTTTAA
A2_ER2_s1	ACTAAATCCAAACTCTACC
A2_ER2_s2	AAAAACCAACCCAAACAATAAT

References

[1] Lauberth SM, Nakayama T, Wu X, Ferris AL, Tang Z, Hughes SH, et al. H3K4me3 interactions with TAF3 regulate preinitiation complex assembly and selective gene activation. Cell 2013;152:1021–36. https://doi.org/10.1016/j.cell.2013.01.052.

[2] Cano-Rodriguez D, Gjaltema RAF, Jilderda LJ, Jellema P, Dokter-Fokkens J, Ruiters MHJ, et al. Writing of H3K4Me3 overcomes epigenetic silencing in a sustained but context-dependent manner. Nat Commun 2016;7:12284. https://doi.org/10.1038/ncomms12284.

[3] Talbert PB, Meers MP, Henikoff S. Old cogs, new tricks: the evolution of gene expression in a chromatin context. Nat Rev Genet 2019;20:283–97. https://doi.org/10.1038/s41576-019-0105-7.

[4] Calo E, Wysocka J. Modification of enhancer chromatin: what, how, and why? Mol Cell 2013;49:825–37. https://doi.org/10.1016/j.molcel.2013.01.038.

[5] Li G, Levitus M, Bustamante C, Widom J. Rapid spontaneous accessibility of nucleosomal DNA. Nat Struct Mol Biol 2005;12:46–53. https://doi.org/10.1038/nsmb869.

[6] Cosgrove MS, Boeke JD, Wolberger C. Regulated nucleosome mobility and the histone code. Nat Struct Mol Biol 2004;11:1037–43. https://doi.org/10.1038/nsmb851.

[7] Shimko JC, North JA, Bruns AN, Poirier MG, Ottesen JJ. Preparation of fully synthetic histone H3 reveals that acetyl-lysine 56 facilitates protein binding within nucleosomes. J Mol Biol 2011;408:187–204. https://doi.org/10.1016/j.jmb.2011.01.003.

[8] Daujat S, Weiss T, Mohn F, Lange UC, Ziegler-Birling C, Zeissler U, et al. H3K64 trimethylation marks heterochromatin and is dynamically remodeled during developmental reprogramming. Nat Struct Mol Biol 2009;16:777–81. https://doi.org/10.1038/nsmb.1629.

[9] Jambhekar A, Dhall A, Shi Y. Roles and regulation of histone methylation in animal development. Nat Rev Mol Cell Biol 2019;20:625–41. https://doi.org/10.1038/s41580-019-0151-1.

[10] Nacev BA, Feng L, Bagert JD, Lemiesz AE, Gao J, Soshnev A, et al. The expanding landscape of 'oncohistone' mutations in human cancers. Nature 2019;567:473. https://doi.org/10.1038/s41586-019-1038-1.

[11] Hemberger M, Hanna CW, Dean W. Mechanisms of early placental development in mouse and humans. Nat Rev Genet 2020;21:27–43. https://doi.org/10.1038/s41576-019-0169-4.

[12] Hajkova P, Ancelin K, Waldmann T, Lacoste N, Lange UC, Cesari F, et al. Chromatin dynamics during epigenetic reprogramming in the mouse germ line. Nature 2008;452:877–81. https://doi.org/10.1038/nature06714.
[13] Liu X, Wang C, Liu W, Li J, Li C, Kou X, et al. Distinct features of H3K4me3 and H3K27me3 chromatin domains in pre-implantation embryos. Nature 2016;537:558–62. https://doi.org/10.1038/nature19362.
[14] van der Heijden GW, Dieker JW, Derijck AAHA, Muller S, Berden JHM, Braat DDM, et al. Asymmetry in histone H3 variants and lysine methylation between paternal and maternal chromatin of the early mouse zygote. Mech Dev 2005;122:1008–22. https://doi.org/10.1016/j.mod.2005.04.009.
[15] Stäubli A, Peters AH. Mechanisms of maternal intergenerational epigenetic inheritance. Curr Opin Genet Dev 2021;67:151–62. https://doi.org/10.1016/j.gde.2021.01.008.
[16] Bernstein BE, Mikkelsen TS, Xie X, Kamal M, Huebert DJ, Cuff J, et al. A bivalent chromatin structure marks key developmental genes in embryonic stem cells. Cell 2006;125:315–26. https://doi.org/10.1016/j.cell.2006.02.041.
[17] Mikkelsen TS, Ku M, Jaffe DB, Issac B, Lieberman E, Giannoukos G, et al. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature 2007;448:553–60. https://doi.org/10.1038/nature06008.
[18] Krivtsov AV, Armstrong SA. MLL translocations, histone modifications and leukaemia stem-cell development. Nat Rev Cancer 2007;7:823–33. https://doi.org/10.1038/nrc2253.
[19] Salz T, Deng C, Pampo C, Siemann D, Qiu Y, Brown K, et al. Histone Methyltransferase hSETD1A Is a Novel Regulator of Metastasis in Breast Cancer. Mol Cancer Res MCR 2015;13:461–9. https://doi.org/10.1158/1541-7786.MCR-14-0389.
[20] Glozak MA, Seto E. Histone deacetylases and cancer. Oncogene 2007;26:5420–32. https://doi.org/10.1038/sj.onc.1210610.
[21] Dell'Aversana C, Lepore I, Altucci L. HDAC modulation and cell death in the clinic. Exp Cell Res 2012;318:1229–44. https://doi.org/10.1016/j.yexcr.2012.01.025.
[22] West AC, Johnstone RW. New and emerging HDAC inhibitors for cancer treatment. J Clin Invest 2014;124:30–9. https://doi.org/10.1172/JCI69738.
[23] Hargreaves DC, Crabtree GR. ATP-dependent chromatin remodeling: genetics, genomics and mechanisms. Cell Res 2011;21:396–420. https://doi.org/10.1038/cr.2011.32.

[24] Cairns BR. Chromatin remodeling: insights and intrigue from single-molecule studies. Nat Struct Mol Biol 2007;14:989–96. https://doi.org/10.1038/nsmb1333.
[25] Ho L, Crabtree GR. Chromatin remodelling during development. Nature 2010;463:474–84. https://doi.org/10.1038/nature08911.
[26] Bultman S, Gebuhr T, Yee D, La Mantia C, Nicholson J, Gilliam A, et al. A Brg1 null mutation in the mouse reveals functional differences among mammalian SWI/SNF complexes. Mol Cell 2000;6:1287–95. https://doi.org/10.1016/s1097-2765(00)00127-1.
[27] Kidder BL, Palmer S, Knott JG. SWI/SNF-Brg1 regulates self-renewal and occupies core pluripotency-related genes in embryonic stem cells. Stem Cells Dayt Ohio 2009;27:317–28. https://doi.org/10.1634/stemcells.2008-0710.
[28] Barisic D, Stadler MB, Iurlaro M, Schübeler D. Mammalian ISWI and SWI/SNF selectively mediate binding of distinct transcription factors. Nature 2019;569:136–40. https://doi.org/10.1038/s41586-019-1115-5.
[29] Aydin ÖZ, Marteijn JA, Ribeiro-Silva C, Rodríguez López A, Wijgers N, Smeenk G, et al. Human ISWI complexes are targeted by SMARCA5 ATPase and SLIDE domains to help resolve lesion-stalled transcription. Nucleic Acids Res 2014;42:8473–85. https://doi.org/10.1093/nar/gku565.
[30] Klement K, Luijsterburg MS, Pinder JB, Cena CS, Del Nero V, Wintersinger CM, et al. Opposing ISWI- and CHD-class chromatin remodeling activities orchestrate heterochromatic DNA repair. J Cell Biol 2014;207:717–33. https://doi.org/10.1083/jcb.201405077.
[31] Iurlaro M, Masoni F, Flyamer IM, Wirbelauer C, Iskar M, Burger L, et al. Systematic assessment of ISWI subunits shows that NURF creates local accessibility for CTCF. Nat Genet 2024;56:1203–12. https://doi.org/10.1038/s41588-024-01767-x.
[32] Clapier CR, Iwasa J, Cairns BR, Peterson CL. Mechanisms of action and regulation of ATP-dependent chromatin-remodelling complexes. Nat Rev Mol Cell Biol 2017;18:407–22. https://doi.org/10.1038/nrm.2017.26.
[33] Mashtalir N, D'Avino AR, Michel BC, Luo J, Pan J, Otto JE, et al. Modular Organization and Assembly of SWI/SNF Family Chromatin Remodeling Complexes. Cell 2018;175:1272-1288.e20. https://doi.org/10.1016/j.cell.2018.09.032.
[34] Sims RJ, Millhouse S, Chen C-F, Lewis BA, Erdjument-Bromage H, Tempst P, et al. Recognition of trimethylated histone H3 lysine 4 facilitates the recruitment of transcription postinitiation factors and pre-mRNA splicing. Mol Cell 2007;28:665–76. https://doi.org/10.1016/j.molcel.2007.11.010.

[35] Flanagan JF, Mi L-Z, Chruszcz M, Cymborowski M, Clines KL, Kim Y, et al. Double chromodomains cooperate to recognize the methylated histone H3 tail. Nature 2005;438:1181–5. https://doi.org/10.1038/nature04290.
[36] Guzman-Ayala M, Sachs M, Koh FM, Onodera C, Bulut-Karslioglu A, Lin C-J, et al. Chd1 is essential for the high transcriptional output and rapid growth of the mouse epiblast. Dev Camb Engl 2015;142:118–27. https://doi.org/10.1242/dev.114843.
[37] Yen K, Vinayachandran V, Pugh BF. SWR-C and INO80 chromatin remodelers recognize nucleosome-free regions near +1 nucleosomes. Cell 2013;154:1246–56. https://doi.org/10.1016/j.cell.2013.08.043.
[38] Neumann FR, Dion V, Gehlen LR, Tsai-Pflugfelder M, Schmid R, Taddei A, et al. Targeted INO80 enhances subnuclear chromatin movement and ectopic homologous recombination. Genes Dev 2012;26:369–83. https://doi.org/10.1101/gad.176156.111.
[39] Papamichos-Chronakis M, Watanabe S, Rando OJ, Peterson CL. Global regulation of H2A.Z localization by the INO80 chromatin-remodeling enzyme is essential for genome integrity. Cell 2011;144:200–13. https://doi.org/10.1016/j.cell.2010.12.021.
[40] Weber CM, Ramachandran S, Henikoff S. Nucleosomes are context-specific, H2A.Z-modulated barriers to RNA polymerase. Mol Cell 2014;53:819–30. https://doi.org/10.1016/j.molcel.2014.02.014.
[41] Mizuguchi G, Shen X, Landry J, Wu W-H, Sen S, Wu C. ATP-driven exchange of histone H2AZ variant catalyzed by SWR1 chromatin remodeling complex. Science 2004;303:343–8. https://doi.org/10.1126/science.1090701.
[42] Serber DW, Runge JS, Menon DU, Magnuson T. The Mouse INO80 Chromatin-Remodeling Complex Is an Essential Meiotic Factor for Spermatogenesis1. Biol Reprod 2016;94:8, 1–9. https://doi.org/10.1095/biolreprod.115.135533.
[43] Wang L, Du Y, Ward JM, Shimbo T, Lackford B, Zheng X, et al. INO80 facilitates pluripotency gene activation in embryonic stem cell self-renewal, reprogramming, and blastocyst development. Cell Stem Cell 2014;14:575–91. https://doi.org/10.1016/j.stem.2014.02.013.
[44] Shen J, Ju Z, Zhao W, Wang L, Peng Y, Ge Z, et al. ARID1A deficiency promotes mutability and potentiates therapeutic antitumor immunity unleashed by immune checkpoint blockade. Nat Med 2018;24:556–62. https://doi.org/10.1038/s41591-018-0012-z.
[45] D'Alesio C, Punzi S, Cicalese A, Fornasari L, Furia L, Riva L, et al. RNAi screens identify CHD4 as an essential gene in breast cancer growth.

Oncotarget 2016;7:80901–15. https://doi.org/10.18632/oncotarget.12646.
[46] Hwang H-W, Mendell JT. MicroRNAs in cell proliferation, cell death, and tumorigenesis. Br J Cancer 2006;94:776–80. https://doi.org/10.1038/sj.bjc.6603023.
[47] Rosa A, Spagnoli FM, Brivanlou AH. The miR-430/427/302 Family Controls Mesendodermal Fate Specification via Species-Specific Target Selection. Dev Cell 2009;16:517–27. https://doi.org/10.1016/j.devcel.2009.02.007.
[48] Souza OF, Popi AF. Role of microRNAs in B-Cell Compartment: Development, Proliferation and Hematological Diseases. Biomedicines 2022;10:2004. https://doi.org/10.3390/biomedicines10082004.
[49] Asangani IA, Rasheed S a. K, Nikolova DA, Leupold JH, Colburn NH, Post S, et al. MicroRNA-21 (miR-21) post-transcriptionally downregulates tumor suppressor Pdcd4 and stimulates invasion, intravasation and metastasis in colorectal cancer. Oncogene 2008;27:2128–36. https://doi.org/10.1038/sj.onc.1210856.
[50] Renou L, Boelle P-Y, Deswarte C, Spicuglia S, Benyoucef A, Calvo J, et al. Homeobox protein TLX3 activates miR-125b expression to promote T-cell acute lymphoblastic leukemia. Blood Adv 2017;1:733–47. https://doi.org/10.1182/bloodadvances.2017005538.
[51] Pintacuda G, Young AN, Cerase A. Function by Structure: Spotlights on Xist Long Non-coding RNA. Front Mol Biosci 2017;4:90. https://doi.org/10.3389/fmolb.2017.00090.
[52] Bhan A, Soleimani M, Mandal SS. Long Noncoding RNA and Cancer: A New Paradigm. Cancer Res 2017;77:3965–81. https://doi.org/10.1158/0008-5472.CAN-16-2634.
[53] Liu X-H, Sun M, Nie F-Q, Ge Y-B, Zhang E-B, Yin D-D, et al. Lnc RNA HOTAIR functions as a competing endogenous RNA to regulate HER2 expression by sponging miR-331-3p in gastric cancer. Mol Cancer 2014;13:92. https://doi.org/10.1186/1476-4598-13-92.
[54] Du F, Yuan P, Zhao ZT, Yang Z, Wang T, Zhao JD, et al. A miRNA-based signature predicts development of disease recurrence in HER2 positive breast cancer after adjuvant trastuzumab-based treatment. Sci Rep 2016;6:33825. https://doi.org/10.1038/srep33825.
[55] Xing C, Hu X, Xie F, Yu Z, Li H, Bin-Zhou null, et al. Long non-coding RNA HOTAIR modulates c-KIT expression through sponging miR-193a in acute myeloid leukemia. FEBS Lett 2015;589:1981–7. https://doi.org/10.1016/j.febslet.2015.04.061.

[56] Delaunay S, Helm M, Frye M. RNA modifications in physiology and disease: towards clinical applications. Nat Rev Genet 2024;25:104–22. https://doi.org/10.1038/s41576-023-00645-2.
[57] Shi H, Chai P, Jia R, Fan X. Novel insight into the regulatory roles of diverse RNA modifications: Re-defining the bridge between transcription and translation. Mol Cancer 2020;19:78. https://doi.org/10.1186/s12943-020-01194-6.
[58] Frye M, Harada BT, Behm M, He C. RNA modifications modulate gene expression during development. Science 2018;361:1346–9. https://doi.org/10.1126/science.aau1646.
[59] Gattazzo F, Urciuolo A, Bonaldo P. Extracellular matrix: a dynamic microenvironment for stem cell niche. Biochim Biophys Acta 2014;1840:2506–19. https://doi.org/10.1016/j.bbagen.2014.01.010.
[60] Delaunay S, Frye M. RNA modifications regulating cell fate in cancer. Nat Cell Biol 2019;21:552–9. https://doi.org/10.1038/s41556-019-0319-0.
[61] Lan J, Rajan N, Bizet M, Penning A, Singh NK, Guallar D, et al. Functional role of Tet-mediated RNA hydroxymethylcytosine in mouse ES cells and during differentiation. Nat Commun 2020;11:4956. https://doi.org/10.1038/s41467-020-18729-6.
[62] Dominissini D, Moshitch-Moshkovitz S, Schwartz S, Salmon-Divon M, Ungar L, Osenberg S, et al. Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. Nature 2012;485:201–6. https://doi.org/10.1038/nature11112.
[63] Zheng G, Dahl JA, Niu Y, Fedorcsak P, Huang C-M, Li CJ, et al. ALKBH5 is a mammalian RNA demethylase that impacts RNA metabolism and mouse fertility. Mol Cell 2013;49:18–29. https://doi.org/10.1016/j.molcel.2012.10.015.
[64] Jia G, Fu Y, Zhao X, Dai Q, Zheng G, Yang Y, et al. N6-Methyladenosine in nuclear RNA is a major substrate of the obesity-associated FTO. Nat Chem Biol 2011;7:885–7. https://doi.org/10.1038/nchembio.687.
[65] Patil DP, Pickering BF, Jaffrey SR. Reading m6A in the Transcriptome: m6A-Binding Proteins. Trends Cell Biol 2018;28:113–27. https://doi.org/10.1016/j.tcb.2017.10.001.
[66] Geula S, Moshitch-Moshkovitz S, Dominissini D, Mansour AA, Kol N, Salmon-Divon M, et al. m6A mRNA methylation facilitates resolution of naïve pluripotency toward differentiation. Science 2015;347:1002–6. https://doi.org/10.1126/science.1261417.
[67] Batista PJ, Molinie B, Wang J, Qu K, Zhang J, Li L, et al. m(6)A RNA modification controls cell fate transition in mammalian embryonic stem cells. Cell Stem Cell 2014;15:707–19. https://doi.org/10.1016/j.stem.2014.09.019.

[68] Li H-B, Tong J, Zhu S, Batista PJ, Duffy EE, Zhao J, et al. m6A mRNA methylation controls T cell homeostasis by targeting the IL-7/STAT5/SOCS pathways. Nature 2017;548:338–42. https://doi.org/10.1038/nature23450.
[69] Yoon K-J, Ringeling FR, Vissers C, Jacob F, Pokrass M, Jimenez-Cyrus D, et al. Temporal Control of Mammalian Cortical Neurogenesis by m6A Methylation. Cell 2017;171:877-889.e17. https://doi.org/10.1016/j.cell.2017.09.003.
[70] Zhang C, Chen Y, Sun B, Wang L, Yang Y, Ma D, et al. m6A modulates haematopoietic stem and progenitor cell specification. Nature 2017;549:273–6. https://doi.org/10.1038/nature23883.
[71] Sun Y, Gong W, Zhang S. METTL3 promotes colorectal cancer progression through activating JAK1/STAT3 signaling pathway. Cell Death Dis 2023;14:1–14. https://doi.org/10.1038/s41419-023-06287-w.
[72] Vu LP, Pickering BF, Cheng Y, Zaccara S, Nguyen D, Minuesa G, et al. The N6-methyladenosine (m6A)-forming enzyme METTL3 controls myeloid differentiation of normal hematopoietic and leukemia cells. Nat Med 2017;23:1369–76. https://doi.org/10.1038/nm.4416.
[73] Barbieri I, Tzelepis K, Pandolfini L, Shi J, Millán-Zambrano G, Robson SC, et al. Promoter-bound METTL3 maintains myeloid leukaemia by m6A-dependent translation control. Nature 2017;552:126–31. https://doi.org/10.1038/nature24678.
[74] Paris J, Morgan M, Campos J, Spencer GJ, Shmakova A, Ivanova I, et al. Targeting the RNA m6A Reader YTHDF2 Selectively Compromises Cancer Stem Cells in Acute Myeloid Leukemia. Cell Stem Cell 2019;25:137-148.e6. https://doi.org/10.1016/j.stem.2019.03.021.
[75] Mapperley C, van de Lagemaat LN, Lawson H, Tavosanis A, Paris J, Campos J, et al. The mRNA m6A reader YTHDF2 suppresses proinflammatory pathways and sustains hematopoietic stem cell function. J Exp Med 2021;218:e20200829. https://doi.org/10.1084/jem.20200829.
[76] Greenberg MVC, Bourc'his D. The diverse roles of DNA methylation in mammalian development and disease. Nat Rev Mol Cell Biol 2019;20:590–607. https://doi.org/10.1038/s41580-019-0159-6.
[77] Li E, Zhang Y. DNA methylation in mammals. Cold Spring Harb Perspect Biol 2014;6:a019133. https://doi.org/10.1101/cshperspect.a019133.
[78] He Y, Ecker JR. Non-CG Methylation in the Human Genome. Annu Rev Genomics Hum Genet 2015;16:55–77. https://doi.org/10.1146/annurev-genom-090413-025437.
[79] Tomizawa S, Kobayashi H, Watanabe T, Andrews S, Hata K, Kelsey G, et al. Dynamic stage-specific changes in imprinted differentially methylated

regions during early mammalian development and prevalence of non-CpG methylation in oocytes. Dev Camb Engl 2011;138:811–20. https://doi.org/10.1242/dev.061416.

[80] Patil V, Ward RL, Hesson LB. The evidence for functional non-CpG methylation in mammalian cells. Epigenetics 2014;9:823–8. https://doi.org/10.4161/epi.28741.

[81] Ramasamy D, Deva Magendhra Rao AK, Rajkumar T, Mani S. Non-CpG methylation—a key epigenetic modification in cancer. Brief Funct Genomics 2021;20:304–11. https://doi.org/10.1093/bfgp/elab035.

[82] Malone CS, Miner MD, Doerr JR, Jackson JP, Jacobsen SE, Wall R, et al. CmC(A/T)GG DNA methylation in mature B cell lymphoma gene silencing. Proc Natl Acad Sci U S A 2001;98:10404–9. https://doi.org/10.1073/pnas.181206898.

[83] Inoue S, Oishi M. Effects of methylation of non-CpG sequence in the promoter region on the expression of human synaptotagmin XI (syt11). Gene 2005;348:123–34. https://doi.org/10.1016/j.gene.2004.12.044.

[84] Ziller MJ, Müller F, Liao J, Zhang Y, Gu H, Bock C, et al. Genomic Distribution and Inter-Sample Variation of Non-CpG Methylation across Human Cell Types. PLoS Genet 2011;7:e1002389. https://doi.org/10.1371/journal.pgen.1002389.

[85] Laurent L, Wong E, Li G, Huynh T, Tsirigos A, Ong CT, et al. Dynamic changes in the human methylome during differentiation. Genome Res 2010;20:320–31. https://doi.org/10.1101/gr.101907.109.

[86] Lister R, Mukamel EA, Nery JR, Urich M, Puddifoot CA, Johnson ND, et al. Global epigenomic reconfiguration during mammalian brain development. Science 2013;341:1237905. https://doi.org/10.1126/science.1237905.

[87] Imamura T, Kerjean A, Heams T, Kupiec J-J, Thenevin C, Pàldi A. Dynamic CpG and non-CpG methylation of the Peg1/Mest gene in the mouse oocyte and preimplantation embryo. J Biol Chem 2005;280:20171–5. https://doi.org/10.1074/jbc.M501749200.

[88] Shirane K, Toh H, Kobayashi H, Miura F, Chiba H, Ito T, et al. Mouse Oocyte Methylomes at Base Resolution Reveal Genome-Wide Accumulation of Non-CpG Methylation and Role of DNA Methyltransferases. PLOS Genet 2013;9:e1003439. https://doi.org/10.1371/journal.pgen.1003439.

[89] Kouidou S, Agidou T, Kyrkou A, Andreou A, Katopodi T, Georgiou E, et al. Non-CpG cytosine methylation of p53 exon 5 in non-small cell lung carcinoma. Lung Cancer Amst Neth 2005;50:299–307. https://doi.org/10.1016/j.lungcan.2005.06.012.

[90] Li C, Xiong W, Liu X, Xiao W, Guo Y, Tan J, et al. Hypomethylation at non-CpG/CpG sites in the promoter of HIF-1α gene combined with enhanced H3K9Ac modification contribute to maintain higher HIF-1α expression in breast cancer. Oncogenesis 2019;8:1–18. https://doi.org/10.1038/s41389-019-0135-1.
[91] Truong M, Yang B, Wagner J, Desotelle J, Jarrard DF. Analysis of Promoter Non-CG Methylation in Prostate Cancer. Epigenomics 2013;5:65–71. https://doi.org/10.2217/epi.12.67.
[92] Bellizzi D, D'Aquila P, Scafone T, Giordano M, Riso V, Riccio A, et al. The control region of mitochondrial DNA shows an unusual CpG and non-CpG methylation pattern. DNA Res Int J Rapid Publ Rep Genes Genomes 2013;20:537–47. https://doi.org/10.1093/dnares/dst029.
[93] Patil V, Cuenin C, Chung F, Aguilera JRR, Fernandez-Jimenez N, Romero-Garmendia I, et al. Human mitochondrial DNA is extensively methylated in a non-CpG context. Nucleic Acids Res 2019;47:10072–85. https://doi.org/10.1093/nar/gkz762.
[94] Weerts MJA, Sieuwerts AM, Smid M, Look MP, Foekens JA, Sleijfer S, et al. Mitochondrial DNA content in breast cancer: Impact on in vitro and in vivo phenotype and patient prognosis. Oncotarget 2016;7:29166–76. https://doi.org/10.18632/oncotarget.8688.
[95] Guitton R, Nido GS, Tzoulis C. No evidence of extensive non-CpG methylation in mtDNA. Nucleic Acids Res 2022;50:9190–4. https://doi.org/10.1093/nar/gkac701.
[96] Mattei AL, Bailly N, Meissner A. DNA methylation: a historical perspective. Trends Genet TIG 2022;38:676–707. https://doi.org/10.1016/j.tig.2022.03.010.
[97] Deaton AM, Bird A. CpG islands and the regulation of transcription. Genes Dev 2011;25:1010–22. https://doi.org/10.1101/gad.2037511.
[98] Yin Y, Morgunova E, Jolma A, Kaasinen E, Sahu B, Khund-Sayeed S, et al. Impact of cytosine methylation on DNA binding specificities of human transcription factors. Science 2017;356:eaaj2239. https://doi.org/10.1126/science.aaj2239.
[99] Hu S, Wan J, Su Y, Song Q, Zeng Y, Nguyen HN, et al. DNA methylation presents distinct binding sites for human transcription factors. eLife 2013;2:e00726. https://doi.org/10.7554/eLife.00726.
[100] Sardina JL, Collombet S, Tian TV, Gómez A, Di Stefano B, Berenguer C, et al. Transcription Factors Drive Tet2-Mediated Enhancer Demethylation to Reprogram Cell Fate. Cell Stem Cell 2018;23:727-741.e9. https://doi.org/10.1016/j.stem.2018.08.016.

[101] Tanay A, O'Donnell AH, Damelin M, Bestor TH. Hyperconserved CpG domains underlie Polycomb-binding sites. Proc Natl Acad Sci U S A 2007;104:5521–6. https://doi.org/10.1073/pnas.0609746104.
[102] Jermann P, Hoerner L, Burger L, Schübeler D. Short sequences can efficiently recruit histone H3 lysine 27 trimethylation in the absence of enhancer activity and DNA methylation. Proc Natl Acad Sci 2014;111:E3415–21. https://doi.org/10.1073/pnas.1400672111.
[103] Hon GC, Hawkins RD, Caballero OL, Lo C, Lister R, Pelizzola M, et al. Global DNA hypomethylation coupled to repressive chromatin domain formation and gene silencing in breast cancer. Genome Res 2012;22:246–58. https://doi.org/10.1101/gr.125872.111.
[104] Wu H, Coskun V, Tao J, Xie W, Ge W, Yoshikawa K, et al. Dnmt3a-dependent nonpromoter DNA methylation facilitates transcription of neurogenic genes. Science 2010;329:444–8. https://doi.org/10.1126/science.1190485.
[105] Rush M, Appanah R, Lee S, Lam LL, Goyal P, Lorincz MC. Targeting of EZH2 to a defined genomic site is sufficient for recruitment of Dnmt3a but not de novo DNA methylation. Epigenetics 2009;4:404–14. https://doi.org/10.4161/epi.4.6.9392.
[106] Leung DC, Dong KB, Maksakova IA, Goyal P, Appanah R, Lee S, et al. Lysine methyltransferase G9a is required for de novo DNA methylation and the establishment, but not the maintenance, of proviral silencing. Proc Natl Acad Sci U S A 2011;108:5718–23. https://doi.org/10.1073/pnas.1014660108.
[107] Fuks F, Burgers WA, Brehm A, Hughes-Davies L, Kouzarides T. DNA methyltransferase Dnmt1 associates with histone deacetylase activity. Nat Genet 2000;24:88–91. https://doi.org/10.1038/71750.
[108] Estève P-O, Chin HG, Smallwood A, Feehery GR, Gangisetty O, Karpf AR, et al. Direct interaction between DNMT1 and G9a coordinates DNA and histone methylation during replication. Genes Dev 2006;20:3089–103. https://doi.org/10.1101/gad.1463706.
[109] Du Q, Luu P-L, Stirzaker C, Clark SJ. Methyl-CpG-binding domain proteins: readers of the epigenome. Epigenomics 2015;7:1051–73. https://doi.org/10.2217/epi.15.39.
[110] Baubec T, Ivánek R, Lienert F, Schübeler D. Methylation-dependent and -independent genomic targeting principles of the MBD protein family. Cell 2013;153:480–92. https://doi.org/10.1016/j.cell.2013.03.011.
[111] Barau J, Teissandier A, Zamudio N, Roy S, Nalesso V, Hérault Y, et al. The DNA methyltransferase DNMT3C protects male germ cells from transposon activity. Science 2016;354:909–12. https://doi.org/10.1126/science.aah5143.

[112] Ooi SKT, Qiu C, Bernstein E, Li K, Jia D, Yang Z, et al. DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. Nature 2007;448:714–7. https://doi.org/10.1038/nature05987.
[113] Otani J, Nankumo T, Arita K, Inamoto S, Ariyoshi M, Shirakawa M. Structural basis for recognition of H3K4 methylation status by the DNA methyltransferase 3A ATRX-DNMT3-DNMT3L domain. EMBO Rep 2009;10:1235–41. https://doi.org/10.1038/embor.2009.218.
[114] Dhayalan A, Rajavelu A, Rathert P, Tamas R, Jurkowska RZ, Ragozin S, et al. The Dnmt3a PWWP domain reads histone 3 lysine 36 trimethylation and guides DNA methylation. J Biol Chem 2010;285:26114–20. https://doi.org/10.1074/jbc.M109.089433.
[115] Goll MG, Kirpekar F, Maggert KA, Yoder JA, Hsieh C-L, Zhang X, et al. Methylation of tRNAAsp by the DNA methyltransferase homolog Dnmt2. Science 2006;311:395–8. https://doi.org/10.1126/science.1120976.
[116] Bourc'his D, Xu GL, Lin CS, Bollman B, Bestor TH. Dnmt3L and the establishment of maternal genomic imprints. Science 2001;294:2536–9. https://doi.org/10.1126/science.1065848.
[117] Zhang Y, Jurkowska R, Soeroes S, Rajavelu A, Dhayalan A, Bock I, et al. Chromatin methylation activity of Dnmt3a and Dnmt3a/3L is guided by interaction of the ADD domain with the histone H3 tail. Nucleic Acids Res 2010;38:4246–53. https://doi.org/10.1093/nar/gkq147.
[118] Jones PA. Functions of DNA methylation: islands, start sites, gene bodies and beyond. Nat Rev Genet 2012;13:484–92. https://doi.org/10.1038/nrg3230.
[119] Jones PA, Liang G. Rethinking how DNA methylation patterns are maintained. Nat Rev Genet 2009;10:805–11. https://doi.org/10.1038/nrg2651.
[120] Bostick M, Kim JK, Estève P-O, Clark A, Pradhan S, Jacobsen SE. UHRF1 plays a role in maintaining DNA methylation in mammalian cells. Science 2007;317:1760–4. https://doi.org/10.1126/science.1147939.
[121] Jeong M, Sun D, Luo M, Huang Y, Challen GA, Rodriguez B, et al. Large conserved domains of low DNA methylation maintained by Dnmt3a. Nat Genet 2014;46:17–23. https://doi.org/10.1038/ng.2836.
[122] Xie W, Schultz MD, Lister R, Hou Z, Rajagopal N, Ray P, et al. Epigenomic analysis of multilineage differentiation of human embryonic stem cells. Cell 2013;153:1134–48. https://doi.org/10.1016/j.cell.2013.04.022.
[123] Pachano T, Sánchez-Gaya V, Ealo T, Mariner-Faulí M, Bleckwehl T, Asenjo HG, et al. Orphan CpG islands amplify poised enhancer regulatory activity and determine target gene responsiveness. Nat Genet 2021;53:1036–49. https://doi.org/10.1038/s41588-021-00888-x.

[124] Sarda S, Hannenhalli S. Orphan CpG islands as alternative promoters. Transcription 2018;9:171–6. https://doi.org/10.1080/21541264.2017.1373209.
[125] Illingworth RS, Gruenewald-Schneider U, Webb S, Kerr ARW, James KD, Turner DJ, et al. Orphan CpG islands identify numerous conserved promoters in the mammalian genome. PLoS Genet 2010;6:e1001134. https://doi.org/10.1371/journal.pgen.1001134.
[126] Weber M, Hellmann I, Stadler MB, Ramos L, Pääbo S, Rebhan M, et al. Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. Nat Genet 2007;39:457–66. https://doi.org/10.1038/ng1990.
[127] Spainhour JC, Lim HS, Yi SV, Qiu P. Correlation Patterns Between DNA Methylation and Gene Expression in The Cancer Genome Atlas. Cancer Inform 2019;18:1176935119828776. https://doi.org/10.1177/1176935119828776.
[128] Spitz F, Furlong EEM. Transcription factors: from enhancer binding to developmental control. Nat Rev Genet 2012;13:613–26. https://doi.org/10.1038/nrg3207.
[129] Kim T-K, Shiekhattar R. Architectural and Functional Commonalities between Enhancers and Promoters. Cell 2015;162:948–59. https://doi.org/10.1016/j.cell.2015.08.008.
[130] Kribelbauer JF, Laptenko O, Chen S, Martini GD, Freed-Pastor WA, Prives C, et al. Quantitative Analysis of the DNA Methylation Sensitivity of Transcription Factor Complexes. Cell Rep 2017;19:2383–95. https://doi.org/10.1016/j.celrep.2017.05.069.
[131] Maurano MT, Wang H, John S, Shafer A, Canfield T, Lee K, et al. Role of DNA Methylation in Modulating Transcription Factor Occupancy. Cell Rep 2015;12:1184–95. https://doi.org/10.1016/j.celrep.2015.07.024.
[132] Domcke S, Bardet AF, Adrian Ginno P, Hartl D, Burger L, Schübeler D. Competition between DNA methylation and transcription factors determines binding of NRF1. Nature 2015;528:575–9. https://doi.org/10.1038/nature16462.
[133] Hon GC, Song C-X, Du T, Jin F, Selvaraj S, Lee AY, et al. 5mC oxidation by Tet2 modulates enhancer activity and timing of transcriptome reprogramming during differentiation. Mol Cell 2014;56:286–97. https://doi.org/10.1016/j.molcel.2014.08.026.
[134] Hodges E, Molaro A, Dos Santos CO, Thekkat P, Song Q, Uren PJ, et al. Directional DNA methylation changes and complex intermediate states accompany lineage specificity in the adult hematopoietic compartment. Mol Cell 2011;44:17–28. https://doi.org/10.1016/j.molcel.2011.08.026.

[135] Shayevitch R, Askayo D, Keydar I, Ast G. The importance of DNA methylation of exons on alternative splicing. RNA N Y N 2018;24:1351–62. https://doi.org/10.1261/rna.064865.117.
[136] Gelfman S, Cohen N, Yearim A, Ast G. DNA-methylation effect on cotranscriptional splicing is dependent on GC architecture of the exon-intron structure. Genome Res 2013;23:789–99. https://doi.org/10.1101/gr.143503.112.
[137] Shukla S, Kavak E, Gregory M, Imashimizu M, Shutinoski B, Kashlev M, et al. CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing. Nature 2011;479:74–9. https://doi.org/10.1038/nature10442.
[138] Maunakea AK, Chepelev I, Cui K, Zhao K. Intragenic DNA methylation modulates alternative splicing by recruiting MeCP2 to promote exon recognition. Cell Res 2013;23:1256–69. https://doi.org/10.1038/cr.2013.110.
[139] Mellén M, Ayata P, Dewell S, Kriaucionis S, Heintz N. MeCP2 binds to 5hmC enriched within active genes and accessible chromatin in the nervous system. Cell 2012;151:1417–30. https://doi.org/10.1016/j.cell.2012.11.022.
[140] Neri F, Rapelli S, Krepelova A, Incarnato D, Parlato C, Basile G, et al. Intragenic DNA methylation prevents spurious transcription initiation. Nature 2017;543:72–7. https://doi.org/10.1038/nature21373.
[141] Carrozza MJ, Li B, Florens L, Suganuma T, Swanson SK, Lee KK, et al. Histone H3 methylation by Set2 directs deacetylation of coding regions by Rpd3S to suppress spurious intragenic transcription. Cell 2005;123:581–92. https://doi.org/10.1016/j.cell.2005.10.023.
[142] Karimi MM, Goyal P, Maksakova IA, Bilenky M, Leung D, Tang JX, et al. DNA methylation and SETDB1/H3K9me3 regulate predominantly distinct sets of genes, retroelements, and chimeric transcripts in mESCs. Cell Stem Cell 2011;8:676–87. https://doi.org/10.1016/j.stem.2011.04.004.
[143] Teissandier A, Bourc'his D. Gene body DNA methylation conspires with H3K36me3 to preclude aberrant transcription. EMBO J 2017;36:1471–3. https://doi.org/10.15252/embj.201796812.
[144] Maupetit-Méhouas S, Montibus B, Nury D, Tayama C, Wassef M, Kota SK, et al. Imprinting control regions (ICRs) are marked by mono-allelic bivalent chromatin when transcriptionally inactive. Nucleic Acids Res 2016;44:621–35. https://doi.org/10.1093/nar/gkv960.
[145] Proudhon C, Bourc'his D. Identification and resolution of artifacts in the interpretation of imprinted gene expression. Brief Funct Genomics 2010;9:374–84. https://doi.org/10.1093/bfgp/elq020.
[146] Williamson CM, Turner MD, Ball ST, Nottingham WT, Glenister P, Fray M, et al. Identification of an imprinting control region affecting the

expression of all transcripts in the Gnas cluster. Nat Genet 2006;38:350–5. https://doi.org/10.1038/ng1731.

[147] Brind'Amour J, Kobayashi H, Richard Albert J, Shirane K, Sakashita A, Kamio A, et al. LTR retrotransposons transcribed in oocytes drive species-specific and heritable changes in DNA methylation. Nat Commun 2018;9:3331. https://doi.org/10.1038/s41467-018-05841-x.

[148] Veselovska L, Smallwood SA, Saadeh H, Stewart KR, Krueger F, Maupetit-Méhouas S, et al. Deep sequencing and de novo assembly of the mouse oocyte transcriptome define the contribution of transcription to the DNA methylation landscape. Genome Biol 2015;16:209. https://doi.org/10.1186/s13059-015-0769-z.

[149] Peaston AE, Evsikov AV, Graber JH, de Vries WN, Holbrook AE, Solter D, et al. Retrotransposons regulate host genes in mouse oocytes and preimplantation embryos. Dev Cell 2004;7:597–606. https://doi.org/10.1016/j.devcel.2004.09.004.

[150] Takahashi N, Coluccio A, Thorball CW, Planet E, Shi H, Offner S, et al. ZNF445 is a primary regulator of genomic imprinting. Genes Dev 2019;33:49–54. https://doi.org/10.1101/gad.320069.118.

[151] Strogantsev R, Krueger F, Yamazawa K, Shi H, Gould P, Goldman-Roberts M, et al. Allele-specific binding of ZFP57 in the epigenetic regulation of imprinted and non-imprinted monoallelic expression. Genome Biol 2015;16:112. https://doi.org/10.1186/s13059-015-0672-7.

[152] Borgel J, Guibert S, Li Y, Chiba H, Schübeler D, Sasaki H, et al. Targets and dynamics of promoter DNA methylation during early mouse development. Nat Genet 2010;42:1093–100. https://doi.org/10.1038/ng.708.

[153] Mochizuki K, Sharif J, Shirane K, Uranishi K, Bogutz AB, Janssen SM, et al. Repression of germline genes by PRC1.6 and SETDB1 in the early embryo precedes DNA methylation-mediated silencing. Nat Commun 2021;12:7020. https://doi.org/10.1038/s41467-021-27345-x.

[154] Velasco G, Hubé F, Rollin J, Neuillet D, Philippe C, Bouzinba-Segard H, et al. Dnmt3b recruitment through E2F6 transcriptional repressor mediates germ-line gene silencing in murine somatic tissues. Proc Natl Acad Sci U S A 2010;107:9281–6. https://doi.org/10.1073/pnas.1000473107.

[155] Auclair G, Guibert S, Bender A, Weber M. Ontogeny of CpG island methylation and specificity of DNMT3 methyltransferases during embryonic development in the mouse. Genome Biol 2014;15:545. https://doi.org/10.1186/s13059-014-0545-5.

[156] Loda A, Collombet S, Heard E. Gene regulation in time and space during X-chromosome inactivation. Nat Rev Mol Cell Biol 2022;23:231–49. https://doi.org/10.1038/s41580-021-00438-7.
[157] Blewitt ME, Gendrel A-V, Pang Z, Sparrow DB, Whitelaw N, Craig JM, et al. SmcHD1, containing a structural-maintenance-of-chromosomes hinge domain, has a critical role in X inactivation. Nat Genet 2008;40:663–9. https://doi.org/10.1038/ng.142.
[158] Gdula MR, Nesterova TB, Pintacuda G, Godwin J, Zhan Y, Ozadam H, et al. The non-canonical SMC protein SmcHD1 antagonises TAD formation and compartmentalisation on the inactive X chromosome. Nat Commun 2019;10:30. https://doi.org/10.1038/s41467-018-07907-2.
[159] Sidhu SK, Minks J, Chang SC, Cotton AM, Brown CJ. X chromosome inactivation: heterogeneity of heterochromatin. Biochem Cell Biol Biochim Biol Cell 2008;86:370–9. https://doi.org/10.1139/o08-100.
[160] Kusmartsev V, Drożdż M, Schuster-Böckler B, Warnecke T. Cytosine Methylation Affects the Mutability of Neighboring Nucleotides in Germline and Soma. Genetics 2020;214:809–23. https://doi.org/10.1534/genetics.120.303028.
[161] Poulos RC, Olivier J, Wong JWH. The interaction between cytosine methylation and processes of DNA replication and repair shape the mutational landscape of cancer genomes. Nucleic Acids Res 2017;45:7786–95. https://doi.org/10.1093/nar/gkx463.
[162] Nishibuchi G, Déjardin J. The molecular basis of the organization of repetitive DNA-containing constitutive heterochromatin in mammals. Chromosome Res Int J Mol Supramol Evol Asp Chromosome Biol 2017;25:77–87. https://doi.org/10.1007/s10577-016-9547-3.
[163] Gaudet F, Rideout WM, Meissner A, Dausman J, Leonhardt H, Jaenisch R. Dnmt1 expression in pre- and postimplantation embryogenesis and the maintenance of IAP silencing. Mol Cell Biol 2004;24:1640–8. https://doi.org/10.1128/MCB.24.4.1640-1648.2004.
[164] Howell CY, Bestor TH, Ding F, Latham KE, Mertineit C, Trasler JM, et al. Genomic imprinting disrupted by a maternal effect mutation in the Dnmt1 gene. Cell 2001;104:829–38. https://doi.org/10.1016/s0092-8674(01)00280-x.
[165] Hashimoto H, Liu Y, Upadhyay AK, Chang Y, Howerton SB, Vertino PM, et al. Recognition and potential mechanisms for replication and erasure of cytosine hydroxymethylation. Nucleic Acids Res 2012;40:4841–9. https://doi.org/10.1093/nar/gks155.
[166] Inoue A, Shen L, Dai Q, He C, Zhang Y. Generation and replication-dependent dilution of 5fC and 5caC during mouse preimplantation

development. Cell Res 2011;21:1670–6. https://doi.org/10.1038/cr.2011.189.
[167] Valinluck V, Sowers LC. Endogenous cytosine damage products alter the site selectivity of human DNA maintenance methyltransferase DNMT1. Cancer Res 2007;67:946–50. https://doi.org/10.1158/0008-5472.CAN-06-3123.
[168] Frauer C, Hoffmann T, Bultmann S, Casa V, Cardoso MC, Antes I, et al. Recognition of 5-hydroxymethylcytosine by the Uhrf1 SRA domain. PloS One 2011;6:e21306. https://doi.org/10.1371/journal.pone.0021306.
[169] Chen C-C, Wang K-Y, Shen C-KJ. DNA 5-methylcytosine demethylation activities of the mammalian DNA methyltransferases. J Biol Chem 2013;288:9084–91. https://doi.org/10.1074/jbc.M112.445585.
[170] Kunimoto H, McKenney AS, Meydan C, Shank K, Nazir A, Rapaport F, et al. Aid is a key regulator of myeloid/erythroid differentiation and DNA methylation in hematopoietic stem/progenitor cells. Blood 2017;129:1779–90. https://doi.org/10.1182/blood-2016-06-721977.
[171] Català-Moll F, Ferreté-Bonastre AG, Li T, Weichenhan D, Lutsik P, Ciudad L, et al. Activation-induced deaminase is critical for the establishment of DNA methylation patterns prior to the germinal center reaction. Nucleic Acids Res 2021;49:5057–73. https://doi.org/10.1093/nar/gkab322.
[172] Wijesinghe P, Bhagwat AS. Efficient deamination of 5-methylcytosines in DNA by human APOBEC3A, but not by AID or APOBEC3G. Nucleic Acids Res 2012;40:9206–17. https://doi.org/10.1093/nar/gks685.
[173] Lorsbach RB, Moore J, Mathew S, Raimondi SC, Mukatira ST, Downing JR. TET1, a member of a novel protein family, is fused to MLL in acute myeloid leukemia containing the t(10;11)(q22;q23). Leukemia 2003;17:637–41. https://doi.org/10.1038/sj.leu.2402834.
[174] Ko M, An J, Bandukwala HS, Chavez L, Aijö T, Pastor WA, et al. Modulation of TET2 expression and 5-methylcytosine oxidation by the CXXC domain protein IDAX. Nature 2013;497:122–6. https://doi.org/10.1038/nature12052.
[175] Good CR, Madzo J, Patel B, Maegawa S, Engel N, Jelinek J, et al. A novel isoform of TET1 that lacks a CXXC domain is overexpressed in cancer. Nucleic Acids Res 2017;45:8269–81. https://doi.org/10.1093/nar/gkx435.
[176] Lu X, Zhao BS, He C. TET Family Proteins: Oxidation Activity, Interacting Molecules, and Functions in Diseases. Chem Rev 2015;115:2225–39. https://doi.org/10.1021/cr500470n.

[177] Hu L, Lu J, Cheng J, Rao Q, Li Z, Hou H, et al. Structural insight into substrate preference for TET-mediated oxidation. Nature 2015;527:118–22. https://doi.org/10.1038/nature15713.
[178] Weber AR, Krawczyk C, Robertson AB, Kuśnierczyk A, Vågbø CB, Schuermann D, et al. Biochemical reconstitution of TET1–TDG–BER-dependent active DNA demethylation reveals a highly coordinated mechanism. Nat Commun 2016;7:10806. https://doi.org/10.1038/ncomms10806.
[179] Spruijt CG, Gnerlich F, Smits AH, Pfaffeneder T, Jansen PWTC, Bauer C, et al. Dynamic readers for 5-(hydroxy)methylcytosine and its oxidized derivatives. Cell 2013;152:1146–59. https://doi.org/10.1016/j.cell.2013.02.004.
[180] An J, Rao A, Ko M. TET family dioxygenases and DNA demethylation in stem cells and cancers. Exp Mol Med 2017;49:e323. https://doi.org/10.1038/emm.2017.5.
[181] Xiong J, Zhang Z, Chen J, Huang H, Xu Y, Ding X, et al. Cooperative Action between SALL4A and TET Proteins in Stepwise Oxidation of 5-Methylcytosine. Mol Cell 2016;64:913–25. https://doi.org/10.1016/j.molcel.2016.10.013.
[182] Kellinger MW, Song C-X, Chong J, Lu X-Y, He C, Wang D. 5-formylcytosine and 5-carboxylcytosine reduce the rate and substrate specificity of RNA polymerase II transcription. Nat Struct Mol Biol 2012;19:831–3. https://doi.org/10.1038/nsmb.2346.
[183] Raiber E-A, Murat P, Chirgadze DY, Beraldi D, Luisi BF, Balasubramanian S. 5-Formylcytosine alters the structure of the DNA double helix. Nat Struct Mol Biol 2015;22:44–9. https://doi.org/10.1038/nsmb.2936.
[184] Dai H-Q, Wang B-A, Yang L, Chen J-J, Zhu G-C, Sun M-L, et al. TET-mediated DNA demethylation controls gastrulation by regulating Lefty-Nodal signalling. Nature 2016;538:528–32. https://doi.org/10.1038/nature20095.
[185] Dawlaty MM, Breiling A, Le T, Barrasa MI, Raddatz G, Gao Q, et al. Loss of Tet enzymes compromises proper differentiation of embryonic stem cells. Dev Cell 2014;29:102–11. https://doi.org/10.1016/j.devcel.2014.03.003.
[186] Hu X, Zhang L, Mao S-Q, Li Z, Chen J, Zhang R-R, et al. Tet and TDG Mediate DNA Demethylation Essential for Mesenchymal-to-Epithelial Transition in Somatic Cell Reprogramming. Cell Stem Cell 2014;14:512–22. https://doi.org/10.1016/j.stem.2014.01.001.
[187] Zhu F, Zhu Q, Ye D, Zhang Q, Yang Y, Guo X, et al. Sin3a-Tet1 interaction activates gene transcription and is required for embryonic stem cell

pluripotency. Nucleic Acids Res 2018;46:6026–40. https://doi.org/10.1093/nar/gky347.
[188] Dawlaty MM, Breiling A, Le T, Raddatz G, Barrasa MI, Cheng AW, et al. Combined deficiency of Tet1 and Tet2 causes epigenetic abnormalities but is compatible with postnatal development. Dev Cell 2013;24:310–23. https://doi.org/10.1016/j.devcel.2012.12.015.
[189] Rasmussen KD, Helin K. Role of TET enzymes in DNA methylation, development, and cancer. Genes Dev 2016;30:733–50. https://doi.org/10.1101/gad.276568.115.
[190] Charlton J, Jung EJ, Mattei AL, Bailly N, Liao J, Martin EJ, et al. TETs compete with DNMT3 activity in pluripotent cells at thousands of methylated somatic enhancers. Nat Genet 2020;52:819–27. https://doi.org/10.1038/s41588-020-0639-9.
[191] Lazarenkov A, Sardina JL. Dissecting TET2 Regulatory Networks in Blood Differentiation and Cancer. Cancers 2022;14:830. https://doi.org/10.3390/cancers14030830.
[192] Adelman ER, Figueroa ME. Human hematopoiesis: aging and leukemogenic risk. Curr Opin Hematol 2021;28:57–63. https://doi.org/10.1097/MOH.0000000000000622.
[193] Tsagaratou A, Lio C-WJ, Yue X, Rao A. TET Methylcytosine Oxidases in T Cell and B Cell Development and Function. Front Immunol 2017;8:220. https://doi.org/10.3389/fimmu.2017.00220.
[194] Kunimoto H, Fukuchi Y, Sakurai M, Takubo K, Okamoto S, Nakajima H. Tet2-mutated myeloid progenitors possess aberrant in vitro self-renewal capacity. Blood 2014;123:2897–9. https://doi.org/10.1182/blood-2014-01-552471.
[195] Joshi K, Zhang L, Breslin S J P, Kini AR, Zhang J. Role of TET dioxygenases in the regulation of both normal and pathological hematopoiesis. J Exp Clin Cancer Res CR 2022;41:294. https://doi.org/10.1186/s13046-022-02496-x.
[196] Farlik M, Halbritter F, Müller F, Choudry FA, Ebert P, Klughammer J, et al. DNA Methylation Dynamics of Human Hematopoietic Stem Cell Differentiation. Cell Stem Cell 2016;19:808–22. https://doi.org/10.1016/j.stem.2016.10.019.
[197] Huerga Encabo H, Aramburu IV, Garcia-Albornoz M, Piganeau M, Wood H, Song A, et al. Loss of TET2 in human hematopoietic stem cells alters the development and function of neutrophils. Cell Stem Cell 2023;30:781-799.e9. https://doi.org/10.1016/j.stem.2023.05.004.
[198] Koh KP, Yabuuchi A, Rao S, Huang Y, Cunniff K, Nardone J, et al. Tet1 and Tet2 regulate 5-hydroxymethylcytosine production and cell lineage

specification in mouse embryonic stem cells. Cell Stem Cell 2011;8:200–13. https://doi.org/10.1016/j.stem.2011.01.008.
[199] Ravichandran M, Lei R, Tang Q, Zhao Y, Lee J, Ma L, et al. Rinf Regulates Pluripotency Network Genes and Tet Enzymes in Embryonic Stem Cells. Cell Rep 2019;28:1993-2003.e5. https://doi.org/10.1016/j.celrep.2019.07.080.
[200] Kallin EM, Rodríguez-Ubreva J, Christensen J, Cimmino L, Aifantis I, Helin K, et al. Tet2 facilitates the derepression of myeloid target genes during CEBPα-induced transdifferentiation of pre-B cells. Mol Cell 2012;48:266–76. https://doi.org/10.1016/j.molcel.2012.08.007.
[201] Huang F, Sun J, Chen W, He X, Zhu Y, Dong H, et al. HDAC4 inhibition disrupts TET2 function in high-risk MDS and AML. Aging 2020;12:16759–74. https://doi.org/10.18632/aging.103605.
[202] Cheng J, Guo S, Chen S, Mastriano SJ, Liu C, D'Alessio AC, et al. An extensive network of TET2-targeting MicroRNAs regulates malignant hematopoiesis. Cell Rep 2013;5:471–81. https://doi.org/10.1016/j.celrep.2013.08.050.
[203] Song SJ, Ito K, Ala U, Kats L, Webster K, Sun SM, et al. The oncogenic microRNA miR-22 targets the TET2 tumor suppressor to promote hematopoietic stem cell self-renewal and transformation. Cell Stem Cell 2013;13:87–101. https://doi.org/10.1016/j.stem.2013.06.003.
[204] Sun J, He X, Zhu Y, Ding Z, Dong H, Feng Y, et al. SIRT1 Activation Disrupts Maintenance of Myelodysplastic Syndrome Stem and Progenitor Cells by Restoring TET2 Function. Cell Stem Cell 2018;23:355-369.e9. https://doi.org/10.1016/j.stem.2018.07.018.
[205] Wang Y, Zhang Y. Regulation of TET protein stability by calpains. Cell Rep 2014;6:278–84. https://doi.org/10.1016/j.celrep.2013.12.031.
[206] Jeong JJ, Gu X, Nie J, Sundaravel S, Liu H, Kuo W-L, et al. Cytokine-Regulated Phosphorylation and Activation of TET2 by JAK2 in Hematopoiesis. Cancer Discov 2019;9:778–95. https://doi.org/10.1158/2159-8290.CD-18-1138.
[207] Bauer C, Göbel K, Nagaraj N, Colantuoni C, Wang M, Müller U, et al. Phosphorylation of TET proteins is regulated via O-GlcNAcylation by the O-linked N-acetylglucosamine transferase (OGT). J Biol Chem 2015;290:4801–12. https://doi.org/10.1074/jbc.M114.605881.
[208] Deplus R, Delatte B, Schwinn MK, Defrance M, Méndez J, Murphy N, et al. TET2 and TET3 regulate GlcNAcylation and H3K4 methylation through OGT and SET1/COMPASS. EMBO J 2013;32:645–55. https://doi.org/10.1038/emboj.2012.357.
[209] Wang X, Rosikiewicz W, Sedkov Y, Martinez T, Hansen BS, Schreiner P, et al. PROSER1 mediates TET2 O-GlcNAcylation to regulate DNA

demethylation on UTX-dependent enhancers and CpG islands. Life Sci Alliance 2022;5:e202101228. https://doi.org/10.26508/lsa.202101228.
[210] Zhang YW, Wang Z, Xie W, Cai Y, Xia L, Easwaran H, et al. Acetylation Enhances TET2 Function in Protecting against Abnormal DNA Methylation during Oxidative Stress. Mol Cell 2017;65:323–35. https://doi.org/10.1016/j.molcel.2016.12.013.
[211] Nakagawa T, Lv L, Nakagawa M, Yu Y, Yu C, D'Alessio AC, et al. CRL4(VprBP) E3 ligase promotes monoubiquitylation and chromatin binding of TET dioxygenases. Mol Cell 2015;57:247–60. https://doi.org/10.1016/j.molcel.2014.12.002.
[212] Chen L-L, Smith MD, Lv L, Nakagawa T, Li Z, Sun S-C, et al. USP15 suppresses tumor immunity via deubiquitylation and inactivation of TET2. Sci Adv 2020;6:eabc9730. https://doi.org/10.1126/sciadv.abc9730.
[213] Ko M, Bandukwala HS, An J, Lamperti ED, Thompson EC, Hastie R, et al. Ten-Eleven-Translocation 2 (TET2) negatively regulates homeostasis and differentiation of hematopoietic stem cells in mice. Proc Natl Acad Sci U S A 2011;108:14566–71. https://doi.org/10.1073/pnas.1112317108.
[214] Quivoron C, Couronné L, Della Valle V, Lopez CK, Plo I, Wagner-Ballon O, et al. TET2 inactivation results in pleiotropic hematopoietic abnormalities in mouse and is a recurrent event during human lymphomagenesis. Cancer Cell 2011;20:25–38. https://doi.org/10.1016/j.ccr.2011.06.003.
[215] Moran-Crusio K, Reavie L, Shih A, Abdel-Wahab O, Ndiaye-Lobry D, Lobry C, et al. Tet2 loss leads to increased hematopoietic stem cell self-renewal and myeloid transformation. Cancer Cell 2011;20:11–24. https://doi.org/10.1016/j.ccr.2011.06.001.
[216] Morante-Palacios O, Godoy-Tena G, Calafell-Segura J, Ciudad L, Martínez-Cáceres EM, Sardina JL, et al. Vitamin C enhances NF-κB-driven epigenomic reprogramming and boosts the immunogenic properties of dendritic cells. Nucleic Acids Res 2022;50:10981–94. https://doi.org/10.1093/nar/gkac941.
[217] Cimmino L, Dolgalev I, Wang Y, Yoshimi A, Martin GH, Wang J, et al. Restoration of TET2 Function Blocks Aberrant Self-Renewal and Leukemia Progression. Cell 2017;170:1079-1095.e20. https://doi.org/10.1016/j.cell.2017.07.032.
[218] Figueroa ME, Abdel-Wahab O, Lu C, Ward PS, Patel J, Shih A, et al. Leukemic IDH1 and IDH2 mutations result in a hypermethylation phenotype, disrupt TET2 function, and impair hematopoietic differentiation. Cancer Cell 2010;18:553–67. https://doi.org/10.1016/j.ccr.2010.11.015.

[219] Agathocleous M, Meacham CE, Burgess RJ, Piskounova E, Zhao Z, Crane GM, et al. Ascorbate regulates haematopoietic stem cell function and leukaemogenesis. Nature 2017;549:476–81. https://doi.org/10.1038/nature23876.
[220] Xiao M, Yang H, Xu W, Ma S, Lin H, Zhu H, et al. Inhibition of α-KG-dependent histone and DNA demethylases by fumarate and succinate that are accumulated in mutations of FH and SDH tumor suppressors. Genes Dev 2012;26:1326–38. https://doi.org/10.1101/gad.191056.112.
[221] Delhommeau F, Dupont S, Della Valle V, James C, Trannoy S, Massé A, et al. Mutation in TET2 in myeloid cancers. N Engl J Med 2009;360:2289–301. https://doi.org/10.1056/NEJMoa0810069.
[222] Tefferi A, Pardanani A, Lim K-H, Abdel-Wahab O, Lasho TL, Patel J, et al. TET2 mutations and their clinical correlates in polycythemia vera, essential thrombocythemia and myelofibrosis. Leukemia 2009;23:905–11. https://doi.org/10.1038/leu.2009.47.
[223] He P, Lei J, Zou L-X, Zhou G-Z, Peng L, Deng Q, et al. Effects of hypoxia on DNA hydroxymethylase Tet methylcytosine dioxygenase 2 in a KG-1 human acute myeloid leukemia cell line and its mechanism. Oncol Lett 2021;22:692. https://doi.org/10.3892/ol.2021.12953.
[224] Laukka T, Mariani CJ, Ihantola T, Cao JZ, Hokkanen J, Kaelin WG, et al. Fumarate and Succinate Regulate Expression of Hypoxia-inducible Genes via TET Enzymes. J Biol Chem 2016;291:4256–65. https://doi.org/10.1074/jbc.M115.688762.
[225] Pantier R, Tatar T, Colby D, Chambers I. Endogenous epitope-tagging of Tet1, Tet2 and Tet3 identifies TET2 as a naïve pluripotency marker. Life Sci Alliance 2019;2:e201900516. https://doi.org/10.26508/lsa.201900516.
[226] Costa Y, Ding J, Theunissen TW, Faiola F, Hore TA, Shliaha PV, et al. NANOG-dependent function of TET1 and TET2 in establishment of pluripotency. Nature 2013;495:370–4. https://doi.org/10.1038/nature11925.
[227] Okashita N, Kumaki Y, Ebi K, Nishi M, Okamoto Y, Nakayama M, et al. PRDM14 promotes active DNA demethylation through the ten-eleven translocation (TET)-mediated base excision repair pathway in embryonic stem cells. Dev Camb Engl 2014;141:269–80. https://doi.org/10.1242/dev.099622.
[228] Guallar D, Bi X, Pardavila JA, Huang X, Saenz C, Shi X, et al. RNA-dependent chromatin targeting of TET2 for endogenous retrovirus control in pluripotent stem cells. Nat Genet 2018;50:443–51. https://doi.org/10.1038/s41588-018-0060-9.

[229] Zhang DE, Zhang P, Wang ND, Hetherington CJ, Darlington GJ, Tenen DG. Absence of granulocyte colony-stimulating factor signaling and neutrophil development in CCAAT enhancer binding protein alpha-deficient mice. Proc Natl Acad Sci U S A 1997;94:569–74. https://doi.org/10.1073/pnas.94.2.569.

[230] van Oevelen C, Collombet S, Vicent G, Hoogenkamp M, Lepoivre C, Badeaux A, et al. C/EBPα Activates Pre-existing and De Novo Macrophage Enhancers during Induced Pre-B Cell Transdifferentiation and Myelopoiesis. Stem Cell Rep 2015;5:232–47. https://doi.org/10.1016/j.stemcr.2015.06.007.

[231] Suzuki T, Shimizu Y, Furuhata E, Maeda S, Kishima M, Nishimura H, et al. RUNX1 regulates site specificity of DNA demethylation by recruitment of DNA demethylation machineries in hematopoietic cells. Blood Adv 2017;1:1699–711. https://doi.org/10.1182/bloodadvances.2017005710.

[232] Rampal R, Alkalin A, Madzo J, Vasanthakumar A, Pronier E, Patel J, et al. DNA hydroxymethylation profiling reveals that WT1 mutations result in loss of TET2 function in acute myeloid leukemia. Cell Rep 2014;9:1841–55. https://doi.org/10.1016/j.celrep.2014.11.004.

[233] Wang Y, Xiao M, Chen X, Chen L, Xu Y, Lv L, et al. WT1 recruits TET2 to regulate its target gene expression and suppress leukemia cell proliferation. Mol Cell 2015;57:662–73. https://doi.org/10.1016/j.molcel.2014.12.023.

[234] Álvarez-Errico D, Vento-Tormo R, Sieweke M, Ballestar E. Epigenetic control of myeloid cell differentiation, identity and function. Nat Rev Immunol 2015;15:7–17. https://doi.org/10.1038/nri3777.

[235] de la Rica L, Rodríguez-Ubreva J, García M, Islam ABMMK, Urquiza JM, Hernando H, et al. PU.1 target genes undergo Tet2-coupled demethylation and DNMT3b-mediated methylation in monocyte-to-osteoclast differentiation. Genome Biol 2013;14:R99. https://doi.org/10.1186/gb-2013-14-9-r99.

[236] Morante-Palacios O, Ciudad L, Micheroli R, de la Calle-Fabregat C, Li T, Barbisan G, et al. Coordinated glucocorticoid receptor and MAFB action induces tolerogenesis and epigenome remodeling in dendritic cells. Nucleic Acids Res 2022;50:108–26. https://doi.org/10.1093/nar/gkab1182.

[237] Mendes K, Schmidhofer S, Minderjahn J, Glatz D, Kiesewetter C, Raithel J, et al. The epigenetic pioneer EGR2 initiates DNA demethylation in differentiating monocytes at both stable and transient binding sites. Nat Commun 2021;12:1556. https://doi.org/10.1038/s41467-021-21661-y.

[238] Madzo J, Liu H, Rodriguez A, Vasanthakumar A, Sundaravel S, Caces DBD, et al. Hydroxymethylation at gene regulatory regions directs

stem/early progenitor cell commitment during erythropoiesis. Cell Rep 2014;6:231–44. https://doi.org/10.1016/j.celrep.2013.11.044.
[239] Zhang Q, Zhao K, Shen Q, Han Y, Gu Y, Li X, et al. Tet2 is required to resolve inflammation by recruiting Hdac2 to specifically repress IL-6. Nature 2015;525:389–93. https://doi.org/10.1038/nature15252.
[240] Lio C-WJ, Shukla V, Samaniego-Castruita D, González-Avalos E, Chakraborty A, Yue X, et al. TET enzymes augment activation-induced deaminase (AID) expression via 5-hydroxymethylcytosine modifications at the Aicda superenhancer. Sci Immunol 2019;4:eaau7523. https://doi.org/10.1126/sciimmunol.aau7523.
[241] Kulis M, Merkel A, Heath S, Queirós AC, Schuyler RP, Castellano G, et al. Whole-genome fingerprint of the DNA methylome during human B cell differentiation. Nat Genet 2015;47:746–56. https://doi.org/10.1038/ng.3291.
[242] Guilhamon P, Eskandarpour M, Halai D, Wilson GA, Feber A, Teschendorff AE, et al. Meta-analysis of IDH-mutant cancers identifies EBF1 as an interaction partner for TET2. Nat Commun 2013;4:2166. https://doi.org/10.1038/ncomms3166.
[243] Tsagaratou A, Äijö T, Lio C-WJ, Yue X, Huang Y, Jacobsen SE, et al. Dissecting the dynamic changes of 5-hydroxymethylcytosine in T-cell development and differentiation. Proc Natl Acad Sci U S A 2014;111:E3306-3315. https://doi.org/10.1073/pnas.1412327111.
[244] Yang R, Qu C, Zhou Y, Konkel JE, Shi S, Liu Y, et al. Hydrogen Sulfide Promotes Tet1- and Tet2-Mediated Foxp3 Demethylation to Drive Regulatory T Cell Differentiation and Maintain Immune Homeostasis. Immunity 2015;43:251–63. https://doi.org/10.1016/j.immuni.2015.07.017.
[245] O'Hagan HM, Wang W, Sen S, Destefano Shields C, Lee SS, Zhang YW, et al. Oxidative damage targets complexes containing DNA methyltransferases, SIRT1, and polycomb members to promoter CpG Islands. Cancer Cell 2011;20:606–19. https://doi.org/10.1016/j.ccr.2011.09.012.
[246] Ding N, Bonham EM, Hannon BE, Amick TR, Baylin SB, O'Hagan HM. Mismatch repair proteins recruit DNA methyltransferase 1 to sites of oxidative DNA damage. J Mol Cell Biol 2016;8:244–54. https://doi.org/10.1093/jmcb/mjv050.
[247] Kafer GR, Li X, Horii T, Suetake I, Tajima S, Hatada I, et al. 5-Hydroxymethylcytosine Marks Sites of DNA Damage and Promotes Genome Stability. Cell Rep 2016;14:1283–92. https://doi.org/10.1016/j.celrep.2016.01.035.

[248] Chen L-L, Lin H-P, Zhou W-J, He C-X, Zhang Z-Y, Cheng Z-L, et al. SNIP1 Recruits TET2 to Regulate c-MYC Target Genes and Cellular DNA Damage Response. Cell Rep 2018;25:1485-1500.e4. https://doi.org/10.1016/j.celrep.2018.10.028.
[249] Song C, Wang L, Wu X, Wang K, Xie D, Xiao Q, et al. PML Recruits TET2 to Regulate DNA Modification and Cell Proliferation in Response to Chemotherapeutic Agent. Cancer Res 2018;78:2475–89. https://doi.org/10.1158/0008-5472.CAN-17-3091.
[250] Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. Nature 2021;596:583–9. https://doi.org/10.1038/s41586-021-03819-2.
[251] Varadi M, Bertoni D, Magana P, Paramval U, Pidruchna I, Radhakrishnan M, et al. AlphaFold Protein Structure Database in 2024: providing structure coverage for over 214 million protein sequences. Nucleic Acids Res 2024;52:D368–75. https://doi.org/10.1093/nar/gkad1011.
[252] Szklarczyk D, Kirsch R, Koutrouli M, Nastou K, Mehryary F, Hachilif R, et al. The STRING database in 2023: protein–protein association networks and functional enrichment analyses for any sequenced genome of interest. Nucleic Acids Res 2022;51:D638–46. https://doi.org/10.1093/nar/gkac1000.
[253] Challen GA, Goodell MA. Clonal hematopoiesis: mechanisms driving dominance of stem cell clones. Blood 2020;136:1590–8. https://doi.org/10.1182/blood.2020006510.
[254] Kunimoto H, Nakajima H. TET2: A cornerstone in normal and malignant hematopoiesis. Cancer Sci 2021;112:31–40. https://doi.org/10.1111/cas.14688.
[255] Buscarlet M, Provost S, Zada YF, Barhdadi A, Bourgoin V, Lépine G, et al. DNMT3A and TET2 dominate clonal hematopoiesis and demonstrate benign phenotypes and different genetic predispositions. Blood 2017;130:753–62. https://doi.org/10.1182/blood-2017-04-777029.
[256] Cai Z, Kotzin JJ, Ramdas B, Chen S, Nelanuthala S, Palam LR, et al. Inhibition of Inflammatory Signaling in Tet2 Mutant Preleukemic Cells Mitigates Stress-Induced Abnormalities and Clonal Hematopoiesis. Cell Stem Cell 2018;23:833-849.e5. https://doi.org/10.1016/j.stem.2018.10.013.
[257] Abdel-Wahab O, Mullally A, Hedvat C, Garcia-Manero G, Patel J, Wadleigh M, et al. Genetic characterization of TET1, TET2, and TET3 alterations in myeloid malignancies. Blood 2009;114:144–7. https://doi.org/10.1182/blood-2009-03-210039.
[258] Papaemmanuil E, Gerstung M, Bullinger L, Gaidzik VI, Paschka P, Roberts ND, et al. Genomic Classification and Prognosis in Acute

Myeloid Leukemia. N Engl J Med 2016;374:2209–21. https://doi.org/10.1056/NEJMoa1516192.
[259] Puig I, Tenbaum SP, Chicote I, Arqués O, Martínez-Quintanilla J, Cuesta-Borrás E, et al. TET2 controls chemoresistant slow-cycling cancer cell survival and tumor recurrence. J Clin Invest 2018;128:3887–905. https://doi.org/10.1172/JCI96393.
[260] Elena C, Gallì A, Such E, Meggendorfer M, Germing U, Rizzo E, et al. Integrating clinical features and genetic lesions in the risk assessment of patients with chronic myelomonocytic leukemia. Blood 2016;128:1408–17. https://doi.org/10.1182/blood-2016-05-714030.
[261] Palomo L, Malinverni R, Cabezón M, Xicoy B, Arnan M, Coll R, et al. DNA methylation profile in chronic myelomonocytic leukemia associates with distinct clinical, biological and genetic features. Epigenetics 2018;13:8–18. https://doi.org/10.1080/15592294.2017.1405199.
[262] Das AB, Kakadia PM, Wojcik D, Pemberton L, Browett PJ, Bohlander SK, et al. Clinical remission following ascorbate treatment in a case of acute myeloid leukemia with mutations in TET2 and WT1. Blood Cancer J 2019;9:82. https://doi.org/10.1038/s41408-019-0242-4.
[263] Zhao H, Zhu H, Huang J, Zhu Y, Hong M, Zhu H, et al. The synergy of Vitamin C with decitabine activates TET2 in leukemic cells and significantly improves overall survival in elderly patients with acute myeloid leukemia. Leuk Res 2018;66:1–7. https://doi.org/10.1016/j.leukres.2017.12.009.
[264] Mingay M, Chaturvedi A, Bilenky M, Cao Q, Jackson L, Hui T, et al. Vitamin C-induced epigenomic remodelling in IDH1 mutant acute myeloid leukaemia. Leukemia 2018;32:11–20. https://doi.org/10.1038/leu.2017.171.
[265] Kulis M, Heath S, Bibikova M, Queirós AC, Navarro A, Clot G, et al. Epigenomic analysis detects widespread gene-body DNA hypomethylation in chronic lymphocytic leukemia. Nat Genet 2012;44:1236–42. https://doi.org/10.1038/ng.2443.
[266] Oakes CC, Seifert M, Assenov Y, Gu L, Przekopowitz M, Ruppert AS, et al. DNA methylation dynamics during B cell maturation underlie a continuum of disease phenotypes in chronic lymphocytic leukemia. Nat Genet 2016;48:253–64. https://doi.org/10.1038/ng.3488.
[267] Beekman R, Chapaprieta V, Russiñol N, Vilarrasa-Blasi R, Verdaguer-Dot N, Martens JHA, et al. The reference epigenome and regulatory chromatin landscape of chronic lymphocytic leukemia. Nat Med 2018;24:868–80. https://doi.org/10.1038/s41591-018-0028-4.
[268] Nanan KK, Sturgill DM, Prigge MF, Thenoz M, Dillman AA, Mandler MD, et al. TET-Catalyzed 5-Carboxylcytosine Promotes CTCF Binding to

Suboptimal Sequences Genome-wide. iScience 2019;19:326–39. https://doi.org/10.1016/j.isci.2019.07.041.
[269] Asmar F, Punj V, Christensen J, Pedersen MT, Pedersen A, Nielsen AB, et al. Genome-wide profiling identifies a DNA methylation signature that associates with TET2 mutations in diffuse large B-cell lymphoma. Haematologica 2013;98:1912–20. https://doi.org/10.3324/haematol.2013.088740.
[270] Rosikiewicz W, Chen X, Dominguez PM, Ghamlouch H, Aoufouchi S, Bernard OA, et al. TET2 deficiency reprograms the germinal center B cell epigenome and silences genes linked to lymphomagenesis. Sci Adv 2020;6:eaay5872. https://doi.org/10.1126/sciadv.aay5872.
[271] Dominguez PM, Ghamlouch H, Rosikiewicz W, Kumar P, Béguelin W, Fontán L, et al. TET2 Deficiency Causes Germinal Center Hyperplasia, Impairs Plasma Cell Differentiation, and Promotes B-cell Lymphomagenesis. Cancer Discov 2018;8:1632–53. https://doi.org/10.1158/2159-8290.CD-18-0657.
[272] Fuks F, Hurd PJ, Wolf D, Nan X, Bird AP, Kouzarides T. The methyl-CpG-binding protein MeCP2 links DNA methylation to histone methylation. J Biol Chem 2003;278:4035–40. https://doi.org/10.1074/jbc.M210256200.
[273] Li H, Rauch T, Chen Z-X, Szabó PE, Riggs AD, Pfeifer GP. The histone methyltransferase SETDB1 and the DNA methyltransferase DNMT3A interact directly and localize to promoters silenced in cancer cells. J Biol Chem 2006;281:19489–500. https://doi.org/10.1074/jbc.M513249200.
[274] Milutinovic S, D'Alessio AC, Detich N, Szyf M. Valproate induces widespread epigenetic reprogramming which involves demethylation of specific genes. Carcinogenesis 2007;28:560–71. https://doi.org/10.1093/carcin/bgl167.
[275] Kawamoto K, Okino ST, Place RF, Urakami S, Hirata H, Kikuno N, et al. Epigenetic modifications of RASSF1A gene through chromatin remodeling in prostate cancer. Clin Cancer Res Off J Am Assoc Cancer Res 2007;13:2541–8. https://doi.org/10.1158/1078-0432.CCR-06-2225.
[276] Ou J-N, Torrisani J, Unterberger A, Provençal N, Shikimi K, Karimi M, et al. Histone deacetylase inhibitor Trichostatin A induces global and gene-specific DNA demethylation in human cancer cell lines. Biochem Pharmacol 2007;73:1297–307. https://doi.org/10.1016/j.bcp.2006.12.032.
[277] Dunican DS, Mjoseng HK, Duthie L, Flyamer IM, Bickmore WA, Meehan RR. Bivalent promoter hypermethylation in cancer is linked to the H327me3/H3K4me3 ratio in embryonic stem cells. BMC Biol 2020;18:25. https://doi.org/10.1186/s12915-020-0752-3.

[278] Schlesinger Y, Straussman R, Keshet I, Farkash S, Hecht M, Zimmerman J, et al. Polycomb-mediated methylation on Lys27 of histone H3 pre-marks genes for de novo methylation in cancer. Nat Genet 2007;39:232–6. https://doi.org/10.1038/ng1950.
[279] Ohm JE, McGarvey KM, Yu X, Cheng L, Schuebel KE, Cope L, et al. A stem cell-like chromatin pattern may predispose tumor suppressor genes to DNA hypermethylation and heritable silencing. Nat Genet 2007;39:237–42. https://doi.org/10.1038/ng1972.
[280] Rajagopalan KN, Chen X, Weinberg DN, Chen H, Majewski J, Allis CD, et al. Depletion of H3K36me2 recapitulates epigenomic and phenotypic changes induced by the H3.3K36M oncohistone mutation. Proc Natl Acad Sci U S A 2021;118:e2021795118. https://doi.org/10.1073/pnas.2021795118.
[281] Lu C, Jain SU, Hoelper D, Bechet D, Molden RC, Ran L, et al. Histone H3K36 mutations promote sarcomagenesis through altered histone methylation landscape. Science 2016;352:844–9. https://doi.org/10.1126/science.aac7272.
[282] Tiedemann RL, Hlady RA, Hanavan PD, Lake DF, Tibes R, Lee J-H, et al. Dynamic reprogramming of DNA methylation in SETD2-deregulated renal cell carcinoma. Oncotarget 2016;7:1927–46. https://doi.org/10.18632/oncotarget.6481.
[283] Jenseit A, Camgöz A, Pfister SM, Kool M. EZHIP: a new piece of the puzzle towards understanding pediatric posterior fossa ependymoma. Acta Neuropathol (Berl) 2022;143:1–13. https://doi.org/10.1007/s00401-021-02382-4.
[284] Yamanaka S, Blau HM. Nuclear reprogramming to a pluripotent state by three approaches. Nature 2010;465:704–12. https://doi.org/10.1038/nature09229.
[285] Graf T. Historical origins of transdifferentiation and reprogramming. Cell Stem Cell 2011;9:504–16. https://doi.org/10.1016/j.stem.2011.11.012.
[286] Ladewig J, Koch P, Brüstle O. Leveling Waddington: the emergence of direct programming and the loss of cell fate hierarchies. Nat Rev Mol Cell Biol 2013;14:225–36. https://doi.org/10.1038/nrm3543.
[287] Bueno-Costa A, Piñeyro D, Soler M, Javierre BM, Raurell-Vila H, Subirana-Granés M, et al. B-cell leukemia transdifferentiation to macrophage involves reconfiguration of DNA methylation for long-range regulation. Leukemia 2020;34:1158–62. https://doi.org/10.1038/s41375-019-0643-1.
[288] Rapino F, Robles EF, Richter-Larrea JA, Kallin EM, Martinez-Climent JA, Graf T. C/EBPα induces highly efficient macrophage transdifferentiation

of B lymphoma and leukemia cell lines and impairs their tumorigenicity. Cell Rep 2013;3:1153–63. https://doi.org/10.1016/j.celrep.2013.03.003.
[289] Stik G, Vidal E, Barrero M, Cuartero S, Vila-Casadesús M, Mendieta-Esteban J, et al. CTCF is dispensable for immune cell transdifferentiation but facilitates an acute inflammatory response. Nat Genet 2020;52:655–61. https://doi.org/10.1038/s41588-020-0643-0.
[290] Valcárcel G, López-Rubio AV, Lazarenkov A, Berenguer C, Calafell J, Rodríguez-Ubreva J, et al. Modulating immune cell fate and inflammation through CRISPR-mediated DNA methylation editing 2024:2024.07.10.599183. https://doi.org/10.1101/2024.07.10.599183.
[291] Weng C, Yu F, Yang D, Poeschla M, Liggett LA, Jones MG, et al. Deciphering cell states and genealogies of human haematopoiesis. Nature 2024;627:389–98. https://doi.org/10.1038/s41586-024-07066-z.
[292] Chan H-W, Kurago ZB, Stewart CA, Wilson MJ, Martin MP, Mace BE, et al. DNA methylation maintains allele-specific KIR gene expression in human natural killer cells. J Exp Med 2003;197:245–55. https://doi.org/10.1084/jem.20021127.
[293] Wiencke JK, Butler R, Hsuang G, Eliot M, Kim S, Sepulveda MA, et al. The DNA methylation profile of activated human natural killer cells. Epigenetics 2016;11:363–80. https://doi.org/10.1080/15592294.2016.1163454.
[294] McClatchy J, Strogantsev R, Wolfe E, Lin HY, Mohammadhosseini M, Davis BA, et al. Clonal hematopoiesis related TET2 loss-of-function impedes IL1β-mediated epigenetic reprogramming in hematopoietic stem and progenitor cells. Nat Commun 2023;14:8102. https://doi.org/10.1038/s41467-023-43697-y.
[295] Rasmussen KD, Berest I, Keβler S, Nishimura K, Simón-Carrasco L, Vassiliou GS, et al. TET2 binding to enhancers facilitates transcription factor recruitment in hematopoietic cells. Genome Res 2019;29:564–75. https://doi.org/10.1101/gr.239277.118.
[296] Tulstrup M, Soerensen M, Hansen JW, Gillberg L, Needhamsen M, Kaastrup K, et al. TET2 mutations are associated with hypermethylation at key regulatory enhancers in normal and malignant hematopoiesis. Nat Commun 2021;12:6061. https://doi.org/10.1038/s41467-021-26093-2.
[297] Ito K, Lee J, Chrysanthou S, Zhao Y, Josephs K, Sato H, et al. Non-catalytic Roles of Tet2 Are Essential to Regulate Hematopoietic Stem and Progenitor Cell Homeostasis. Cell Rep 2019;28:2480-2490.e4. https://doi.org/10.1016/j.celrep.2019.07.094.
[298] Wiederschain D, Wee S, Chen L, Loo A, Yang G, Huang A, et al. Single-vector inducible lentiviral RNAi system for oncology target validation. Cell Cycle Georget Tex 2009;8:498–504. https://doi.org/10.4161/cc.8.3.7701.

[299] Broome R, Chernukhin I, Jamieson S, Kishore K, Papachristou EK, Mao S-Q, et al. TET2 is a component of the estrogen receptor complex and controls 5mC to 5hmC conversion at estrogen receptor cis-regulatory regions. Cell Rep 2021;34. https://doi.org/10.1016/j.celrep.2021.108776.
[300] Wang L, Ozark PA, Smith ER, Zhao Z, Marshall SA, Rendleman EJ, et al. TET2 coactivates gene expression through demethylation of enhancers. Sci Adv 2018;4:eaau6986. https://doi.org/10.1126/sciadv.aau6986.
[301] Piccolo FM, Bagci H, Brown KE, Landeira D, Soza-Ried J, Feytout A, et al. Different roles for Tet1 and Tet2 proteins in reprogramming-mediated erasure of imprints induced by EGC fusion. Mol Cell 2013;49:1023–33. https://doi.org/10.1016/j.molcel.2013.01.032.
[302] Adams GE, Chandru A, Cowley SM. Co-repressor, co-activator and general transcription factor: the many faces of the Sin3 histone deacetylase (HDAC) complex. Biochem J 2018;475:3921–32. https://doi.org/10.1042/BCJ20170314.
[303] Zhu C, Xiao F, Lin W. EFTUD2 on innate immunity. Oncotarget 2015;6:32313–4.
[304] Irimia M, Weatheritt RJ, Ellis JD, Parikshak NN, Gonatopoulos-Pournatzis T, Babor M, et al. A highly conserved program of neuronal microexons is misregulated in autistic brains. Cell 2014;159:1511–23. https://doi.org/10.1016/j.cell.2014.11.035.
[305] Tapial J, Ha KCH, Sterne-Weiler T, Gohr A, Braunschweig U, Hermoso-Pulido A, et al. An atlas of alternative splicing profiles and functional associations reveals new regulatory programs and genes that simultaneously express multiple major isoforms. Genome Res 2017;27:1759–68. https://doi.org/10.1101/gr.220962.117.
[306] Tian TV, Sardina JL. Uncovering Sequence-Specific Transcription FactorsTranscription factors Interacting with TET2Ten-eleven translocation 2 (TET2). In: Bogdanovic O, Vermeulen M, editors. TET Proteins DNA Demethylation Methods Protoc., New York, NY: Springer US; 2021, p. 239–50. https://doi.org/10.1007/978-1-0716-1294-1_14.
[307] Choi J, Lysakovskaia K, Stik G, Demel C, Söding J, Tian TV, et al. Evidence for additive and synergistic action of mammalian enhancers during cell fate determination. eLife 2021;10:e65381. https://doi.org/10.7554/eLife.65381.
[308] Schwalb B, Michel M, Zacher B, Frühauf K, Demel C, Tresch A, et al. TT-seq maps the human transient transcriptome. Science 2016;352:1225–8. https://doi.org/10.1126/science.aad9841.
[309] Seegren PV, Harper LR, Downs TK, Zhao X-Y, Viswanathan SB, Stremska ME, et al. Reduced mitochondrial calcium uptake in

macrophages is a major driver of inflammaging. Nat Aging 2023;3:796–812. https://doi.org/10.1038/s43587-023-00436-8.
[310] Chan SM, Majeti R. Role of DNMT3A, TET2, and IDH1/2 mutations in pre-leukemic stem cells in acute myeloid leukemia. Int J Hematol 2013;98:648–57. https://doi.org/10.1007/s12185-013-1407-8.
[311] Heyes E, Wilhelmson AS, Wenzel A, Manhart G, Eder T, Schuster MB, et al. TET2 lesions enhance the aggressiveness of CEBPA-mutant acute myeloid leukemia by rebalancing GATA2 expression. Nat Commun 2023;14:6185. https://doi.org/10.1038/s41467-023-41927-x.
[312] Tschan MP, Shan D, Laedrach J, Eyholzer M, Leibundgut EO, Baerlocher GM, et al. NDRG1/2 expression is inhibited in primary acute myeloid leukemia. Leuk Res 2010;34:393–8. https://doi.org/10.1016/j.leukres.2009.08.037.
[313] Onagoruwa OT, Pal G, Ochu C, Ogunwobi OO. Oncogenic Role of PVT1 and Therapeutic Implications. Front Oncol 2020;10:17. https://doi.org/10.3389/fonc.2020.00017.
[314] Nowak I, Sarshad AA. Argonaute Proteins Take Center Stage in Cancers. Cancers 2021;13:788. https://doi.org/10.3390/cancers13040788.
[315] Bird A. DNA methylation patterns and epigenetic memory. Genes Dev 2002;16:6–21. https://doi.org/10.1101/gad.947102.
[316] Borlenghi E, Cattaneo C, Bertoli D, Cerqui E, Archetti S, Passi A, et al. Prognostic Relevance of NPM1 and FLT3 Mutations in Acute Myeloid Leukaemia, Longterm Follow-Up-A Single Center Experience. Cancers 2022;14:4716. https://doi.org/10.3390/cancers14194716.
[317] Avagyan S, Zon LI. Clonal hematopoiesis and inflammation – the perpetual cycle. Trends Cell Biol 2023;33:695–707. https://doi.org/10.1016/j.tcb.2022.12.001.
[318] Meisel M, Hinterleitner R, Pacis A, Chen L, Earley ZM, Mayassi T, et al. Microbial signals drive pre-leukaemic myeloproliferation in a Tet2-deficient host. Nature 2018;557:580–4. https://doi.org/10.1038/s41586-018-0125-z.
[319] Christou-Kent M, Cuartero S, Garcia-Cabau C, Ruehle J, Naderi J, Erber J, et al. CEBPA phase separation links transcriptional activity and 3D chromatin hubs. Cell Rep 2023;42:112897. https://doi.org/10.1016/j.celrep.2023.112897.
[320] Chen Q, Chen Y, Bian C, Fujiki R, Yu X. TET2 promotes histone O-GlcNAcylation during gene transcription. Nature 2013;493:561–4. https://doi.org/10.1038/nature11742.
[321] Baralle FE, Giudice J. Alternative splicing as a regulator of development and tissue identity. Nat Rev Mol Cell Biol 2017;18:437–51. https://doi.org/10.1038/nrm.2017.27.

[322] Gurnari C, Pagliuca S, Visconte V. Alternative Splicing in Myeloid Malignancies. Biomedicines 2021;9:1844. https://doi.org/10.3390/biomedicines9121844.
[323] Theofilatos D, Ho T, Waitt G, Äijö T, Schiapparelli LM, Soderblom EJ, et al. Deciphering the TET3 interactome in primary thymic developing T cells. iScience 2024;27:109782. https://doi.org/10.1016/j.isci.2024.109782.
[324] de la Rica L, Deniz Ö, Cheng KCL, Todd CD, Cruz C, Houseley J, et al. TET-dependent regulation of retrotransposable elements in mouse embryonic stem cells. Genome Biol 2016;17:234. https://doi.org/10.1186/s13059-016-1096-8.
[325] Flores JC, Sidoli S, Dawlaty MM. Tet2 regulates Sin3a recruitment at active enhancers in embryonic stem cells. iScience 2023;26:107170. https://doi.org/10.1016/j.isci.2023.107170.
[326] Guo L, Hong T, Lee Y-T, Hu X, Pan G, Zhao R, et al. Perturbing TET2 condensation promotes aberrant genome-wide DNA methylation and curtails leukaemia cell growth. Nat Cell Biol 2024. https://doi.org/10.1038/s41556-024-01496-7.
[327] Zhou H-L, Luo G, Wise JA, Lou H. Regulation of alternative splicing by local histone modifications: potential roles for RNA-guided mechanisms. Nucleic Acids Res 2014;42:701–13. https://doi.org/10.1093/nar/gkt875.
[328] Wong JJ-L, Gao D, Nguyen TV, Kwok C-T, van Geldermalsen M, Middleton R, et al. Intron retention is regulated by altered MeCP2-mediated splicing factor recruitment. Nat Commun 2017;8:15134. https://doi.org/10.1038/ncomms15134.
[329] Gao D, Pinello N, Nguyen TV, Thoeng A, Nagarajah R, Holst J, et al. DNA methylation/hydroxymethylation regulate gene expression and alternative splicing during terminal granulopoiesis. Epigenomics 2019;11:95–109. https://doi.org/10.2217/epi-2018-0050.
[330] Green ID, Pinello N, Song R, Lee Q, Halstead JM, Kwok C-T, et al. Macrophage development and activation involve coordinated intron retention in key inflammatory regulators. Nucleic Acids Res 2020;48:6513–29. https://doi.org/10.1093/nar/gkaa435.
[331] Akerman M, Fregoso OI, Das S, Ruse C, Jensen MA, Pappin DJ, et al. Differential connectivity of splicing activators and repressors to the human spliceosome. Genome Biol 2015;16:119. https://doi.org/10.1186/s13059-015-0682-5.
[332] Cobo I, Tanaka TN, Chandra Mangalhara K, Lana A, Yeang C, Han C, et al. DNA methyltransferase 3 alpha and TET methylcytosine dioxygenase 2 restrain mitochondrial DNA-mediated interferon signaling in

macrophages. Immunity 2022;55:1386-1401.e10. https://doi.org/10.1016/j.immuni.2022.06.022.
[333] Gentile C, Berlivet S, Mayran A, Paquette D, Guerard-Millet F, Bajon E, et al. PRC2-Associated Chromatin Contacts in the Developing Limb Reveal a Possible Mechanism for the Atypical Role of PRC2 in HoxA Gene Expression. Dev Cell 2019;50:184-196.e4. https://doi.org/10.1016/j.devcel.2019.05.021.
[334] Kraft K, Yost KE, Murphy SE, Magg A, Long Y, Corces MR, et al. Polycomb-mediated genome architecture enables long-range spreading of H3K27 methylation. Proc Natl Acad Sci U S A 2022;119:e2201883119. https://doi.org/10.1073/pnas.2201883119.
[335] Laurent A, Madigou T, Bizot M, Turpin M, Palierne G, Mahé E, et al. TET2-mediated epigenetic reprogramming of breast cancer cells impairs lysosome biogenesis. Life Sci Alliance 2022;5:e202101283. https://doi.org/10.26508/lsa.202101283.
[336] Tian H, Luan P, Liu Y, Li G. Tet-mediated DNA methylation dynamics affect chromosome organization. Nucleic Acids Res 2024;52:3654–66. https://doi.org/10.1093/nar/gkae054.
[337] Xu Y, Wu F, Tan L, Kong L, Xiong L, Deng J, et al. Genome-wide regulation of 5hmC, 5mC, and gene expression by Tet1 hydroxylase in mouse embryonic stem cells. Mol Cell 2011;42:451–64. https://doi.org/10.1016/j.molcel.2011.04.005.
[338] Hughes AL, Kelley JR, Klose RJ. Understanding the interplay between CpG island-associated gene promoters and H3K4 methylation. Biochim Biophys Acta Gene Regul Mech 2020;1863. https://doi.org/10.1016/j.bbagrm.2020.194567.
[339] Adams FF, Heckl D, Hoffmann T, Talbot SR, Kloos A, Thol F, et al. An optimized lentiviral vector system for conditional RNAi and efficient cloning of microRNA embedded short hairpin RNA libraries. Biomaterials 2017;139:102–15. https://doi.org/10.1016/j.biomaterials.2017.05.032.
[340] Hoffmann T, Hörmann A, Corcokovic M, Zmajkovic J, Hinterndorfer M, Salkanovic J, et al. Precision RNAi using synthetic shRNAmir target sites. eLife 2023;12:RP84792. https://doi.org/10.7554/eLife.84792.
[341] Ramdas B, Lakshmi Reddy P, Mali RS, Pasupuleti SK, Zhang J, Kelley MR, et al. Combined heterozygosity of FLT3 ITD, TET2, and DNMT3A results in aggressive leukemia. JCI Insight 2022;7:e162016. https://doi.org/10.1172/jci.insight.162016.
[342] Caiado F, Kovtonyuk LV, Gonullu NG, Fullin J, Boettcher S, Manz MG. Aging drives Tet2+/- clonal hematopoiesis via IL-1 signaling. Blood 2023;141:886–903. https://doi.org/10.1182/blood.2022016835.

[343] Lin T-L, Nagata Y, Kao H-W, Sanada M, Okuno Y, Huang C-F, et al. Clonal leukemic evolution in myelodysplastic syndromes with TET2 and IDH1/2 mutations. Haematologica 2014;99:28–36. https://doi.org/10.3324/haematol.2013.091249.
[344] Stölzel F, Fordham SE, Nandana D, Lin W-Y, Blair H, Elstob C, et al. Biallelic TET2 mutations confer sensitivity to 5′-azacitidine in acute myeloid leukemia. JCI Insight n.d.;8:e150368. https://doi.org/10.1172/jci.insight.150368.
[345] Awada H, Nagata Y, Goyal A, Asad MF, Patel B, Hirsch CM, et al. Invariant phenotype and molecular association of biallelic TET2 mutant myeloid neoplasia. Blood Adv 2019;3:339–49. https://doi.org/10.1182/bloodadvances.2018024216.
[346] Zhang X, Su J, Jeong M, Ko M, Huang Y, Park HJ, et al. DNMT3A and TET2 compete and cooperate to repress lineage-specific transcription factors in hematopoietic stem cells. Nat Genet 2016;48:1014–23. https://doi.org/10.1038/ng.3610.
[347] López-Moyado IF, Tsagaratou A, Yuita H, Seo H, Delatte B, Heinz S, et al. Paradoxical association of TET loss of function with genome-wide DNA hypomethylation. Proc Natl Acad Sci 2019;116:16933–42. https://doi.org/10.1073/pnas.1903059116.
[348] Kwon SY, Lee J, Kim B, Park J-W, Kwon TK, Kang SH, et al. Complexity in regulation of microRNA machinery components in invasive breast carcinoma. Pathol Oncol Res POR 2014;20:697–705. https://doi.org/10.1007/s12253-014-9750-5.
[349] Piroozian F, Bagheri Varkiyani H, Koolivand M, Ansari M, Afsa M, AtashAbParvar A, et al. The impact of variations in transcription of DICER and AGO2 on exacerbation of childhood B-cell lineage acute lymphoblastic leukaemia. Int J Exp Pathol 2019;100:184–91. https://doi.org/10.1111/iep.12316.
[350] Lopez-Orozco J, Fayad N, Khan JQ, Felix-Lopez A, Elaish M, Rohamare M, et al. The RNA Interference Effector Protein Argonaute 2 Functions as a Restriction Factor Against SARS-CoV-2. J Mol Biol 2023;435:168170. https://doi.org/10.1016/j.jmb.2023.168170.
[351] Naoghare PK, Tak YK, Kim MJ, Han E, Song JM. Knock-down of argonaute 2 (AGO2) induces apoptosis in myeloid leukaemia cells and inhibits siRNA-mediated silencing of transfected oncogenes in HEK-293 cells. Basic Clin Pharmacol Toxicol 2011;109:274–82. https://doi.org/10.1111/j.1742-7843.2011.00716.x.
[352] Iosue I, Quaranta R, Masciarelli S, Fontemaggi G, Batassa EM, Bertolami C, et al. Argonaute 2 sustains the gene expression program driving

human monocytic differentiation of acute myeloid leukemia cells. Cell Death Dis 2013;4:e926–e926. https://doi.org/10.1038/cddis.2013.452.
[353] Ramabadran R, Wang JH, Reyes JM, Guzman AG, Gupta S, Rosas C, et al. DNMT3A-coordinated splicing governs the stem state switch towards differentiation in embryonic and haematopoietic stem cells. Nat Cell Biol 2023;25:528–39. https://doi.org/10.1038/s41556-023-01109-9.
[354] Zakrzewska A, Cui C, Stockhammer OW, Benard EL, Spaink HP, Meijer AH. Macrophage-specific gene functions in Spi1-directed innate immunity. Blood 2010;116:e1–11. https://doi.org/10.1182/blood-2010-01-262873.
[355] Fontana MF, Baccarella A, Kellar D, Oniskey TK, Terinate P, Rosenberg SD, et al. Myeloid expression of the AP-1 transcription factor JUNB modulates outcomes of type 1 and type 2 parasitic infections. Parasite Immunol 2015;37:470–8. https://doi.org/10.1111/pim.12215.
[356] Pundhir S, Bratt Lauridsen FK, Schuster MB, Jakobsen JS, Ge Y, Schoof EM, et al. Enhancer and Transcription Factor Dynamics during Myeloid Differentiation Reveal an Early Differentiation Block in Cebpa null Progenitors. Cell Rep 2018;23:2744–57. https://doi.org/10.1016/j.celrep.2018.05.012.
[357] Huber W, Carey VJ, Gentleman R, Anders S, Carlson M, Carvalho BS, et al. Orchestrating high-throughput genomic analysis with Bioconductor. Nat Methods 2015;12:115–21. https://doi.org/10.1038/nmeth.3252.
[358] Yu G, Wang L-G, He Q-Y. ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. Bioinforma Oxf Engl 2015;31:2382–3. https://doi.org/10.1093/bioinformatics/btv145.
[359] Yu G, Wang L-G, Han Y, He Q-Y. clusterProfiler: an R package for comparing biological themes among gene clusters. Omics J Integr Biol 2012;16:284–7. https://doi.org/10.1089/omi.2011.0118.
[360] Ramírez F, Dündar F, Diehl S, Grüning BA, Manke T. deepTools: a flexible platform for exploring deep-sequencing data. Nucleic Acids Res 2014;42:W187–91. https://doi.org/10.1093/nar/gku365.
[361] Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinforma Oxf Engl 2010;26:841–2. https://doi.org/10.1093/bioinformatics/btq033.
[362] Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. Bioinforma Oxf Engl 2016;32:2847–9. https://doi.org/10.1093/bioinformatics/btw313.
[363] Ramírez F, Ryan DP, Grüning B, Bhardwaj V, Kilpert F, Richter AS, et al. deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Res 2016;44:W160-165. https://doi.org/10.1093/nar/gkw257.

[364] Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res 2015;43:e47. https://doi.org/10.1093/nar/gkv007.
[365] Conway JR, Lex A, Gehlenborg N. UpSetR: an R package for the visualization of intersecting sets and their properties. Bioinforma Oxf Engl 2017;33:2938–40. https://doi.org/10.1093/bioinformatics/btx364.
[366] Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 2014;30:2114–20. https://doi.org/10.1093/bioinformatics/btu170.
[367] Krueger F, Andrews SR. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. Bioinformatics 2011;27:1571–2. https://doi.org/10.1093/bioinformatics/btr167.
[368] Aryee MJ, Jaffe AE, Corrada-Bravo H, Ladd-Acosta C, Feinberg AP, Hansen KD, et al. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. Bioinformatics 2014;30:1363–9. https://doi.org/10.1093/bioinformatics/btu049.
[369] Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. Bioinforma Oxf Engl 2013;29:15–21. https://doi.org/10.1093/bioinformatics/bts635.
[370] Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinforma Oxf Engl 2014;30:923–30. https://doi.org/10.1093/bioinformatics/btt656.
[371] Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 2014;15:550. https://doi.org/10.1186/s13059-014-0550-8.
[372] Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods 2012;9:357–9. https://doi.org/10.1038/nmeth.1923.
[373] Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, et al. Model-based Analysis of ChIP-Seq (MACS). Genome Biol 2008;9:R137. https://doi.org/10.1186/gb-2008-9-9-r137.
[374] Li Q, Brown JB, Huang H, Bickel PJ. Measuring reproducibility of high-throughput experiments. Ann Appl Stat 2011;5:1752–79. https://doi.org/10.1214/11-AOAS466.
[375] Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, et al. MEME SUITE: tools for motif discovery and searching. Nucleic Acids Res 2009;37:W202-208. https://doi.org/10.1093/nar/gkp335.
[376] Servant N, Varoquaux N, Lajoie BR, Viara E, Chen C-J, Vert J-P, et al. HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. Genome Biol 2015;16:259. https://doi.org/10.1186/s13059-015-0831-x.

[377] Ramírez F, Bhardwaj V, Arrigoni L, Lam KC, Grüning BA, Villaveces J, et al. High-resolution TADs reveal DNA sequences underlying genome organization in flies. Nat Commun 2018;9:189. https://doi.org/10.1038/s41467-017-02525-w.